TESIS DOCTORAL

AÑO 2023

INNER FABRIC: modelo bioinspirado para la re-presentación como mapa de prominencia del tejido interno de la composición de imágenes artísticas

ÓSCAR SÁNCHEZ CESTEROS

PROGRAMA DE DOCTORADO EN SISTEMAS INTELIGENTES

DIRECTOR:

Dr. MARIANO RINCÓN ZAMORANO

«'Eso es lo malo', dijo, 'que fue pintado así para siempre y ahora nos quedamos sin saber lo que pasa.»,

JAVIER MARÍAS, Corazón tan Blanco

A mi familia

Resumen

A lo largo de la historia, la humanidad se ha interesado por las imágenes y su composición con la finalidad de controlar su creación y de comprender su contemplación. Pero el análisis de la composición ha sido, y sigue siendo, una tarea compleja y ardua, donde la subjetividad y la necesidad de tener un enfoque más objetivo para su análisis ha sido su talón de Aquiles. La existencia de una estructura subyacente, por debajo de los elementos visuales y representando los pesos visuales como tensiones, facilita este análisis, ya que puede reducir la composición a un mapa de prominencia. Esta estructura es definida desde la psicología del arte como el tejido interno de la composición.

La visión artificial ha utilizado la extracción de características visuales para analizar la composición a través de distintas técnicas. Sin embargo, no se han planteado soluciones con una estructura en un nivel semántico bajo, como la del tejido interno. En esta tesis, se presenta un modelo que permite desde una imagen obtener una representación del tejido interno como mapa de prominencia. Para este fin, se ha llevado a cabo una investigación transversal entre la neurociencia y la psicología del arte, aplicando una metodología bioinspirada para construir un modelo de visión artificial.

Para construir el modelo, se ha estudiado el sistema de percepción visual, localizando el área que podría procesar el tejido interno y cómo se representa y se ha analizado su estructura y funcionalidad. Además, se ha conectado la atención con procesos definidos desde la psicología del arte como la agudización y nivelación. Uno de los aspectos relevantes es que en la estructura de la composición existen dos focos de atracción: el central y el externo, y que, además, existe una preponderancia de la región inferior izquierda Estas cuestiones se han relacionado con la excentricidad y anisotropía de los mapas retinotópicos del área visual del cerebro. Por otro lado, para poder establecer el peso visual que evalúa la prominencia de una región, se ha construido un sistema de color bioinspirado en la retina y el núcleo geniculado lateral, y basado en el sistema de color de Schopenhauer, donde se crea una escala jerárquica del color según la actividad neuronal.

El resultado es un modelo con el nombre de «Inner Fabric» que, a través del escaneo de la imagen, es capaz de localizar las regiones prominentes y construir el mapa progresivamente, región a región El modelo se ha aplicado en un clasificador de tipo de composición y en un buscador de imágenes utilizando el mapa del tejido interno como criterio de búsqueda para recuperar imágenes con composiciones similares con independencia de los estilos, temáticas o contenidos visuales. Los resultados muestran las capacidades del modelo para procesar la composición de la imagen en tareas de visión artificial.

Abstract

Throughout history, humanity has been interested in images and their composition for the purpose of controlling their creation and understanding their contemplation. However, the analysis of composition has been, and continues to be, a complex and challenging task, where subjectivity and the need for a more objective approach to analysis have posed a significant challenge. The presence of an underlying structure beneath the visual elements, representing visual weights as tensions, facilitates this analysis by allowing the composition to be reduced to a saliency map. This structure is defined in the field of art psychology as the inner fabric of composition.

Computer vision has utilized the extraction of visual features to analyze composition through various techniques. Nevertheless, solutions with a low-level semantic structure, such as the inner fabric, have not been proposed. In this thesis, a model is presented that enables the extraction of the inner fabric as a saliency map from an image. To achieve this, a cross-disciplinary investigation bridging neuroscience and art psychology has been conducted, applying a bioinspired methodology to construct a computer vision model.

To construct the model, the visual perception system has been studied, pinpointing the area that could process the inner fabric and how it is represented and analyzing its structure and functionality. Furthermore, attention has been linked to processes defined in the field of art psychology, such as sharpening and leveling. One of the relevant aspects is that within the composition structure, there are two focal points of attraction: the central and the external, and moreover, there is a predominance of the lower-left region. These aspects have been related to the eccentricity and anisotropy of retinotopic maps in the visual area of the brain. On the other hand, to establish the visual weight that assesses the saliency of a region, a bioinspired color system has been constructed based on the retina and the lateral geniculate nucleus, drawing from Schopenhauer's color system, where a hierarchical color scale is created according to neuronal activity.

The result is a model named "Inner Fabric" that, through image scanning, can locate saliency regions and progressively build the map, region by region. The model has been applied in a composition type classifier and an image search engine, using the inner fabric map as a query to retrieve images with similar compositions, irrespective of styles, themes, or visual content. The results demonstrate the model's capabilities in processing image composition in artificial vision tasks.

Agradecimientos

En primer lugar, quiero expresar mi más sincero agradecimiento al director de la tesis y amigo, Mariano Rincón, por su esfuerzo y paciencia. Agradezco profundamente que aceptara el desafío de construir un modelo de visión artificial desde la perspectiva de la neurociencia y la psicología del arte, así como por su contribución al rigor científico cuando fue necesario. También deseo manifestar mi gratitud al departamento de Inteligencia Artificial de la UNED por proporcionar recursos y colaborar en diversos momentos de la investigación.

Quiero rendir homenaje al doctor José Mira, in memoriam, y agradecer que apoyara mi ingreso en el antiguo programa de doctorado en Inteligencia Artificial. Aprecio enormemente el tiempo y esfuerzo que dedicó para comprender una forma de trabajo que se basaba más en procesos creativos propios del arte que de la ciencia.

Expreso mi agradecimiento a la catedrática de filosofía, Pilar López de Santa María, por su traducción al castellano del libro de Schopenhauer «Sobre la visión y los colores». Las conversaciones que mantuvimos me permitieron comprender mejor a Schopenhauer y su visión del mundo. Su teoría, que relaciona el color con la actividad neuronal, constituye uno de los pilares fundamentales de esta tesis.

Agradezco a mi familia el apoyo, cariño y paciencia en todo el proceso de elaborar esta tesis.

Por último, pero no menos importante, agradezco a todas y cada una de las influencias que me han ayudado a entender y comprender lo que implica la composición en la creación y contemplación de las imágenes.

Índice general

Glosario de abreviaturas y acrónimos

BAAT: « *Best Artworks of All Time* », *dataset* de pinturas de 50 autores relevantes desde el Renacimiento al arte Pop. Es utilizado en los casos de uso.

BRF: *Brainstem Reticular Formation* (formación reticular del tronco encefálico) es un conjunto de núcleos interconectados que se encuentran a lo largo del tronco encefálico. Desempeña un papel importante en la regulación de funciones vitales como la conciencia, el sueño, la vigilia y la atención.

CIE LAB: sistema de color basado en la percepción humana de los colores.

CNN: *Convolutional neural networks*, Redes neuronales convolucionales.

CRA: Campo Receptivo Agudizado.

CBIR: *Content-Based Image Retrieve*. Sistema de visión artificial para la recuperación de imágenes a partir del contenido.

CSI: *Color Selectivite Index*. Índice de de selectividad al color de una neurona en las rede neuronales artificiales. Cuantifica la dependencia de una neurona a un estímulo en color en comparación con el mismo estímulo en escala de grises.

DoG: *Difference of Gaussians*, modelo que simula los campos receptivos de las neuronas con un centro estimulador y un entorno inhibidor a partir de las diferencia de dos gaussianas.

FT_CNN: *Algorithm-Based Fault Tolerance for Convolutional Neural Networks.* Variante de una red neuronal de convoluciones que aplica un algoritmo basado en la tolerancia de fallos para mantener la estabilidad del proceso de inferencia frente a errores débiles.

HLS: sistema de color que representa el matiz (H), la luminosidad (L) y la saturación (S).

KNN: algoritmo que utiliza la proximidad para hacer clasificaciones o predicciones.

Log-map: mapa logarítmico.

LSC: *Long Skip Connection*: conexión entre bloques no consecutivos en las redes neuronales de convolución.

NGL: Núcleo Geniculado Lateral.

OCC: *Opponent and Complementary Color*, sistema de color usado esta tesis que relaciona el color con un escala de actividad neuronal.

ReLU: función de activación de las redes neuronales artificiales donde los valores negativos se convierten a 0.

RESOM: variante de los SOM (*Self-organized maps*) que utilizan mapas retinotópicos bioinspirados en la retina.

RGB: Red, Green y Blue, sistema de color tricromático basado en cómo la retina capta las ondas de la luz en tres rangos de longitud: larga, media y corta.

RMSE: *Root Mean Square Error*, la raíz del error cuadrático medio, utilizado para medir la cantidad de error que hay entre dos conjuntos de datos.

SOM: *Self-organized maps*, mapas autorganizados. Modelo de red neuronal artificial con aprendizaje sin supervisión.

VSE: *Visual Semantic Embed*, sistema de visión artificial que etiqueta semánticamente características visuales.

WTA: «*Winner takes All*» algoritmo usado para seleccionar el elemento en un grupo que maximiliza una propiedad.

Índice de figuras

Índice de tablas

Prólogo

Las imágenes siempre nos han fascinado. Imaginemos que caminamos por la oscuridad de la cueva y llegamos a un espacio más amplio, donde en sus paredes, ahora visibles por la luz que portamos, surgen imágenes pintadas miles de años atrás. Reconocemos las formas, comprendemos lo que representan y, además, a través de sus trazos y en la manera en que se sitúan en el espacio, intuimos cómo eran sus autores y cómo se relacionaban con el mundo que les rodeaba. Lo mismo sucede con una fotografía hecha en la otra esquina del mundo, nos puede producir rabia, ilusión, dolor, felicidad o placer. Al igual que las pinturas de la cueva, pintadas hace miles de años, existe un lenguaje universal que comprendemos por el mero hecho de poder ver. La psicología del arte estudió este lenguaje universal a mediados del siglo XX y, con la teoría de la Gestalt, intentó describir su funcionamiento a través de experimentos y análisis con una perspectiva más científica que las aproximaciones anteriores, pero con muchas limitaciones. Con el tiempo, hemos ido comprendiendo este lenguaje y los avances en la neurociencia, sobre todo en neuroimagen, han posibilitado relacionar muchas de aquellas leyes y principios con funciones neuronales. De esta manera, la relación entre este lenguaje universal y las operaciones neuronales implicadas en la visión es muy cercana, y los elementos visuales (líneas, formas, colores o texturas) son analizados ahora como procesos neuronales dentro de áreas muy especializadas de la corteza visual. La relación entre esas operaciones y la imagen es inevitable, tanto para el creador, como para el espectador, ambos usando el mismo sistema.

Cuando un artista visual se sitúa delante de su creación, suele pasar más tiempo mirando que trabajando sobre ella. Existe una dinámica muy parecida a la que tiene un tendero cuando pesa el género en una balanza y busca el equilibrio de los dos platos añadiendo pesos a un lado o moviendo el punto de equilibrio. Esta comparación fue planteada por D. Dondis en «A Primer of Visual Literacy» (Dondis, 1974) y fue el germen de «Image Equilibrium: A Global Image Property for Human-Centered Image Analysis» (Sánchez & Rincón, 2009). El objetivo principal era aplicar conceptos de la percepción visual, estudiados desde la psicología del arte, en la solución de problemas de visión artificial, en concreto, el salto semántico. Como indica Kandisky (Kandinsky, 1969), una línea según su pendiente tiene un significado u otro: si es horizontal representaba «quietud», pero si es vertical, «inquietud». El objetivo era utilizar esas etiquetas de bajo contenido semántico asociadas a líneas, colores o formas, para reducir el salto. Aquella propuesta implementaba la sintaxis visual de Dondis para crear un mapa de tensiones de las regiones prominentes con dos valores, uno local y otro global, con el fin de etiquetarlas posteriormente por su característica visual (contornos, formas, color, textura, etc.) en un nivel semántico bajo.

Aunque aquel modelo era operativo, tenía bastantes limitaciones, sobre todo en la determinación de qué características visuales eran más relevantes en cada región y cómo implementar las relaciones espaciales como la preferencia de la región inferior izquierda o la atracción del centro en la percepción. En noviembre de 2013, hubo que situar la

investigación en la frontera de la psicología del arte con la visión artificial. Se pasó de la imagen a la percepción, de ahí al cerebro en el terreno de la bioinspiración, para comprender a la imagen desde su análisis y establecer una relación diferente entre objeto (imagen) y sujeto (aplicación informática).

En los últimos treinta años, los avances en neuroimagen y su aplicación en el estudio de la anatomía y funcionalidad del sistema de percepción visual han situado a esta disciplina en una posición relevante para el avance de la visión artificial. Había dos cuestiones importantes que resolver: por un lado, la existencia de dos focos de atracción, el centro de la imagen y el exterior; y por el otro, la estructura subyacente a la composición, regulada sólo por tensiones y no por elementos visuales como contornos, formas, color, textura, etc. Rudolf Arnheim había introducido la idea de dos focos de atracción (Arnheim, 1956) a todas las fuerzas y tensiones visuales, lo cual invitaba a pensar que en el sistema de percepción visual existía, de alguna manera, esa estructura y que no podía ser algo inherente sólo a la imagen, sino que era consecuencia de una arquitectura funcional del cerebro. El primer paso fue abandonar la idea de que la imagen era sólo un objeto material que analizar para pasar a que la imagen era un objeto visual representado en el cerebro. El segundo paso fue cambiar el objetivo hacia una estructura subyacente de la composición de la imagen, que Anton Enhrenzweig denominó «tejido interno» (Ehrenzweig, 1967). Esta estructura plantea una percepción horizontal basada en las tensiones de las regiones de la imagen sin extraer características como contornos, formas, color, textura, etc.

En 2015, implementamos un modelo computacional robusto que era capaz de «mirar» una imagen recorriendo su espacio para detectar las regiones más prominentes sin la extracción de características visuales. Es decir, se podría automatizar la obtención del tejido interno, algo inconsciente y obviamente difícil de etiquetar por humanos, y usarlo en tareas de visión artificial. A partir de este momento, la investigación se centró en la justificación de cada elemento y su mejora, con el fin de obtener el modelo que se presenta en este trabajo con el nombre de Inner Fabric.

Todas estas investigaciones han llevado a tener una visión más completa de las relaciones que existen entre la imagen, su composición y su percepción, las cuales son, en realidad, inseparables. Este trabajo se ha realizado con una perspectiva bioinspirada, pero con dos focos: uno desde el sistema de percepción visual, y otro, desde la psicología del arte, centrado en el comportamiento de los sujetos ante las imágenes.

1 Introducción

1.1 Motivación y origen del problema

Las imágenes creadas por seres humanos están presentes allá donde vayamos; las vemos en carteles, en libros, en museos, en las paradas del autobús, en grandes carteles a las entradas de las ciudades, en camisetas y, cómo no, en nuestros móviles. Sobre las imágenes se ha escrito bastante desde distintos enfoques y disciplinas y, hoy en día, superada la semiótica y el estructuralismo—los cuales pretendían analizarlas a partir de estructuras lingüísticas—, viven una edad de oro (Mitchell, 1994) como lo que son: imágenes. Ahondar sobre su influjo y por qué nos cautivan no es fácil, ya que lo visual está tan presente en nuestras vidas y estamos tan inmersos, que es difícil establecer un discurso objetivo sobre ellas. Pensemos en un color y juntemos a diez personas, tendremos diez definiciones y diez nombres distintos, pero si lo analizamos con un dispositivo que detecte el color tendremos una definición estandarizada e inequívoca y diez nombres distintos. Si hablamos de cómo se interpreta el significado de una imagen, el problema es más complejo aún, y no sólo por las sensaciones y sentimientos que producen a los espectadores. Una obra de arte abstracto, por ejemplo, de Hartung, genera opiniones opuestas y también reacciones distintas según la persona: habrá quien lo aborrezca por el sólo hecho de no localizar formas reconocibles, y quien lo alabe por producirle emociones. Las imágenes no están ajenas a la polémica y a la discusión. Sin embargo, hay una parte estable que depende del hecho de que la imagen es un objeto, por ejemplo, un lienzo con pintura acrílica en el caso del cuadro de Hartung, y que el sistema de percepción tiene una estructura neuronal que no varía.

Otro aspecto importante, es la realidad visual que percibe y a la cual se ha adaptado. No es lo mismo la realidad visual de un insecto que sólo distingue cambios de intensidades de la luz (de blanco a negro) que la de algunos mamíferos capaces de captar tres rangos

Figura 1 Criterios compositivos inconscientes en una fotografía hecha al azar.
Nota: (a) fotografía hecha al azar; y (b) estructura compositiva existente en la fotografía.

Figura 2 Representación del tejido interno y de la composición de una imagen.
Nota: (a) representación del tejido interno como mapa de prominencia y (b) representación de la estructura de la composición de la imagen con ejes de equilibrio, regiones y guías.

de longitud de onda de la luz (color). La realidad visual condiciona tanto la arquitectura como la funcionalidad del sistema de percepción, ya que no es lo mismo procesar la intensidad de la luz que el color. Una hormiga no podría percibir el cuadro de Hartung, por muy abstracto que sea y no haya patrones o formas reconocibles, ya que en su realidad visual sólo hay cambios de intensidad de luz.

El interés de la humanidad por las imágenes artísticas y con criterios estéticos es muy antiguo, en los tiempos remotos del paleolítico ya se creaban imágenes, e incluso eran objeto de culto (Calabrese, 1985). Cualquier imagen tiene una composición, incluso las que no son artísticas, por ejemplo, una foto que tomamos al azar desde nuestro móvil (ver Figura 1). La foto contiene procesos de decisión inconscientes que determinan el punto de vista y el encuadre, aunque sólo hayamos levantado el móvil. Esas «decisiones inconscientes» se representan en una estructura subyacente a la composición que se encuentra por debajo de las características visuales de la imagen (contornos, formas, colores o texturas). Cuando vemos dos imágenes distintas por la temática, encontramos similitudes entre ellas porque comparten una misma composición que intuimos en la estructura subyacente en un nivel inconsciente. Esta estructura es como un mapa que representa las tensiones y, además, es independiente al contenido semántico. Da igual que en la escena se muestre un paisaje o un retrato, o incluso formas abstractas, podemos relacionar ambas imágenes reconociendo un mismo patrón de tensiones que no podemos verbalizar con facilidad.

La estructura subyacente de la composición fue estudiada por Ehrenzweig (Ehrenzweig, 1967) en los años 60 desde el psicoanálisis. Esta estructura se relaciona, según el autor, con una percepción horizontal con la que el artista visual gestiona los pulsos inconscientes facilitando el proceso creativo para componer la obra. Es decir, con las decisiones que se deben de tomar para que una región u otra tengan una mayor o menor tensión dependiendo del objetivo creativo. A su vez, esta estructura establece una comunicación entre el creador y espectador de tipo inconsciente y, por lo tanto, no controlada conscientemente (tanto por el creador como por el espectador). Ehrenzweig

denominó a esta estructura «tejido interno» (Inner Fabric en inglés) porque actúa como un tejido entrelazado de la composición estableciendo las fuerzas y contrapesos.

Desde el punto de vista del análisis de la composición, el tejido interno se representa con un mapa que muestra las regiones por su tensión, como podemos ver en el ejemplo de la Figura 2.a. Por el contrario, la estructura de la composición se representa, por ejemplo, con ejes de equilibrio o con regiones que representan un patrón de elementos visuales (ver Figura 2.b).

Otra cuestión es cómo percibimos esta estructura. Imaginemos que acabamos de entrar en una sala de un museo donde hay una pintura y nos situamos frente a ella. Nuestra mirada se sitúa en el punto central, el cuadro está ahí, lo percibimos como un todo con claridad, aunque sólo vemos el detalle de esa región central. Pronto recorremos con nuestra mirada el cuadro sin poder evitarlo, nuestros ojos se lanzan a cada región del espacio visual sin que los controlemos, o por lo menos sin que seamos conscientes de que se mueven de un lado a otro. El hecho de escanear la escena visual es en realidad inconsciente, controlado por operaciones distintas e incluso opuestas, como el procesamiento guiado de abajo hacia arriba, a partir del contenido de la imagen, o sea de las propias tensiones y patrones de los elementos visuales. Pero, también de arriba hacia abajo, a partir de nuestro objetivo final en la percepción, o sea interpretar y comprender el significado. Esto lo podemos comprobar en la Figura 3.a fijando la atención en el punto rojo situado en la región central. En seguida recorremos cada región sin poder parar el proceso y mantener la mirada fija en el punto rojo. Esta batalla entre mover los ojos (inconsciente) y mantener la mirada fija (consciente) está perdida, ya que las regiones de la imagen luchan por captar nuestra atención. Por otro lado, este escaneo de la imagen es necesario, ya que no es posible obtener un conocimiento completo desde una sola región. En la Figura 3.b, se ha etiquetado aleatoriamente cada región con un número para poder comprobar que no es posible ver toda la secuencia completa de números, sino que desde cada región se percibe uno solo.

Figura 3 El movimiento inconsciente de los ojos por la imagen.
Nota: (a) proceso de escaneo de la imagen inconsciente, la prueba es intentar mantener la mirada fija en el punto rojo; y (b) procesamiento por regiones de la imagen a partir del movimiento de ojos, la prueba es intentar ver toda la secuencia de números manteniendo la mirada fija en un punto cualquiera de la imagen.

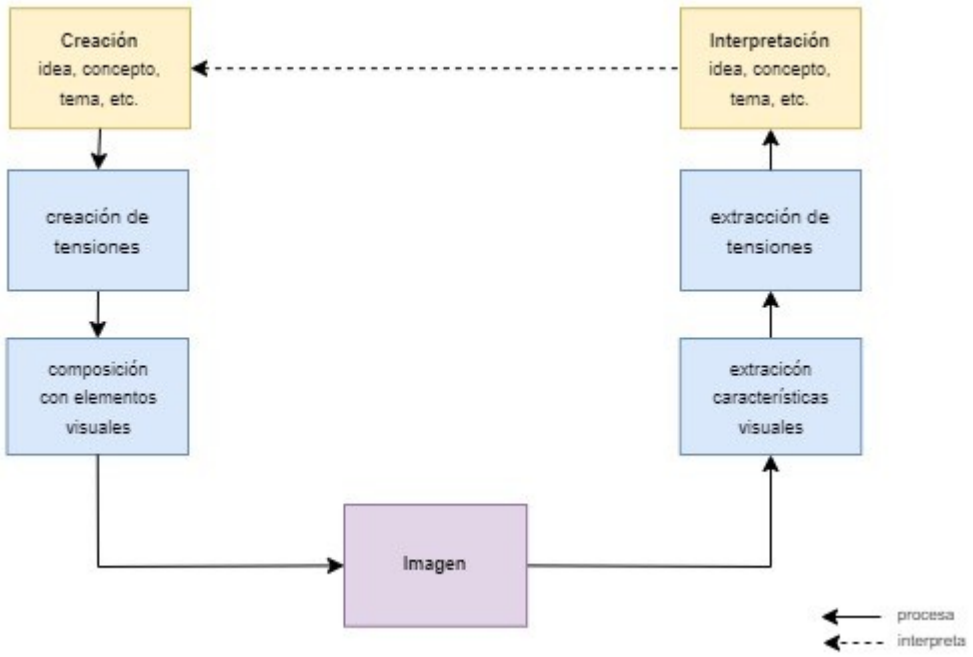


Figura 4 La creación e interpretación de imágenes.

Cada región compite para captar la atención de nuestra mirada sin que nos demos cuenta. Es un proceso de búsqueda, aunque sea inconsciente, y también una experiencia en el tiempo, ya que necesitamos estar en frente del cuadro, mirando, para construir en nuestro cerebro la estructura de la composición y localizar las regiones de interés.

Sin embargo, la percepción horizontal planteada por Ehrenzweig del tejido interno está relacionada con una atención general y global de la imagen y desde un plano inconsciente, mientras que la composición la percibimos conscientemente con una atención focalizada en unas regiones más que otras a través de una percepción vertical. Debemos pensar, por consiguiente, que ambas percepciones están relacionadas de alguna manera permitiéndonos percibir la composición y, a su vez, la estructura subyacente. La cuestión es si es posible separarlas y tener una representación de la estructura subyacente, o sea, un mapa del tejido interno de la composición.

Si fuéramos capaces de resolver esta cuestión, el tejido interno aportaría para el estudio y procesamiento de imágenes una serie de ventajas, ya que facilitaría obtener una representación universal y global, independiente de su contenido semántico. La Figura 4 muestra un esquema que resume esta idea: en la fase de creación, las ideas, los conceptos y el tema son convertidos en una estructura subyacente que representa al patrón de tensiones; después, este patrón se formaliza en la composición a partir de los elementos visuales (puntos, líneas, formas, colores, texturas, etc.) según el tema y contenido. A su vez, el espectador, a través de la percepción, extrae las características visuales conscientemente (contornos, formas, colores, texturas, etc.) e inconscientemente la estructura subyacente con las tensiones de las regiones para interpretar las ideas, conceptos, temas, etc.

Esta representación subyacente es más estable en el tiempo que la composición que podamos percibir de una manera consciente, ya que esta última está condicionada por nuestra manera de comprender e interpretar la imagen. Por ejemplo, cuando vemos un cuadro de un museo y leemos la descripción que hay en la parte inferior, nuestra percepción se ve condicionada por lo que leemos, ya que nos obliga a fijarnos en detalles y características visuales concretas. Además, la estructura subyacente representa los pulsos inconscientes del autor relacionados con la selección de las regiones donde se sitúan los elementos visuales. Es como la huella que queda de la suela del zapato cuando se pisa la nieve o el barro.

Las imágenes viven a pesar de nosotros. Algunas, como las del paleolítico, han estado durante siglos ocultas, pero expectantes a que una luz las iluminara y unos ojos fijaran la atención en ellas para que los pulsos inconscientes del autor llegaran a través del tiempo al nuevo espectador. El tejido interno se nos presenta como una representación ideal para la comprensión de aspectos ajenos a los estilos, las variantes culturales, sociales o incluso políticas, donde los pulsos inconscientes de nuestros cerebros se comunican directamente en un lenguaje directo y natural, y por supuesto, universal.

El tejido interno de la composición plantea un análisis de las imágenes diferente al que se realiza a partir de las características visuales. Esto es más importante en imágenes artísticas y con criterios estéticos, porque las regiones de la imagen son analizadas a partir de si un contorno, una forma, el color o la textura, o todos combinados, son relevantes, y después se establece una estructura de conexiones entre las regiones. Por lo tanto, este tipo de análisis tiene dos etapas: una primera enfocada en las regiones y sus características visuales, y una segunda en la conexión entre las regiones. Sin embargo, el tejido interno plantea un análisis centrado en las tensiones de las regiones de la imagen con independencia de las características visuales presentes y las conexiones que existan. El problema principal es que, primero, hay que descubrir cómo se obtiene la tensión y su relevancia sin las características visuales y, segundo, cómo se representa para su comprensión.

1.2 Objetivo y alcance

En los últimos años, la visión artificial ha demostrado su capacidad para gestionar una gran cantidad de datos, codificando y descodificando información visual para su aplicación en procesos de computación, relacionando incluso características visuales con información verbal. Si bien, aunque su enfoque ha estado centrado en las características visuales (desde la detección de contornos, la segmentación, el reconocimiento de patrones visuales de formas, texturas o colores, etc.), el interés por otros aspectos de la percepción ha tenido un importante desarrollo, sobre todo en la atención. En este campo, destacan los mapas de prominencia (*maps of saliency*) introducidos por (Koch & Ullman, 1987), los cuales detectan y representan las regiones relevantes en el proceso de percepción. Si bien no existe ninguna solución que permita obtener un mapa del tejido interno de una composición, es viable como modelo su reinterpretación para alcanzar este fin. La Figura 5 muestra la implementación de los mapas de prominencia en el es-

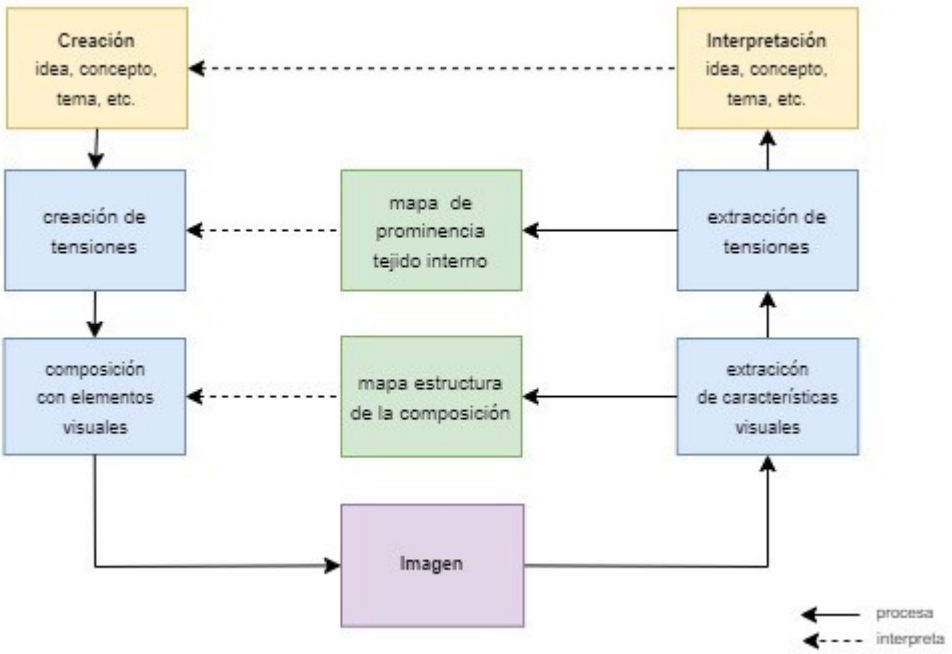


Figura 5 Mapas de prominencia en la creación e interpretación de imágenes.

quema presentado en la Figura 4 en relación con la composición en las fases de creación e interpretación de la imagen. La principal cuestión que se plantea es cómo obtener un mapa de prominencia de la estructura subyacente desde este enfoque.

Para alcanzar este objetivo, este trabajo de investigación se divide en dos fases. La primera, teórica, se centra en analizar el conocimiento existente sobre la visión y su relación con la estructura de la composición de una imagen, con el fin de describir funcionalmente el tejido interno desde la perspectiva de la neurociencia, la psicología del arte y la visión artificial. La segunda fase, de naturaleza computacional, tiene como finalidad, a partir de los conocimientos adquiridos en la primera fase, construir un modelo de visión artificial que obtenga el mapa del tejido interno a partir de una imagen.

El alcance no se limita a la descripción del modelo, sino, además, a su configuración y parametrización para su implementación en herramientas tanto para el análisis de la composición de la imagen como para la búsqueda, clasificación o generación de imágenes. Aunque el modelo puede ser aplicado a cualquier tipo de imagen, tanto la configuración como los casos de estudio, se realizarán en imágenes artísticas como pinturas, dibujos o ilustraciones, y con criterios estéticos.

Si bien el objetivo principal es obtener el mapa del tejido interno, los objetivos secundarios son los siguientes:

- Establecer una representación del tejido interno como el mapa de prominencias para su uso por expertos y para su aplicación en programas informáticos.

- Determinar cómo es la relación funcional de la estructura subyacente con la composición.
- Analizar e implementar la computación de las tensiones de las regiones sin procesar características visuales.

1.3 Estructura de la tesis y secciones del documento

En los últimos años, la transversalidad entre disciplinas se ha puesto encima de la mesa como una oportunidad de avance a partir de intercambios de ideas, metodologías y enfoques entre áreas de conocimiento, a veces opuestas y de difícil conexión. Sin embargo, a pesar del riesgo y con la idea de poder, en ese intercambio, resolver problemas y aportar soluciones no exploradas con anterioridad, esta tesis se origina en la frontera entre las ciencias de la computación, la neurociencia y la psicología del arte con la visión artificial como guía, y en gran medida, receptora de los posibles avances que surgieran.

Una parte importante del proceso de investigación es el estudio y análisis de las distintas teorías y avances sobre la visión y las imágenes desde las tres disciplinas. El capítulo 2 «Fundamentos de la visión, la imagen y la composición» recoge este trabajo, el cual es una pieza teórica importante para comprender, desde un enfoque transversal, cómo afrontar la construcción de un modelo del tejido interno.

La bioinspiración en la visión artificial es un enfoque ideal tanto para integrar las teorías de la neurociencia como las de la psicología del arte, pero no existe una metodología estandarizada que facilite el trabajo. En este sentido, el estudio «Bio-inspired computer vision: Towards a synergistic approach of artificial and biological vision» sobre la bioinspiración en visión artificial por Medathati y colaboradores (Medathati y otros, 2016) es el más avanzado y es el que se utilizará en la investigación; aparece descrito en el capítulo 3.

El objetivo es, desde el principio, construir un modelo de computación que permita, a partir de una imagen, obtener el mapa del tejido interno de la composición. En el capítulo 4 «Modelo Inner Fabric» se describen los distintos componentes como representaciones, estructuras de datos y operaciones, así como el algoritmo completo que describe la implementación desde la entrada de la imagen, hasta la obtención del mapa del tejido interno. Otro aspecto clave del modelo, son sus parámetros y coeficientes configurables, los cuales son necesarios para, a través de su ajuste, alcanzar el objetivo u objetivos que se planteen. Para llevar a cabo la configuración, es necesario crear un proceso de ajuste con el fin de que el modelo pueda obtener el mapa del tejido interno para un tipo de imagen concreto. La configuración se realizará para imágenes artísticas con criterios estéticos para las pruebas y casos de uso.

El tejido interno es una estructura subyacente de la composición de una imagen visible para un experto. Para el resto lo es también, pero con dificultad. Esto presenta un hándicap para la tesis, ya que el resultado del modelo es difícilmente evaluable con

herramientas convencionales. Además, por su propia complejidad, no existen *datasets* etiquetados o modelos previos con los que comparar, con lo que la solución es realizar experimentos con algunos componentes del modelo y con los mapas del tejido obtenidos en tareas habituales de visión artificial como la clasificación o búsqueda de imágenes.

Los experimentos se describen en el capítulo 5 «Casos de uso». En el primer bloque, se implementan dos aplicaciones relacionadas con componentes específicos del modelo:

- Una conversión de color a escala de grises que utiliza el sistema creado para obtener a la tensión de cada región a partir de los píxeles en RGB.
- La aplicación de una *long skip connection* a las arquitecturas de las redes neuronales convolucionales para mejorar su selectividad al color.

En el segundo bloque, se presentan dos aplicaciones del tejido interno en tareas de visión artificial:

- Una clasificación del tipo de composición para pinturas que utiliza una red neuronal artificial con el mapa de prominencia del tejido interno como entrada y entrenada con un *dataset* sintético, creado gracias a la estructura simple de estos mapas.
- Una busqueda CBIR (recuperador de imágenes basado en el contenido) en una base de datos de pinturas utilizando los mapas de prominencia del tejido interno como criterio de búsqueda.

En resumen, el capítulo 2 se titula «Fundamentos de la visión, la imagen y la composición» y tiene como objetivo recopilar y analizar el conocimiento sobre la visión, la imagen y la composición desde las perspectivas de la neurociencia, la psicología del arte y la visión artificial. El capítulo 3, «Metodología de la bioinspiración», describe cómo se utilizará para construir el modelo. El capítulo 4, «Modelo Inner Fabric», aplica la metodología para describir computacionalmente todos los componentes y su implementación como programa informático, así como la configuración necesaria para obtener los mapas de prominencia del tejido interno de imágenes artísticas. Por último, el capítulo 5, «Casos de uso», presenta varias aplicaciones basadas en el modelo Inner Fabric configurado para imágenes artísticas.

2 Fundamentos de la visión, la imagen y la composición

El interés por el estudio de la visión es antiguo, lo que ha llevado a diversas aproximaciones a lo largo del tiempo, desde el estudio de la anatomía de los órganos visuales hasta el análisis del comportamiento psicológico, pasando por la investigación de los fenómenos físicos y químicos involucrados. En el campo de la visión artificial, este interés ha sido y sigue siendo multidisciplinar, de ahí que la relación con la neurociencia y la psicología haya sido estrecha y bidireccional. Muchos de los modelos implementados en visión artificial se han utilizado para poner a prueba teorías psicológicas y realizar investigaciones neurocientíficas, y viceversa, en un enfoque que también involucra la bioinspiración.

En este capítulo, se recopilan las teorías y avances científicos sobre la visión desde las perspectivas de la psicología del arte, la neurociencia y la visión artificial, especialmente en relación con la composición de la imagen y el tejido interno. El capítulo se divide en tres subsecciones según cada área: en la primera, se definen y analizan las imágenes desde la perspectiva de la psicología de la percepción y la psicología del arte, así como la teoría de la atención, todos relacionados con la composición de la imagen; en la segunda, se presentan los avances en el conocimiento anatómico y fisiológico del sistema de percepción visual humano, con un enfoque especial en el área precortical (retina y núcleo geniculado lateral o NGL) y los esquemas de referencia del área parietal; y en la tercera, se exponen los principales problemas y paradigmas de la visión artificial en relación con la visión, con un énfasis en la bioinspiración como enfoque para crear modelos implementables y funcionales.

Con el fin de facilitar la comprensión y la independencia entre los capítulos de la tesis, esta sección se centra únicamente en la exposición de los conocimientos sin formular hipótesis, que se reservan para el capítulo del modelo «Inner Fabric». En ese capítulo, se hará referencia a estas teorías y descubrimientos sobre la visión en la implementación directa de cada componente desde una perspectiva computacional. Por lo tanto, este capítulo puede tener dos propósitos: proporcionar una comprensión básica de los aspectos relacionados con la visión desde diversas perspectivas en relación con la composición de imágenes y su tejido interno, y servir como una referencia posterior de las teorías utilizadas, comparadas y aplicadas en el modelo.

2.1 Bases de la psicología de la percepción visual y de la psicología del arte

El estudio de la psicología de la percepción visual, y en especial del arte, tuvo un avance significativo con el surgimiento de la teoría de la Gestalt a mediados del siglo XX por (Kofka, 1955). Inicialmente, muchos estudios se centraron en las imágenes del arte vi-

sual, pero donde tuvo una mayor aplicación posteriormente fue en la publicidad y la comunicación visual. Aunque al principio sólo se analizaban los efectos visuales que percibían los sujetos durante la percepción a través de experimentos, con los avances en neurociencia a finales del siglo XX, se ha comenzado a investigar en los procesos neuronales del cerebro como en el trabajo de Pina Baingio (Baingio, 2013).

La importancia del estudio y definición de las imágenes ha estado siempre presente, pero es la psicología del arte la que más ha evolucionado en el análisis de cómo se crean a partir de la composición. En la evolución de las últimas décadas, la teoría de la atención o el estudio de los movimientos de los ojos han jugado un papel relevante junto a la teoría del color y la sintaxis de la imagen.

En esta sección, se resumen las principales teorías y estudios sobre las imágenes y la percepción visual desde la psicología del arte para este proyecto. En primer lugar, se analiza la imagen como objeto visual, tanto como concepto (definición de lo qué es una imagen en la actualidad) como realidad formal (lo que implica una imagen como un objeto visual perceptible). En segundo lugar, se introduce la teoría de la Gestalt y su influencia en la psicología del arte. En tercer lugar, se define y analiza qué es la composición en las imágenes desde el punto de vista de la psicología del arte. En cuarto lugar, se presenta la sintaxis visual y su importancia en el análisis y creación de imágenes. En quinto lugar, se analiza la estructura subyacente de la composición conocida como «tejido interno» y su relación con la psicología del arte. En sexto lugar, se describe la teoría del color, tanto en la evolución de su estudio y los diferentes enfoques, como en la descripción de sus principales características e implicaciones en la imagen. En séptimo lugar, se describe la teoría de la atención sobre todo en relación con la percepción visual. Y, por último, se presentan los principales estudios y descubrimientos del movimiento de ojos en relación con la atención y la percepción visual.

2.1.1 La Imagen como objeto visual

Si reflexionamos sobre la palabra imagen, a la cabeza nos vienen objetos como pinturas, fotografías, dibujos, pero a la vez —y de ahí la complejidad de establecer una definición concreta— la representación que tenemos en la mente no sólo de estos objetos, sino de escenas y situaciones que, si nuestra capacidad expresiva nos lo permitiera, podríamos convertir a su vez en pinturas o dibujos. Si leemos un libro, en nuestra mente se generan imágenes, si pensamos en alguien, por ejemplo, un actor famoso, también construimos una imagen visual. En todos estos casos, hay un nexo común principal: la «realidad visual» donde la imagen habita, bien como un objeto físico o bien como una representación mental.

Miramos a la pared y allí hay una pintura (ver Figura 6), está en un espacio físico y delante de nuestros ojos, a unos metros, y no dudamos de su existencia. Además, la imagen se encuentra en nuestro cerebro distribuida en circuitos neuronales. Existe un objeto físico, sí, de hecho, está colgado en la pared, pero también, si cerramos los ojos, se encuentra ahí, en algún lugar de nuestro cerebro. En ambos casos, ¿dudamos de la existencia real de alguna de las dos?, no, la imagen pervive por igual en la pared y en

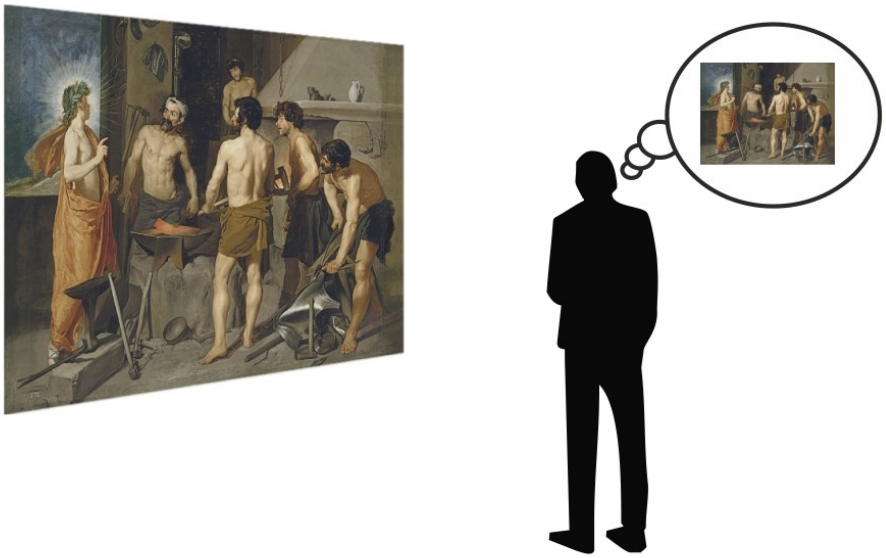


Figura 6 Ejemplo de una imagen como objeto físico y su representación mental.

nuestro cerebro. La realidad visual que la caracteriza se encuentra en ambos casos, es decir, que las características visuales que definen a la imagen están tanto en la pintura como en el cerebro del espectador, lo cual nos permite concluir que esta realidad no está definida sólo por aspectos físicos (la pintura al óleo sobre el lienzo), sino también por operaciones neuronales que se realizan en nuestro cerebro.

Por otro lado, una imagen creada por una persona, por ejemplo, una pintura, es el resultado de un proceso creativo que se genera en su cerebro. Pero esto no sucede sólo durante la creación a partir de una actividad manual como con una pintura, sino también sucede cuando cogemos una cámara de fotos, seleccionamos un punto de vista, situamos una altura y apretamos el botón.

Hablar sobre imágenes es un acto en sí mismo contranatural, ya que el mero hecho de verbalizarlas implica trabajar con sustitutos, en este caso palabras que intentan acercarnos a su realidad de naturaleza visual. Este problema es en definitiva el principal, el que se plantea cuando se intenta establecer definiciones sobre lo que es una imagen, lo que no es una imagen, e incluso la no-imagen como plantea Elkins (Elkins, 2001), es decir, lo que nunca podría ser una imagen en contraposición a lo que no es una imagen, que en algún momento sí podría llegar a serlo. Por ejemplo, imaginemos una palabra, no es una imagen, pero si se convierte en un logotipo y lo ponemos en un cartel, se convierte en una imagen. La palabra no es una imagen, pero se puede convertir en una imagen. Por otro lado, un olor, no es una imagen y no se puede convertir en una imagen.

Para Platón, las imágenes son representaciones de las ideas, en el sentido de reflejo físico de las mismas. Las sombras proyectadas en las paredes de la caverna son las imágenes de las ideas, las cuales permanecen en esencia, pero a su vez incompletas por el mero hecho de «ser representaciones de». Para Aristóteles, son accidentes, elementos

físicos imperfectos y temporales —esta noción es la más extendida en nuestra cultura. En muchas religiones, las imágenes tienen un papel relevante ya que actúan no sólo como «representaciones de», sino que adquieren entidad propia al convertirse en objetos de culto por sí mismas. Desde las distintas ramas que han estudiado las imágenes –la estética principalmente–, el debate se ha establecido en su funcionalidad. Es decir, se ha determinado qué tipo de imágenes son objeto de interés o de estudio según su finalidad. Por otro lado, en un segundo plano, el interés se ha centrado en la fase de creación. En el caso de la estética, el centro de interés son las «imágenes bellas» y, en general, su capacidad de seducción y el efecto que produce en el espectador su contemplación. En las «Lecciones de Estética» de Hegel (Hegel, 1845) se diferencian dos tipos de objetos de contemplación y por lo tanto de belleza: la naturaleza y las obras de arte. Para Hegel, la intencionalidad del hombre a la hora de crear imágenes bellas debe ser el objeto del estudio de la estética.

La evolución del concepto de qué es una imagen, o mejor dicho de qué queremos que sea una imagen, ha sido un viaje que parte desde la dependencia al contenido temático hasta la libertad o independencia como objeto (visual). Esto ha sido más evidente a partir del arte del siglo XX, por ejemplo, con la pintura abstracta. La experiencia nos ha demostrado que las imágenes creadas en el pasado no sólo han superado al tiempo en que fueron creadas, sino también a sus autores o las causas de su creación. La imagen como viajera en el tiempo, desempeñando roles, ha sido también superada por la idea general de que la imagen es un objeto en sí mismo, con sus propias limitaciones. Esto es analizado por Bredekamp (Bredekamp, 2004), Boehm (Boehm, 1994) o Mitchell (Mitchell, 1994). De la imagen que depende del espectador a la imagen que existe por sí misma, con independencia del entorno cultural en que fue creada. De esta manera, se establece una evolución que parte desde el contenido, realmente exento a la propia imagen, hasta la naturaleza bidimensional y material de la propia imagen y su semántica. Al fin al cabo una imagen es un objeto físico compuesto con elementos plásticos (en la pintura, las pinceladas o en el dibujo los trazos) y dispuestos según un orden concreto. Dentro de este proceso el establecimiento como objeto visual es un sólo paso más. Por otro lado, el carácter invariable de la imagen, desde el punto de vista físico, es o ha sido otro de los rasgos que ha marcado su configuración como objeto. Las diversas corrientes de estudio de las artes siempre concebían una parte física inmutable en la imagen, es decir que nunca cambiaba, y otra parte mutable, relacionada con lo externo a la propia imagen y que tiene que ver con su contenido. En los niveles planteados por Panofsky en la iconología (Panofsky, 1962), lo primero entraría a formar parte del nivel preiconográfico, mientras que lo segundo del iconográfico. Un cuadro del siglo XVI llega hasta el siglo XXI físicamente casi igual —no tenemos en cuenta el problema del paso del tiempo y el de las restauraciones—, mientras que su temática, lo que representaba o su simbología (aun siendo el mismo objeto físico), implica distintas interpretaciones en cada época. La Figura 7 muestra el cuadro de La Giaconda y sus cambios funcionales en el tiempo como objeto, desde su origen, un retrato, hasta la actualidad, como símbolo, pasando por un icono o pieza de museo. El objeto físico es siempre el mismo, como se observa en la figura, pero varía su interpretación y funcionalidad para cada generación.

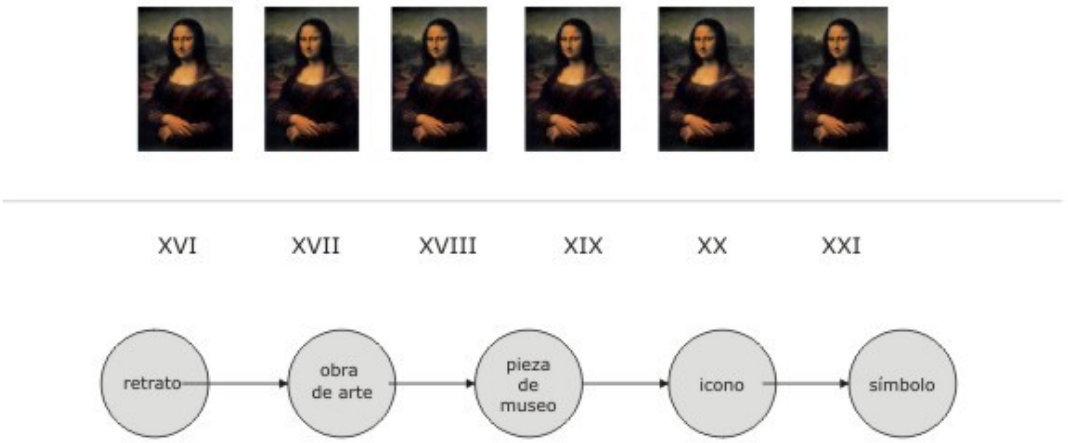


Figura 7 Los cambios de la funcionalidad de una imagen en la historia.

La definición de la imagen como objeto visual se plantea desde los estudios visuales como una nueva manera de comprender su realidad, como vemos en el trabajo de Brea (Brea, 2006), Elkins (Elkins, 2001) y Mitchell (Mitchell, 2002), por un lado, inmutable en cuanto a sus aspectos formales, y, por el otro, variable en su semántica. La idea inicial de que la imagen tiene diversos roles es parte de la necesidad de entender esa variabilidad semántica e intentar establecer una descripción más completa. No se interpreta igual una imagen en una época que en otra, y en muchos casos se debe al hecho de que la imagen original tuvo una motivación muy fuerte hacia un tema muy concreto, que con el paso del tiempo se ha desligado por los cambios culturales o sociales. Realmente se puede decir que lo que cambia es su escenario cultural, más que su contenido semántico ligado a lo formal, ya que éste sigue siendo el mismo. La Gioconda sigue siendo el mismo retrato del siglo XVI, pero su interpretación varía tanto como cambia la cultura, sin que esto provoque ninguna modificación en la imagen, ya que el cambio se produce en su contemplación e interpretación.

2.1.2 La Gestalt, la pregnancia de la buena forma

A mediados del siglo XX, la psicología de la percepción visual tuvo un avance importante con la aparición de la Gestalt, introducida por Kofka (Kofka, 1955). La Gestalt, a partir de ensayos y experimentos, descubrió distintas leyes y principios psicológicos que permitían entrever cómo funcionaba el sistema visual en los seres humanos, aunque sin poder explicar su funcionamiento interno y sin llegar a generalizarlo. Para la Gestalt, el cerebro era una caja negra de la que sólo se puede conocer el resultado de los procesos y no cómo se producen.

El principio de la Gestalt sobre el que gira su teoría es «la pregnancia de la buena forma», el cual indica que en la percepción visual existe una atracción por las formas equilibradas, regulares, simples, etc. A partir de este principio universal, surgen una serie de leyes y principios más concretos que la Gestalt demostró con experimentos reprodu-

cibles fácilmente. Sin embargo, los mecanismos internos del cerebro escapaban del objeto de la Gestalt, con lo que, a pesar de tener un gran desarrollo en el campo del arte y la estética a mediados de siglo XX, fue pronto refutada por no explicar científicamente los procesos. Su influencia en el campo del diseño y del arte, tanto en su desempeño como en las fases de aprendizaje, fue importante desde el principio, y podemos encontrarla en la base de la psicología del arte y la estética desde mediados del siglo XX, desde Gombrich hasta Arnheim, y en el desarrollo de la comunicación visual, como indican Rock y Palmer en su estudio sobre el legado de la Gestalt (Rock & Palmer, 1990). En la actualidad, gracias a los avances de la neurociencia y la psicología cognitiva, se comienzan a relacionar procesos cerebrales con los principios y las leyes. En este sentido existen nuevos enfoques, como el de Pina Baingio (Baingio, 2013), o aplicaciones en la inteligencia artificial (Calì, 2013).

La Gestalt no tenía como objeto explicar cómo el cerebro desarrolla el proceso de la percepción visual, ya que para ella es una caja negra. Sin embargo, ante un estímulo dado, analizaba la respuesta para generalizar por inferencia las leyes. Los principios y leyes fundamentales de la Gestalt son:

- La globalidad o totalidad. El todo es más que la suma de las partes, es decir, el todo no es divisible sin perder sentido.
- Figura y fondo. La parte de mayor pregnancia actúa como figura y el resto como fondo, en caso de igualdad de condiciones, se elige la forma más simple, regular, simétrica, etc. como figura.
- Principio de la buena forma o contraste. Se percibe mejor una forma si se diferencia del fondo.
- Cierre. Cuanto más cerrado está el contorno mejor se percibe una forma.
- Complementación. Si no está el contorno cerrado, se complementa.
- Noción de pregnancia. Es la fuerza que tiene la forma, y cuanto más esté presente su singularidad, mejor se percibirá.
- Simplicidad. Los elementos menos complejos tienen mayor pregnancia.
- Equilibrio. Los elementos se organizan según un centro de equilibrio.
- Continuidad. Los elementos que siguen un patrón pueden desarrollar una forma, de la cual forma parte.
- Principio de proximidad. Se agrupan los elementos cercanos.
- Principio de similitud. Se agrupan los elementos parecidos formalmente.
- Principio de memoria. Cuanto más se parecen las formas a percepciones anteriores, estas se distinguen mejor.

Abraham Moles (Moles, 1971) (Moles, 1973) incluyó otros principios más universales que los de la Gestalt, que denominó como «leyes de infralogía visual» por estar en una capa inferior a las de la Gestalt. Estas leyes se consideran una extensión de las de la Gestalt. Las principales son:

- Ley de centralidad. Los elementos que están en el centro de una figura son más importantes.
- Ley de infinidad. Si existen al menos tres elementos idénticos se entienden como parte de un conjunto limitado, si son superiores a siete de un conjunto infinito.
- Ley de percepción de la complejidad. Emerge cuando el número de elementos distintos que aparecen en la escena es superior a siete.
- Ley del dominio del ángulo recto. Los elementos con ángulos rectos parecen más elaborados que los que están formados por otro tipo de ángulos.

2.1.3 La composición

Dentro del campo del arte, y de una manera general en el de cualquier actividad creativa como el diseño gráfico, la composición es la herramienta esencial para la creación de imágenes donde todas las regiones forman parte de una única estructura. Esto implica que, si en una imagen aparece un círculo, un cuadrado y un triángulo, cada uno de ellos por separado tiene significado por sí mismo, pero sólo los tres, como un conjunto, dan sentido a esa imagen.

Por consiguiente, la composición es el resultado de situar y relacionar elementos para obtener un todo que, como conjunto, funciona con significado propio, pero dividido o seccionado, no. La estructura de la composición de una imagen se establece a partir de los elementos visuales, como la línea, la forma, el color o la textura de cada región, y a partir de las relaciones entre ellos. En este sentido, la composición se ha relacionado con la creación de estrategias a modo de receta por parte de los creadores, que han sido utilizadas en la historia de las artes visuales, y en muchos casos se han llegado a generalizar como cánones de belleza —entendidas como buena composición. Aunque una buena composición podría considerarse «fea» por un grupo concreto de personas sin que necesariamente estuviera mal compuesta.

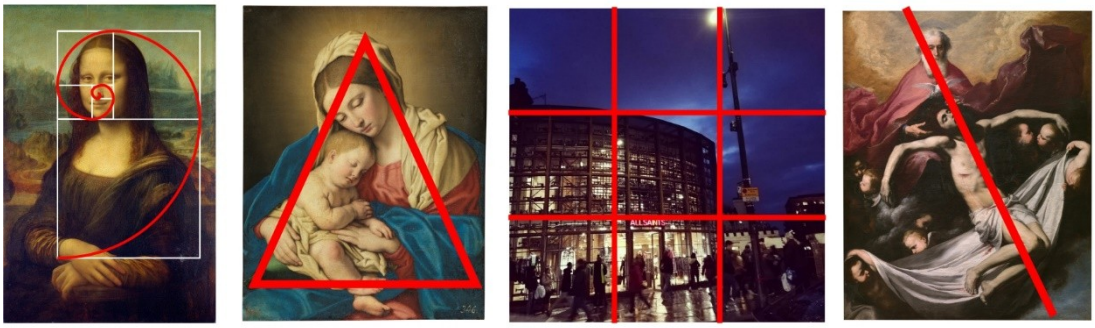


Figura 8 Ejemplo de tipos de composición.
Nota: (a) sección áurea («La Giaconda» de Leonardo dan Vinci; (b) triangular («La Virgen con el Niño dormido» de Sassoferrato); (c) regla de los tercios (Mercado All Saints Camden Town, Londrés 2019, foto del autor); y (d) eje diagonal («La Trinidad» de José de Ribera).

Las estrategias y métodos más comunes han sido los que relacionan la posición de los elementos en el espacio de la imagen dependiendo de un criterio geométrico. El uso de la sección de oro, a partir de la sucesión de Finobacci, se ha aplicado tanto para la relación de ancho y alto, como para la organización de los elementos formales de la imagen (ver Figura 8.a). La aplicación de estructuras basadas en formas como triángulos, cuadrados, rectángulos o círculos se han asociado a temas específicos, por ejemplo, la representación de la Virgen con el niño suele usar un triángulo (ver Figura 8.b) o la Última Cena un rectángulo, como ejemplo de dos temas muy comunes en la historia del arte. Además, existen estrategias basadas en las relaciones espaciales y la percepción visual, como la regla de los tercios, que establece una estructura fija en la cual si se sitúan los principales elementos de la composición para conseguir un equilibrio y estabilidad (ver Figura 8.c). Obviamente, existen estrategias para lo contrario, como el uso de diagonales (ver Figura 8.d).

2.1.4 Sintaxis de la imagen

En el último tercio del siglo XX, la experta en diseño gráfico Donis Dondis, desarrolló una sintaxis de la imagen para crear una gramática visual (Dondis, 1974). Para Dondis, aunque muchas de las leyes y principios que se aplican en la interpretación visual de una imagen son innatas, es necesaria una «alfabetización visual» tanto para la correcta creación de imágenes como para su comprensión. Además de los elementos visuales o de las leyes y principios de la composición basados en las teorías de la Gestalt, la sintaxis de la imagen se caracteriza por el análisis del peso visual y los ejes de equilibrio. El peso visual se asocia a la tensión o capacidad de atracción que tiene en el espectador una región concreta de la imagen dependiendo de los elementos visuales, pero también de la posición que ocupan en la imagen. En la imagen de la Figura 10, los ejes de equilibrio parten de la fragua (Figura 10.c), mientras que las fuerzas visuales de cada región, construidas a partir de los elementos visuales (Figura 10.b), se dirigen hacia la figura humana de la izquierda (Figura 10.d).

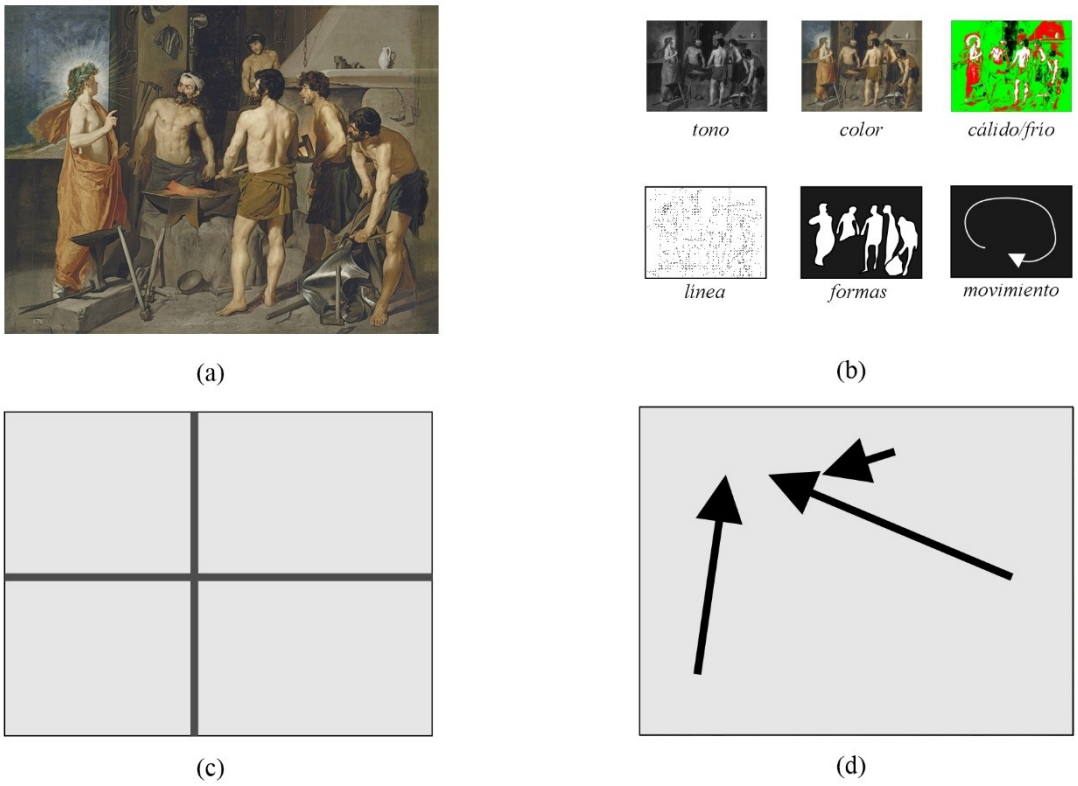


Figura 10 Representación de una composición aplicando la sintaxis visual.
Nota: (a) imagen; (b) elementos visuales, (c) ejes de equilibrio; y (d) fuerzas visuales.

Imaginemos un círculo de color rojo (Figura 9.a) y un cuadrado de color azul (Figura 9.b). En una composición donde el círculo rojo se encuentra embebido dentro del cuadrado azul, el primero parece potenciarse ya que el contraste entre el azul, con tendencia a contraerse, y el rojo, con tendencia a expandirse, favorece al círculo rojo (Figura 9.c). Sin embargo, si intercambiamos los colores, se produce un equilibrio en el contraste que hace que se estabilice la predominancia de uno sobre el otro (Figura 9.d). Esta relación entre elementos visuales puede variar el peso visual de la combinación, aumentando en el caso del círculo rojo embebido en el cuadro azul y disminuyendo en el círculo azul embebido en el cuadro rojo.

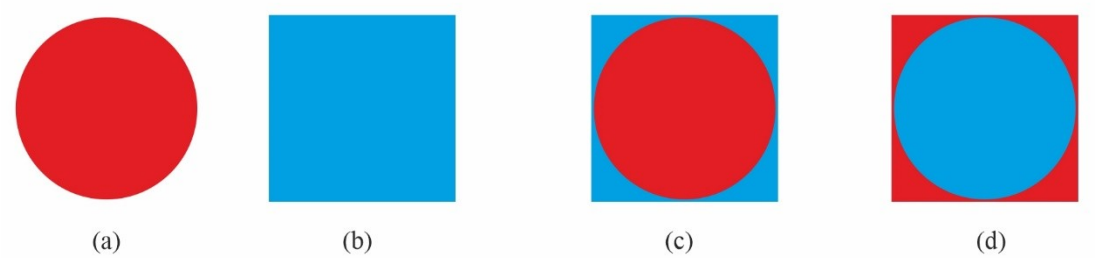


Figura 9 Combinación de elementos visuales y su influencia en la composición.
Nota: (a) círculo rojo; (b) cuadrado azul; (c) círculo rojo sobe cuadrado azul; y (d) círculo azul sobre cuadrado rojo.

Cada región de una imagen tiene dos tipos de tensiones: una local, provocada por los elementos visuales, y otra global por la posición que ocupa en la imagen y por la relación que se establece con el resto de las regiones. Desde la psicología del arte y desde la de la percepción visual, se han estudiado este tipo de tensiones, las cuales se regulan con un proceso que Dondis denomina como «agudización y nivelación». En este proceso, una región es agudizada por su peso visual (tensión local) y nivelada a través del equilibrio con el resto de las regiones (tensión global). De hecho, esa relación de tensiones no sólo depende de los elementos visuales y de la posición de cada región en la imagen, sino que el propio espacio de la imagen establece diferencias entre el centro y el exterior de la composición. Es lo que Arnheim define como los focos de atracción del marco estructural de la imagen, siendo este marco la estructura del espacio de la composición y que obliga a gestionar, además, un equilibrio «centro-exterior» (Arnheim, 1983).

Por consiguiente, la composición de la imagen como estructura es un todo indivisible con dos focos de atracción, uno central y otro externo, donde en cada región existe un peso visual local (que agudiza) y otro global (que nivela). En Figura 11, mostramos un ejemplo de Dondis utilizando una balanza de pesos, donde se consigue equilibrio si sus pesos son iguales (Figura 11.a), y el desequilibrio si no lo son (Figura 11.b). Para obtener un equilibrio en la situación segunda, se modifica el punto de equilibrado (Figura 11.c) o se incluyen más pesos para compensar (Figura 11.d).

El equilibrio como principio universal ha sido tratado desde la filosofía griega y se encuentra presente en muchas de las esferas del saber humano. Según la sintaxis visual de Dondis, el marco estructural de la imagen se establece a partir de dos ejes de equilibrio, uno vertical y otro horizontal, generando cuatro cuadrantes. Existe una preferencia

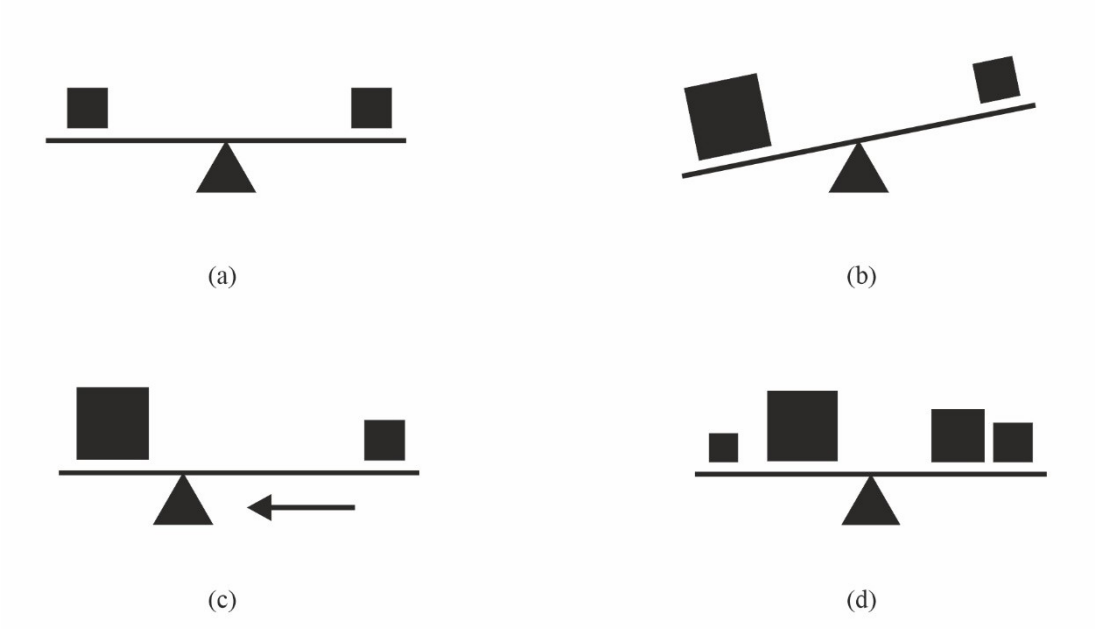


Figura 11 El proceso de equilibrado en una composición según la sintaxis visual.
Nota: (a) equilibrado simétrico; (b) desequilibrio entre los elementos; (c) equilibrado modificando el punto de equilibrio y (d) equilibrado compensando los elementos.

hacia el cuadrante inferior izquierdo, donde los elementos están en un mayor equilibrio, mientras que en la región opuesta (región superior derecha) sucede lo contrario y actúa como contrapeso. Esto genera un segundo sistema de fuerzas que actúa en conjunción con los ejes de equilibrio, en donde los elementos que estén más lejos del centro de equilibrio y en el cuadrante superior derecho serán los más difíciles de equilibrar, mientras que los que estén en la región inferior izquierda, cercana el centro del marco estructural, los que menos.

Los principios de la percepción visual y de la teoría de la Gestalt que aplica la sintaxis de la imagen son:

- **Equilibrio**. Es la finalidad de cualquier composición.
- **Peso visual**. Es la tensión que se produce en las regiones de la imagen tanto por su posición como por la relación de sus elementos visuales.
- **Agudización y nivelación**. Dos conceptos opuestos que tienen que ver con lo sorpresivo (agudización) y lo previsible (nivelación). En ausencia de equilibrio tendemos a la búsqueda de este o sea a la nivelación.
- **Preponderancia por la región inferior izquierda**. La región inferior izquierda es la región más equilibrada, de tal manera que los elementos que se sitúan en ella tienden a parecer equilibrados, en contraposición con la región superior derecha.
- **Atracción y agrupamiento**. Este principio se basa en la relación de elementos similares y su poder de atracción y agrupación.
- **Positivo y negativo.** Tiene que ver con la idea de figura y fondo, o entre elemento de interés y entorno. El primero actúa como positivo, mientras que el resto actúa como un entorno nebuloso, negativo.

Cada elemento formal de la composición se construye principalmente por los siguientes elementos visuales de la composición:

- El **punto** es la unidad mínima formal.
- La **línea** es, por definición, la unión de una serie de puntos estableciendo una continuidad. La línea tiene un propósito o intencionalidad. Los tipos de línea pueden ser: rectas, curvas o la mezcla de ambos, pudiendo ser regulares o irregulares y quebradas o lisas.
- El **contorno y forma** se establece como el resultado del cerramiento de una línea. Existen tres tipos de contornos básicos: círculo, cuadrado y triángulo. Cualquier contorno que encontremos derivará de estos tipos básicos.
- **La dirección.** Depende de la relación de puntos, líneas o contornos. Existen tres tipos: horizontal-vertical, diagonal y curvo. El primer caso establece con claridad un sentido de equilibrio en la escena; el segundo, más tensión que se

resuelve en la nivelación de los elementos; y el último plantea una continua tensión, sin solución, que deriva en un movimiento.

- **Tono.** Es un elemento que representa la intensidad de la luz. El tono se compone de dos extremos, máxima intensidad (blanco) y mínima intensidad (negro).
- **Color**. Es un elemento relacionado con el tono, la intensidad de la luz, pero con la combinación de longitudes de onda diferentes. Las principales propiedades que lo definen son el matiz, la luminosidad y la saturación.
- **Textura.** Es un concepto más bien táctil, pero visualmente es la distribución de los tonos y colores debido al aspecto de las superficies. La rugosidad produce la incidencia de la luz y la sombra con mucha discontinuidad, mientras que lo liso no genera sombras, sino que domina un solo tono.
- **Escala.** La relación que se establece entre los distintos elementos de la composición (puntos, líneas, contornos, formas, etc.). Este concepto es importante para entender el espacio y la lógica compositiva de los objetos y formas.
- **Dimensión.** Una imagen está compuesta por dos dimensiones (alto y ancho) en donde se suelen representan escenas que generan la ilusión de tres dimensiones (la profundidad es la tercera). La perspectiva es una de las principales herramientas para la creación de la ilusión de la tercera dimensión a partir de la línea de horizonte, puntos de fuga, proyección cónica, etc.
- **Movimiento**. Es el resultado de la combinación de los procesos de agudización y nivelación, donde se produce la ilusión de movimiento a través de la tensión de las regiones de la imagen y su equilibrado.

2.1.5 La estructura subyacente de la composición: el tejido interno

Con el inicio del psicoanálisis y el interés por el inconsciente a mediados del siglo XX, pronto surgió la idea de la existencia de una estructura subyacente en la composición. Anton Ehrenzweig, investigador del arte moderno y experto en psicología del arte, en su libro «The hidden order of the art» (Ehrenzweig, 1967) describe la existencia de una estructura subyacente en la composición denominada «tejido interno», la cual representa de una manera global las distintas tensiones de la composición sin la necesidad de analizar las características visuales. Ese tejido interno actúa como la estructura de un edificio, invisible tras los materiales y acabados, pero presente formalmente en la esencia. Ehrenzweig relaciona esta estructura interna con una percepción horizontal donde se capta lo global y es de tipo inconsciente. Por el contrario, la composición se relaciona con una percepción vertical de tipo consciente donde se captan los elementos aislados. Con la percepción horizontal, la composición se percibe como un todo donde no se reconocen formas, sino que se obtiene una especie de mapa estructural, mientras que, con la

vertical, se destacan unos elementos visuales sobre otros y se reconocen formas y patrones. Según Ehrenzweig, la percepción horizontal se caracteriza por:

- Ser polifónica, existen varias regiones a la vez como relevantes.
- Ser conjuntiva, selecciona todas las regiones a la vez.
- Ser desenfocada, tiene varios focos de interés.
- Ser intuitiva, no existe un análisis semántico.
- Ignora la «Gestalt», un todo donde cada parte tenga que estar equilibrada.
- No está presente la Ley del cierre que completa formas o patrones incompletos.

Mientras que la percepción vertical se caracteriza por ser:

- Ser monofónica, se centra en una región como figura y el resto como fondo.
- Ser disyuntiva, selecciona una región sobre otras.
- Se interesa por los detalles de cada una de las formas o patrones reconocidos.
- Ser enfocada, se concentra en una sola región.
- Ser intelectual, existe un análisis semántico.
- Busca la «Gestalt», un todo donde todas sus partes están equilibradas.
- Está presente la Ley del cierre que completa formas o patrones incompletos.

La percepción horizontal plantea una complejidad importante para un ser humano que intenta analizar el tejido interno de la composición, ya que la percepción visual es un proceso principalmente consciente. Ehrenzweig ahonda en la relación entre esta estructura interna y los «pulsos inconscientes del artista», donde, según el autor, el creador, libre de los elementos visuales (líneas, formas, colores, texturas, etc.), establece relaciones inconscientes que facilitan la creatividad. Además, indica que la genialidad de algunos artistas se encuentra precisamente en la capacidad de aplicar una percepción horizontal en el proceso creativo. En la Figura 12.a, vemos el resultado de lo que sería una percepción horizontal, donde todas las regiones tienen el mismo tratamiento, y en la Figura 12.b, vemos un ejemplo de lo que sería el resultado de la percepción vertical, donde se muestran nítidos tres focos mientras que el resto aparece difuminado.

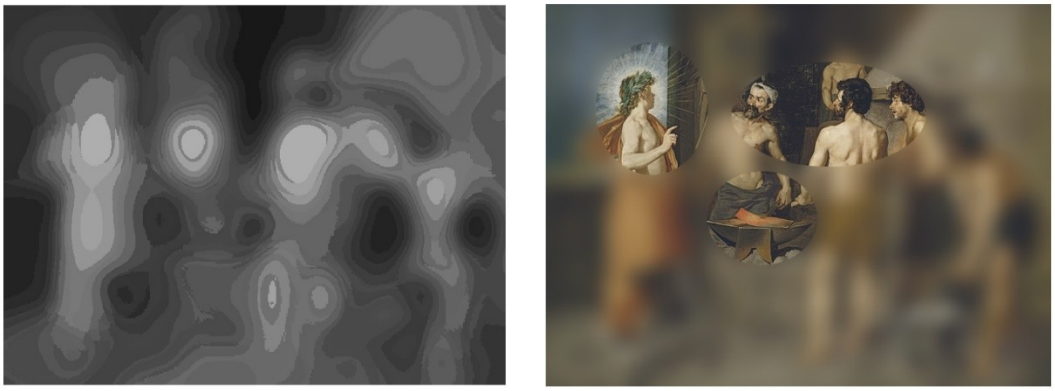


Figura 12 Ejemplos de la percepción horizontal y vertical.
Nota: (a) percepción horizontal, las regiones de la imagen no tienen detalles o formas, sólo niveles de tensión; y (b) percepción vertical, existen varios focos de atención en regiones concretas con detalle, mientras que el resto están difuminadas.

Después de Ehrenzweig, hay estudios teóricos, como (Milner, 1987), pero no existen avances en el desarrollo de técnicas o métodos para determinar este tipo de estructura, debido, principalmente a la complejidad que plantea la percepción horizontal.

2.1.6 La teoría del color

El color ha estado siempre presente en la curiosidad humana. Ya Platón o Aristóteles estudiaron sus propiedades y elaboraron conceptos como el flujo visual en el sistema de percepción o que todos los colores partían de la mezcla de cuatro, respectivamente. Sin embargo, las teorías modernas comenzaron con Newton y su *Óptica* con la descripción del espectro del color (Newton, 1730). Posteriormente, por un lado, Thomas Young estableció las bases de la teoría tricromática (Young. , 1801), donde la retina capta tres rangos diferentes de longitud de onda, y por el otro, Von Helmholtz (Helmholtz, 1852), al final del siglo XIX, definió la teoría completa del sistema de color tricromático (sistema RGB). Munsell desarrolló un sistema (Munsell, 1915) basado en la esfera de color de Runge (Runge, 2010), por el cual, cada color se identificaba a través de sus propiedades de matiz, luminosidad y saturación (sistema HLS). El sistema substractivo CMYK surgiría de las teorías de color desarrolladas en este contexto por Itten (Itten, 1992) y Klee (Klee, 1961), entre otros. Por otro lado, al inicio del siglo XIX, Goethe presentó su teoría del color 1810 (Goethe, 1840), donde indica que la percepción visual desempeña un papel activo y el color emerge en el cerebro. Un joven Arthur Schopenhauer analizaría esta idea y decidiría elaborar una teoría más avanzada en 1816 (Schopenhauer, 1816) donde relaciona la actividad de la retina con los colores a partir de un sistema de procesos de colores opuestos usando una división cualitativa (por la relación de los colores opuestos) y cuantitativa (por la intensidad de la actividad). Hering (Hering, 1885) describió este proceso basándose en «Teoría del color» de Goethe e indirectamente en «sobre la Visión y el Color» de Schopenhauer. En esta teoría, cada color se relaciona con su pareja opuesta, por ejemplo, rojo con verde, o amarillo con azul.

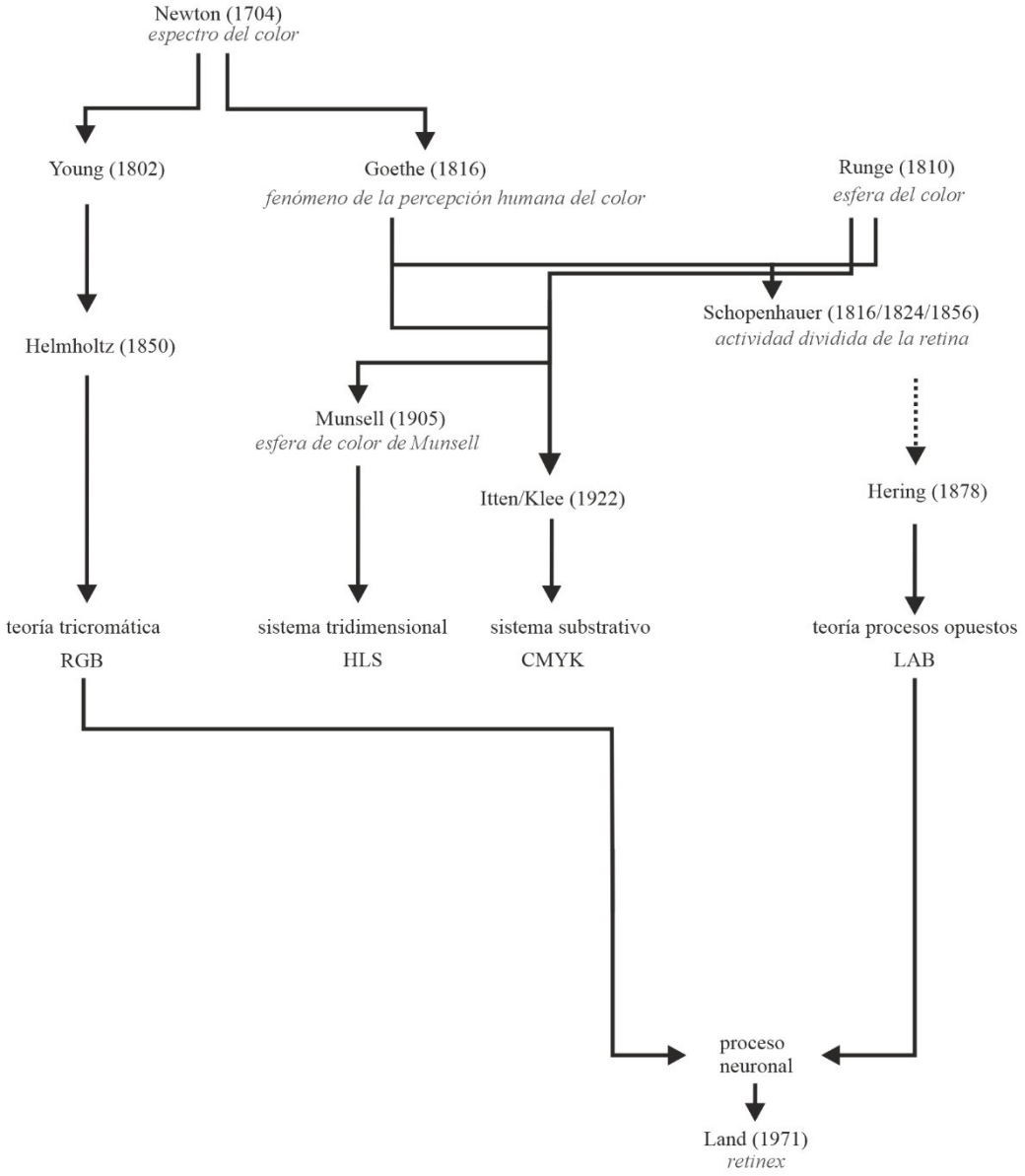


Figura 13 La evolución de las teorías del color.
Nota: Existen dos ramas: la de la teoría tricromática y la de la teoría de procesos de colores opuestos, las cuales convergen en la teoría Retinex.

Durante ciento cincuenta años, existieron dos teorías sobre el color: la tricromática y la de procesos de colores opuestos. Con el avance de la neurociencia, se pudo demostrar a mediados de los años cincuenta que, si bien el ojo capta la luz con un sistema tricromático, la señal es convertida a un sistema de procesos de colores opuestos a través de operaciones neuronales (Hubel, 1995). El color en la percepción es, por tanto, un proceso activo neuronal, tal y como propusiera Land con su teoría Retinex (Land, 1977) en relación con el problema de la constancia del color. La Figura 13 muestra en un mapa la relación de todas estas teorías y la evolución hasta llegar a una teoría general.

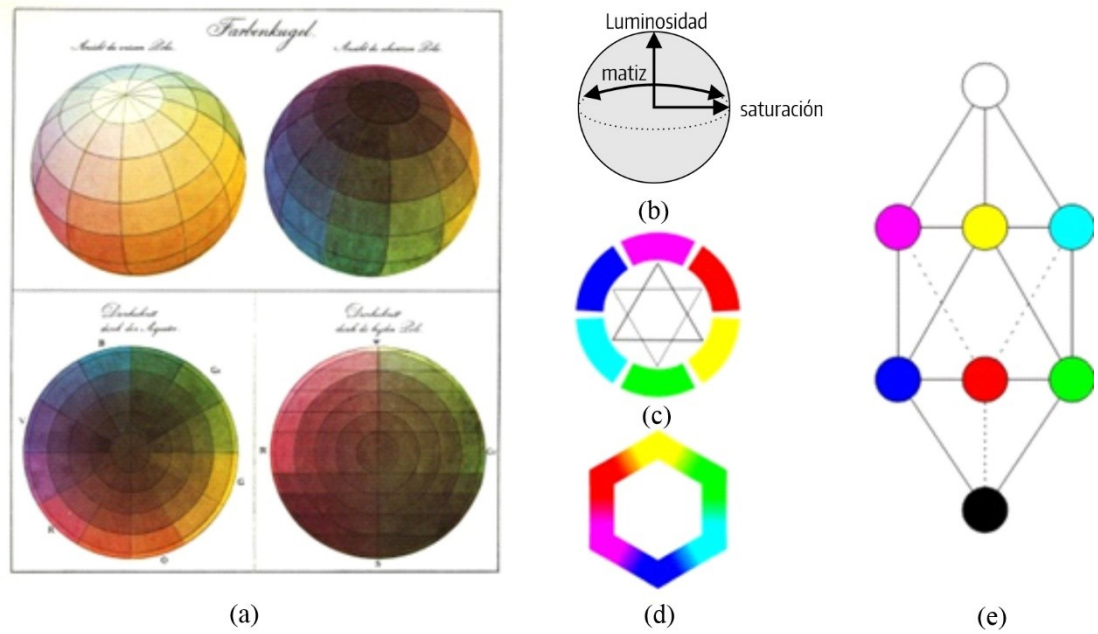


Figura 14 Ejemplos de tipos de representación del color.
Nota: (a) la esfera de color de Runge; (b) la esfera HSL; (c) rueda de color de complementarios de Goethe y Hölzel. En esta rueda, hay una tricromía de primarios con el rojo, verde y azul, y de secundarios con el amarillo, cian y magenta; (d) el hexágono de Kueppers; y (e) romboedro de Kueppers y su relación jerárquica.

La jerarquía entre los colores es necesaria para construir la composición y crear la estructura con criterios estéticos como el contraste y las relaciones de vecindad, oposición o complementariedad. Munsell determinó un sistema (Munsell, 1915) basado en la esfera de color de Runge (Runge, 2010) (ver Figura 14a), por el cual, cada color se posiciona en la espera a través de sus propiedades de matiz, luminosidad y saturación (base del sistema HLS) (ver Figura 14.b). Por otro lado, es muy común la creación de relaciones en forma de rueda de color como las de Itten y Hölzel, (Itten, 1992) (ver Figura 14.c), y estrellas elementales o pirámides dobles por Klee, (Klee, 1961). Kueppers estableció su hexágono (Kueppers, 1982), donde se relacionan los colores a través de sus vértices (ver Figura 14.d), y posteriormente con su romboedro creó una relación jerárquica entre colores, con el amarillo, el cian y el magenta en la parte superior, el verde, el rojo y el azul en la inferior y donde el blanco y el negro establecen los extremos (ver Figura 14.e).

Los conceptos de colores opuestos y complementarios se han usado indistintamente a lo largo de la historia, llegando a ser confusos, aunque existen estudios actuales que replantean las diferencias entre ambos como (Manzotti, 2017), (Zeki y otros, 2017) o (Pridmore, 2008). Tratar con colores complementarios es simple en el sistema tricromático, ya que los colores se complementan para obtener el blanco (valor máximo en los canales RGB), sin embargo, no es tan directo tratar con los colores opuestos.

Berlin and Kay (Berlin & Kay, 1991) demostraron, después de estudiar lenguajes alrededor del mundo, que los colores son universales e independientes a sensibilidades culturales y determinaron dieciséis categorías básicas. E. Rosch (Rosch, 1975) descubrió una tribu en Nueva Guinea (Los Danis) con solamente dos categorías: claro-cálido y

frío-oscuro. En experimentos posteriores con esta tribu, se comprobó que aprendieron el color rojo, verde, azul y amarillo (los colores primarios para los procesos de colores opuestos) mucho más rápido que los otros.

En psicología del arte, Arnheim (Arnheim, 1983), Dondis (Dondis, 1974) o Gombrich (Gombrich, 1959) analizaron composiciones artísticas usando leyes y principios de la Gestalt (Kofka, 1955) para explicar elementos visuales como el color e introdujeron conceptos como el peso visual o el equilibrio, los cuales relacionan el color con propiedades físicas.

2.1.7 La teoría de la atención

La teoría de la atención es un área de estudio de la psicología y la neurociencia que investiga sobre los procesos cognitivos que permiten al cerebro enfocarse en ciertos estímulos ignorando otros. No hay un consenso para una definición única, aunque existen enfoques que tratan de explicar sus causas y efectos. Las corrientes más relevantes son:

- La atención selectiva. Los primeros estudios sobre atención indicaban que el ser humano era un procesador de información con una sola vía y con una capacidad limitada de procesamiento, como indica Posner en su estudio sobre atención y rendimiento computacional (Posner, 1993). En esta línea, destaca el modelo de filtro atencional (Broadbent, 2013), el cual plantea la existencia de un filtro en los niveles cercanos a los estímulos que permite la selección de los más relevantes, así como el almacenamiento en la memoria a corto plazo para su posterior procesamiento.

- La atención dividida. En este modelo, influenciado por la perspectiva cognitiva en la psicología, los autores defienden la existencia de procesos automáticos y controlados (Bennett & Flach, 1992).

- La atención sostenida. En este enfoque, la atención en un estímulo se mantiene en el tiempo a pesar de la existencia de otros estímulos distractores (Davies & Parasuraman, 1982).

Desde el punto de vista del procesamiento y control, la teoría de la atención establece dos tipos:

- El procesamiento de arriba hacia abajo, controlado por áreas del cerebro y que permite seleccionar entre los estímulos de entrada los más relevantes.

- El procesamiento de abajo hacia arriba que permite que los estímulos relevantes de la entrada sean los que guíen la atención.

Lo habitual es que ambos convivan, y que entre ellos exista inhibición y competencia. Para la percepción visual, existe una relación entre la atención y el lugar donde se mira que depende de los dos tipos de procesamiento: bien porque alguna región del campo visual inicia el proceso de atención en el sujeto (abajo hacia arriba) o bien porque el

sujeto fija su mirada con intencionalidad en alguna región (arriba hacia abajo). Debido a esta relación entre mirada y atención, el estudio de los movimientos oculares ha tenido un importante desarrollo, como analiza Wade en su estudio (Wade, 2010).

2.1.8 Movimiento de ojos

El interés por el estudio del movimiento de los ojos es bastante antiguo, empezando por Aristóteles o Ibn Al-Haitham hasta Yarbus, y pasando por Purkinje, Mach, Crum Brown, Hering, Javal, Dodge, Huey, Stratton o Buswell. Nicholas J. Wade realiza un análisis en profundidad de estos estudios en «Pioneers of eye movement research» (Wade, 2010) donde indica que los principales intereses de los primeros investigadores del área se centraron en la anatomía, el movimiento binocular conjunto de ambos ojos y la torsión, además de la sincronización con los movimientos de la cabeza y el cuerpo (aspectos como la rotación y translación de las órbitas o la sincronización con la cabeza especialmente). Un aspecto relevante fue la determinación del «plano medial» y después del «plano transversal», o la simetría de ambos planos (izquierdo y derecho), relacionados con los dos mapas retinotópicos en que se divide el espacio visual. Émile Javal introdujo el término «saccades» referido a los movimientos rápidos que realizan los ojos durante el escaneo del espacio visual, saltando de una zona a otra, que se complementan con las denominadas «fijaciones», que son las paradas que realizan los ojos en zonas concretas (Javal, 1878).

En cuanto a las imágenes y su relación con el movimiento de ojos (visualización de fotografías o pinturas), se empezaron a estudiar en los trabajos de Stratton. Los trabajos (Stratton, 1902) descubrieron que los trazados que realizan los ojos en el escaneo de las imágenes, a diferencia de la lectura de texto donde existen movimientos organizados en una misma dirección, pasan de una zona a otra de una manera irregular, siendo interrumpidos por algunos instantes, como descansando en algunas regiones concretas. Además, realizó un importante avance cuando relacionó el movimiento de ojos con la percepción y los intereses cognitivos de alto nivel con procesos inconscientes de bajo nivel (Stratton, 1906). Buswell se centraría en la observación de sujetos experimentales mientras contemplaban imágenes para registrar la duración y posición de las fijaciones, tanto en imágenes simples como complejas (Busswell, 1935). Descubrió que existían «centros de interés», donde se concentraban las fijaciones, y que, además, estaban relacionados con la imagen. Posteriormente, contabilizó las medias en los tiempos de fijación entre personas formadas en arte y sin formar, entre niños o adultos, concluyendo que no existían diferencias concluyentes. En relación con el efecto que podría tener una composición concreta en cuanto al movimientos de ojos, determinó que era menor que del que se preveía.

Posteriormente, Yarbus demostró en un experimento en el que un sujeto observaba la misma imagen con diferentes tareas indicadas (Yarbus, 1967), que los trazados realizados por los ojos variaban, lo que le llevó a concluir que los procesos de arriba-abajo y de abajo-arriba intervenían en el control. Autores posteriores se centraron en las fijaciones, sobre todo en los movimientos involuntarios, y en la relación entre la postimagen y la dificultad de mantener la fijación en un punto concreto.

En la Figura 15, podemos comprobar fácilmente este problema realizando un experimento de percepción visual. En primer lugar, fijamos la mirada en el cuadro verde durante al menos 30 segundos. En segundo lugar, tapamos el cuadro verde con un papel blanco y fijamos la mirada en el cuadro del centro. A los pocos segundos, veremos como el cuadro se rellena de magenta. En tercer lugar, tapamos también el cuadro del centro y fijamos la mirada en el espacio derecha de la figura. El cuadro magenta empezará a moverse en el espacio blanco sin que lo podamos controlar (normalmente de abajo hacia arriba flotando). En el caso del cuadro que se rellena de magenta, un proceso de tipo arriba hacia abajo prevalece y facilita nuestra atención en el cuadro. En el segundo caso, el cuadro moviéndose, prevalece un proceso de abajo hacia arriba al estar presente un espacio vacío (el cuadro que vemos magenta en movimiento no es un elemento presente en el espacio visual sino una retroalimentación anterior). El reconocer un patrón (cuadrado del centro de la figura) hace prevalecer un proceso de arriba hacia abajo y que la fijación se centre en él, mientras que la ausencia de un patrón hace prevalecer un proceso de abajo hacia arriba y, por consiguiente, movimientos sacádicos por el espacio vacío.

En 2007, Henderson y colaboradores (Henderson y otros, 2007), realizaron un experimento con varios sujetos en tareas de búsqueda visual activa en imágenes, y pudieron comprobar que no existía una relación entre los movimientos sacádicos y las posiciones y las regiones de prominencia de las imágenes, que por estadística eran las de interés en la búsqueda. Concluyeron que, por un lado, los procesos de abajo hacia arriba se centraban en el escaneo del espacio visual a través de un proceso de «pasos que llevan a otros pasos» y por otro lado los procesos de arriba hacia abajo se centraban en la detección de patrones concretos. La interrupción de ambos procesos entre ellos provocaba el movimiento de ojos impredecible que los autores habían constatado.

El estudio de Wade concluía con los avances que aportan las nuevas tecnologías en el seguimiento de los movimientos de ojos. Aunque se mantienen muchas incógnitas sobre los movimiento sacádicos y las fijaciones, actualmente, los investigadores tienen

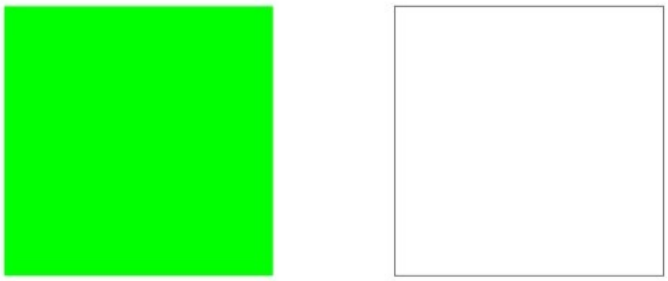


Figura 15 Ejemplo de postimagen.
Nota: mirar fijamente al cuadro verde de la izquierda durante unos treinta segundos, después taparlo y fijar la mirada en el cuadro blanco del centro (veremos con el cuadro pasa del blanco a un magenta. Volver a repetir la acción, pero ahora después de tapar el cuadro verde, fijar la mirada en el espacio vacío de la izquierda, tapando también el cuadro blanco. Veremos que, en este segundo caso, el cuadro magenta se moverá y lo seguiremos. En el primer caso, se desarrolla una tarea de tipo arriba hacia abajo y al mantener la mirada fija en el cuadro, mientras que, en el segundo caso, no es posible mantener la fijación, y el proceso de tipo abajo hacia arriba prevalece estableciéndose un nuevo movimiento.

evidencias de que los movimientos no son tan continuos como se pensaba y que están íntimamente relacionados con procesos cognitivos de la percepción.

2.2 Bases neurocientíficas

El estudio biológico de la visión ha estado tradicionalmente centrado tanto en la anatomía como en el procesamiento de la información. Por lo tanto, la neurociencia se ha enfocado en el análisis de las áreas cerebrales y en la experimentación de la percepción visual en sujetos. En las últimas décadas, el desarrollo de sistemas de visualización de la actividad cerebral mediante neuroimagen ha representado un avance significativo en este campo, ya que ha permitido localizar las áreas activas o inactivas del cerebro en respuesta a estímulos específicos.

La luz constituye el punto de partida de cualquier proceso de percepción visual, ya sea generada por una fuente natural o artificial. El objetivo principal de cualquier sistema de percepción visual es extraer la máxima cantidad de información posible de esta fuente inicial de datos, que proviene de la luz, y establecer las estrategias más adecuadas para lograr dicho objetivo. Durante mucho tiempo, se consideraba que la percepción era un proceso de captación de información sin un procesamiento cognitivo. Sin embargo, en la década de 1950, autores como Rudolf Arnheim plantearon que la percepción era un proceso activo de cognición y no una simple captación pasiva de datos para su posterior procesamiento (Arnheim, 1969). Este cambio de perspectiva fue crucial, ya que condujo al estudio de la percepción visual como un proceso completo, incluyendo la identificación de un área visual en la corteza donde se procesaba la información en interacción con otros órganos involucrados.

En esta sección, se resumen los principales descubrimientos y teorías sobre el funcionamiento del sistema de percepción visual que se han estudiado en el contexto de este proyecto. En primer lugar, se analiza la percepción visual desde diversas perspectivas. En segundo lugar, se describe la anatomía y fisiología del sistema de percepción visual. En tercer lugar, se examina la retina, tanto en términos de su estructura como de su funcionalidad. En cuarto lugar, se aborda el núcleo geniculado lateral (NGL) del tálamo y su papel como puerta de entrada y control de la corteza visual primaria. En quinto lugar, se exploran los mapas retinotópicos. Y, por último, se profundiza en el procesamiento de la percepción visual.

2.2.1 La percepción visual

Un rayo de luz es el dato que llega a la retina. La intensidad de la luz es la primera característica que debemos tener en cuenta y se ve determinada por la fuente de origen, pero también por la segunda, los cambios de longitud de onda debido a la reflexión o refracción con el medio físico. Ambas características de la luz, como dato visual, son la base para la evolución y creación del sistema de percepción visual en un ser vivo, ya que es éste el que se adapta a estas circunstancias y no al revés. De esta manera, la retina se especializa en detectar ciertos rangos de longitud de onda de la luz y también en proce-

sarlos. En el caso del sistema visual de muchos de los primates, estas vías son procesadas principalmente por tres tipos de células ganglionares: magnocélulas, parvocélulas y koniocélulas, las cuales se especializan en la detección de un rango de longitud de onda concreto: largo, medio o corto respectivamente, tal y como los neurocientíficos lo han descrito y Davila Teller recopiló en tu trabajo póstumo sobre la visión (Teller, 2014).

La realidad física de la luz y sus propiedades determinan la anatomía y funcionalidad del sistema de percepción visual y, por consiguiente, una mayor cantidad de objetivos y dificultad para conseguirlos implica una mayor complejidad. En este sentido, la realidad visual que un sistema de percepción capta depende de esta adecuación anatómica y funcional. Por ejemplo, en el sistema de percepción visual de la rana, un objetivo es detectar a un insecto con el fin de cazarlo o a un depredador para huir de él. En el trabajo de Lettwin y colaboradores. «*What the Frog's Eye Tells the Frog's Brain*» (Lettvin y otros, 1959) se investigó la retina de la rana concluyendo que su estructura se compone de células ganglionares que son capaces de discriminar los cambios de intensidad que se producen cuando los objetos se interponen entre la fuente de luz y la rana en dos intervalos de tiempo consecutivos. Dependiendo del tamaño, el objeto podría ser un mosquito (una presa) o un ave (depredador). La velocidad de reacción es importante, por lo que las células ganglionares de la retina tienen la capacidad cognitiva de discriminar ese cambio de intensidad y, al estar conectadas directamente con el sistema motor, estimular la lengua para cazar o las patas para saltar y huir. Por consiguiente, «la realidad visual» de la rana se relaciona con la captación de los cambios de intensidad en los rayos de luz por objetos que se interponen cuando estos pasan de cierto umbral. Si el cielo es azul, las copas de los árboles están llenas de hojas o los pájaros vuelan alto, no forman parte de la realidad visual de la rana salvo que se produzca un cambio de intensidad puntual al interponerse un objeto y, según el tamaño, la acción a realizar sea bien cazar o bien huir.

2.2.2 Anatomía y fisiología del sistema de percepción visual

La neurociencia ha tenido un importante avance en los últimos treinta años, tanto en el conocimiento funcional como en la estructura y anatomía de las áreas del cerebro. Uno de los aspectos más relevantes descubiertos es el de la especialización de cada área en una funcionalidad concreta, lo que implica que sus neuronas y su estructura dependan de la tarea a realizar, es decir, que la forma y función estén relacionados tal y como indica Semir Zeki en su trabajo «The visual association cortex» de 1993 (Zeki, 1993). En este trabajo, en relación con la percepción visual en tareas de reconocimiento de formas, se constata que, además de esta especialización, la parte frontal, asociada al control y la planificación, y que se asumía como coordinadora de las tareas de visión, permanece inactiva. Este hecho deja claro que no existe un proceso muy coordinado en la percepción visual, ni siquiera en tareas que determinábamos como conscientes. A su vez y por otro lado, Zeki indica la existencia de conexiones desde las áreas de procesamiento en los niveles altos, tanto de planificación como del área frontal o de la memoria, con las áreas de procesamiento en niveles bajos en la visión temprana V1 y V2. Con estos avances, la visión se empezó a estudiar como un conjunto de procesos

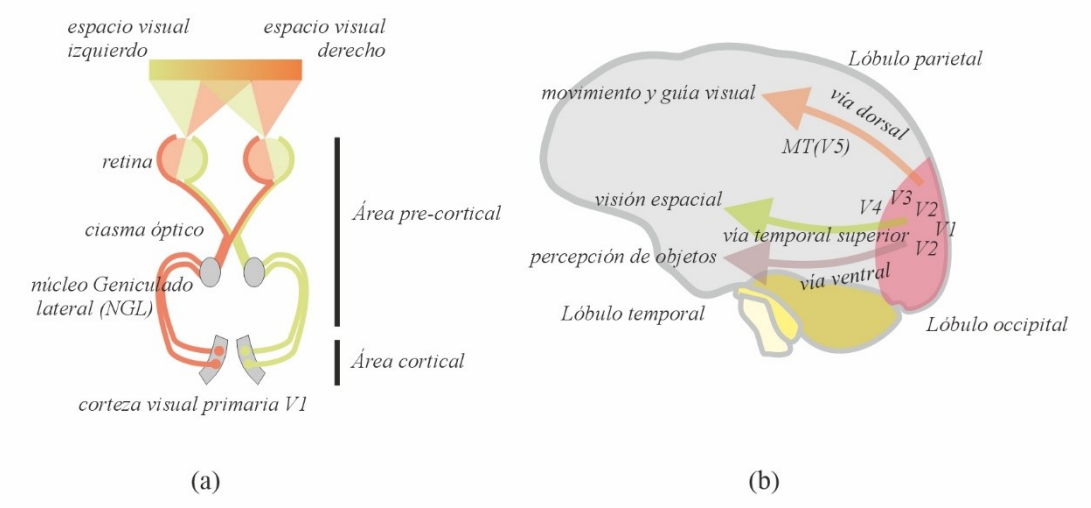


Figura 16 Vías de procesamiento de la información en el sistema de visión.
Nota: (a) área precortical; y (b) corteza visual primaria, principales áreas y vías de procesamiento.

interrelacionados que implican a varios órganos y áreas del cerebro, y en donde la información es procesada por áreas especializadas.

En la Figura 16, mostramos un esquema anatómico del sistema visual humano describiendo los principales flujos de procesamiento. Existen dos áreas básicas: la precortical, retina y NGL (ver Figura 16.a), y la corteza visual (ver Figura 16.b), donde la información se distribuye según la región del espacio visual de la cual proviene —la región de la izquierda es procesada en el hemisferio derecho del cerebro y a la inversa con la región de la derecha. La información es trasladada en un proceso de abajo hacia arriba a la corteza visual desde la retina, pasando por el NGL hasta el área de la corteza visual primaria, V1 (Hubel D. H., 1995). Aquí existe una relación con las áreas visuales V2, V3 y V4, pero también con el V5 y otras áreas, todas especializadas en algún proceso muy concreto: detección de bordes, asociación de patrones y formas, determinación de la profundidad, color, texturas, movimiento, reconocimiento facial, etc. Aunque entre ellas hay retroalimentaciones, en sí, actúan como módulos independientes en relación con sus objetivos, los cuales se agrupan a través de las tres vías de procesamiento como se muestra en el gráfico de la Figura 16.b:

- Dorsal (en relación con el movimiento).
- Temporal superior (visión espacial).
- Ventral (percepción de objetos).

Por otro lado, la especialización funcional también sucede entre las distintas láminas que compone la corteza visual. El modelo *LAMINART,* desarrollado a principios del siglo XXI, plantea una computación que transcurre horizontalmente a la vez que de abajo hacia arriba y de arriba hacia abajo, recibiendo señales de otras láminas, e incluso desde otros módulos, como indica Grossberg (Grossberg, 2003). Dentro de FACADE (Grossberg, 1990) —modelo computacional creado para ampliar los conocimientos

neuronales del funcionamiento de la visión temprana originalmente con un planteamiento no laminar—, el establecimiento del modelo LAMINART resuelve muchas de las cuestiones de cómo computacionalmente se detectan bordes, se establece las distancias en el espacio visual, se determinan formas, etc.

Desde los años 70, existe una variante de la neurociencia denominada neuroestética que estudia la relación entre la anatomía del cerebro y su funcionalidad para cuestiones estéticas. Esta variante trabaja hoy en día desde un enfoque multidisciplinar, intentando correlacionar las respuestas a nivel neuronal con percepciones estéticas. En «Aesthetics and psychobiology» de Berlyne (Berlyne, 1973) se planteó la creación de una ciencia de la estética en este sentido, pero no será hasta los 90, con el avance de la neuroimagen, cuando emergió la neuroestética, donde Semir Zeki es una de las principales figuras.

2.2.3 Retina

Davila Teller, indica que la principal función de la retina es convertir la señal tricromática captada de la señal lumínica a un sistema de procesos de colores opuestos (Teller, 2014). En su análisis anatómico, describe que el ojo humano tiene dos tipos de células para captar la luz: bastones y conos. Los primeros son sensibles a un segmento de longitud de onda media y presentan un campo receptivo grande, vinculándose con la visión nocturna. Los segundos, relacionados con la visión del color, son sensibles a tres tipos de longitud de onda: larga, media y corta y presentan campos receptivos menores. La señal, captada por los bastones y conos, es procesada por las células bipolares e inhibida por las neuronas horizontales que tratan las señales del campo receptivo (ver Figura 17.a). En el campo receptivo, el centro se estimula por señales intensas, mientras que el entorno es inhibido por unas señales que reducen su intensidad según la distancia del centro (ver Figura 17.b). Además, se procesa la señal en dos vías paralelas: ON y OFF, tal y como se describe en los trabajos de Schiller y colaboradores (Schiller y otros, 1986) y Hubel (Hubel, 1988) entre otros científicos. Ambas vías representan la intensidad del estímulo (ON) y su ausencia (OFF). Cada vía contiene la misma estructura de células, pero con valores opuestos, ya que para el procesamiento posterior es necesario tener tanto la información de actividad como de inactividad. Cada célula ganglionar recibe las señales de su campo receptivo (ver ver Figura 17.c), en el cual existe un centro estimulador y un entorno inhibidor, y varía en relación con el rango de longitud de onda:

- Parvocélulas (células midget), hay dos tipos, las que son estimuladas por ondas de longitud larga (denominadas L) en el centro y son inhibidas por ondas de longitud media (denominadas M) en el entorno y, al contrario, por ondas de longitud media en el centro e inhibidas por ondas de longitud larga en el entorno.

- Koniocélulas (biestratificadas), las que son estimuladas por ondas de longitud corta (denominadas S) en el centro, y son inhibidas con la suma de ondas de longitud larga y media en el entorno.

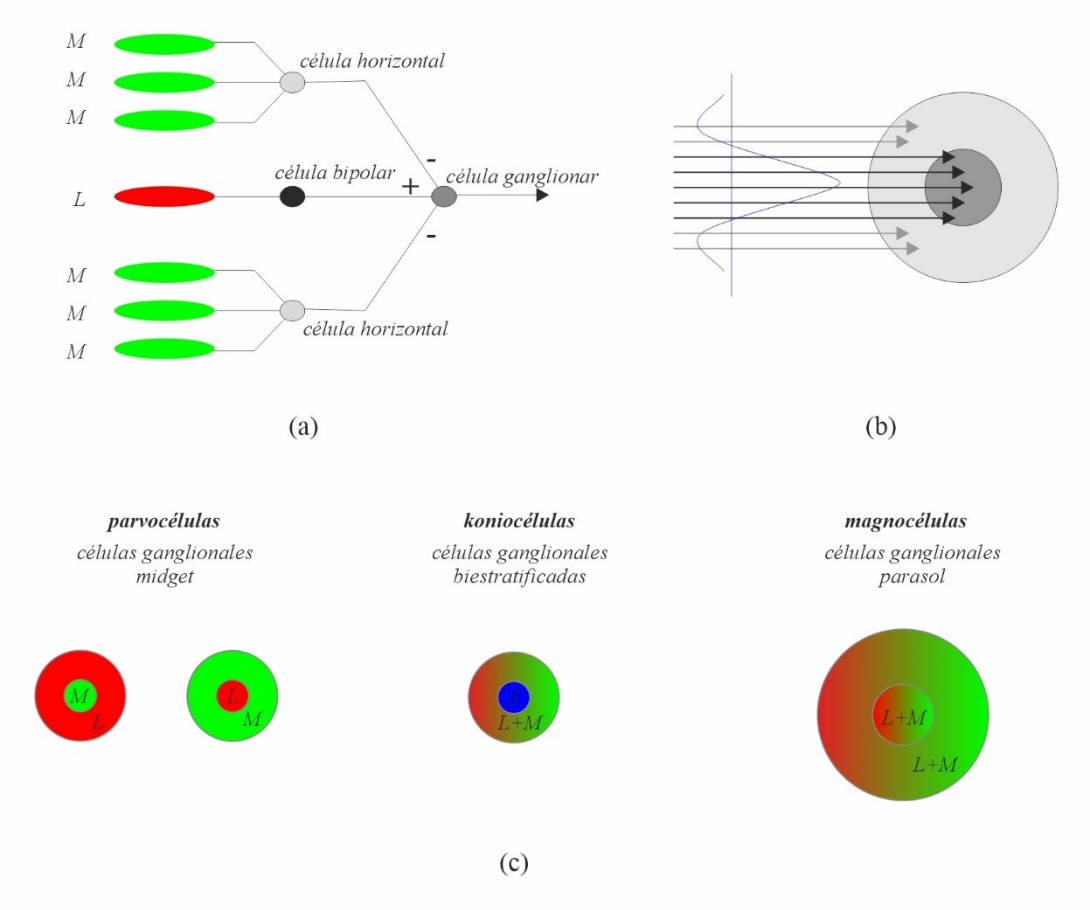


Figura 17 Campo receptivo de una célula ganglional de la retina.
Nota: (a) Esquema de células del campo receptivo para una parvocélula sensible a la longitud de onda larga; (b) modelización del campo receptivo; y (c) campos receptivos según el tipo de células ganglionares.

- Magnocélulas (parasol), las que son estimuladas por la suma de la longitud de onda larga y media (denominadas LM) en el centro y son inhibidas por la suma de la longitud de onda larga y media en el entorno.

2.2.4 NGL

Durante mucho tiempo, el NGL fue un desconocido al que se le atribuía un papel de transmisor de información entre la retina y la corteza visual (Crick, 1994). Sin embargo, en los últimos veinte años, se ha avanzado en su conocimiento, y hoy en día se le atribuye la funcionalidad de «puerta de datos» más allá del de «transmisión de datos». Distanrtas investigaciones de (Teller, 2014), (Mel y otros, 1998), (Sherman, 2005) o (Einevoll, 2003) han avanzado el estudio tanto de la estructura y las relaciones de retroalimentación desde el área primaria visual como de la funcionalidad y la computación.

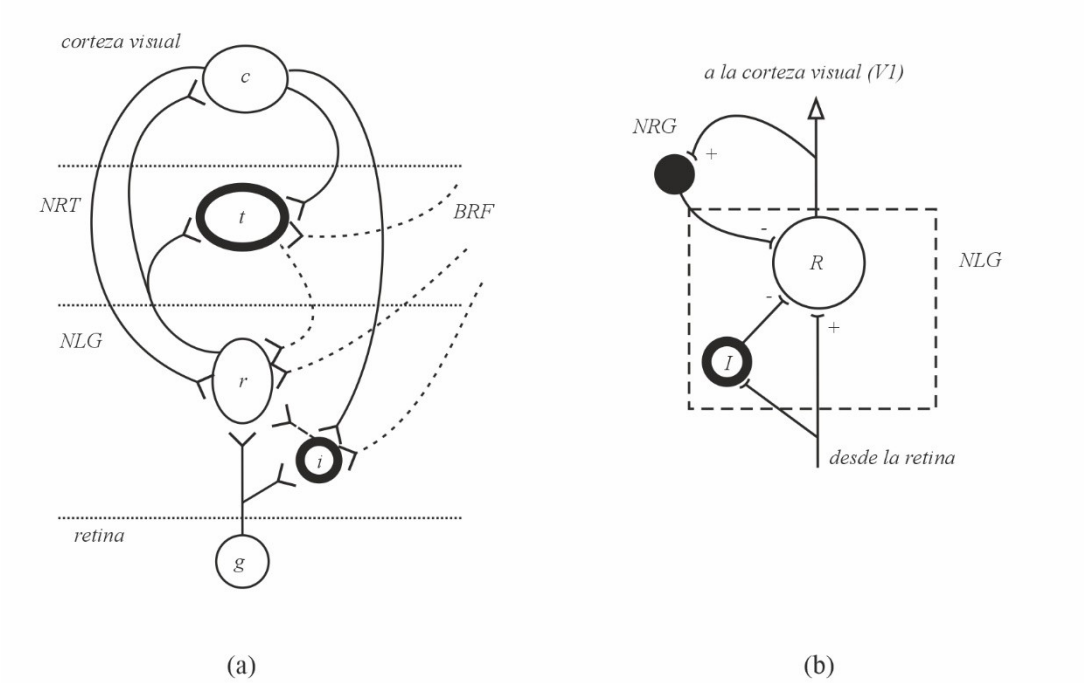


Figura 18 Esquemas de la funcionalidad relé en el NGL según Einevoll y Teller.
Nota: (a) modelo de Einevoll; y (b) modelo de Teller.

La función de relé es quizás uno de los aspectos más analizados de los campos receptivos del NGL, ya que permite, por un lado, inhibir, como describe Alitto y colaboradores. en su estudio sobre las conexiones de retroalmentación entre el tálamo y la corteza (Alitto & Usrey, 2003) y, por el otro, modular, amplificando o disminuyendo la entrada, como describe por Agarwal y Sarma en su estudio sobre la funcionalidad relé (Agarwal & Sarma, 2011). Esta función se desarrolla en el NGL con cierta complejidad, ya que cerca del 70% de las conexiones que tiene con la neorocrteza son de retroalimentación. La Figura 18.a muestra el esquema de conexiones planteado por Einvoll (Einevoll, 2003) con un enfoque biológico donde las funciones de modulación e inhibición provienen de la formación reticular del tronco encefálico (BRF). Einvoll indica que la inhibición es realizada por el núcleo retícular del tálamo, que controla, entre otras funcionalidades, el movimiento de los ojos y la coordinación de ambos, que a su vez recibe flujos de estimulación de la corteza visual.

Por otro lado, Davila Teller (Teller, 2014) describe una estructura más simple (Figura 18.b), donde el campo receptivo (denominado unidad relé) recibe un flujo de datos desde las células ganglionales, tanto de estimulación en la región central como de inhibición desde la denominada intercélula que recoge la señal de su entorno —ambos forman el campo receptivo—, y una segunda vía de modulación desde núcleo reticular del tálamo (NRG) que proviene de la señal de salida anterior. En esta propuesta de Teller, se establece, por un lado, la entrada de datos con el campo receptivo y, por el otro, la modulación por la salida anterior de la unidad relé. Esto determina el procesamiento de la señal temporal que describe Norheim y colaboradores. (Norheim y otros, 2012) en el mecanismo del proceso de la señal temporal en el NGL. Existen más posibilidades de control aparte de las de modulación e inhibición sobre la entrada recibida, pero estas dos definen perfectamente la funcionalidad del NGL para el control de los datos.

2.2.5 Representación de la información: los mapas retinotópicos

Los mapas retinotópicos fueron descubiertos por Tatsuji Inuye y Gordon Holmes (Fishman, 1997) cuando ambos, independientemente, llegaron a las mismas conclusiones tras analizar las áreas dañadas en la corteza visual de pacientes que habían perdido la visión en regiones específicas del campo visual. Existía una relación topográfica entre las distintas regiones del espacio visual no visibles por parte de los pacientes y las áreas dañadas de la corteza visual, de tal manera que se podían establecer una correspondencia entre ellas. Su estudio avanzó posteriormente con otras investigaciones a través de las cuales se han establecido las siguientes características principales:

- La excentricidad. Existe una reducción de las neuronas que representan el espacio visual desde la región central hacia la externa, de tal manera que más de la mitad de ellas representan un grado de ángulo del campo visual, lo que equivale a la región que procesa la fóvea en la retina.
- La división de espacio visual en dos regiones. La existencia de un mapa para la región visual izquierda en el hemisferio derecho del cerebro, e igual para la región derecha en el izquierdo.

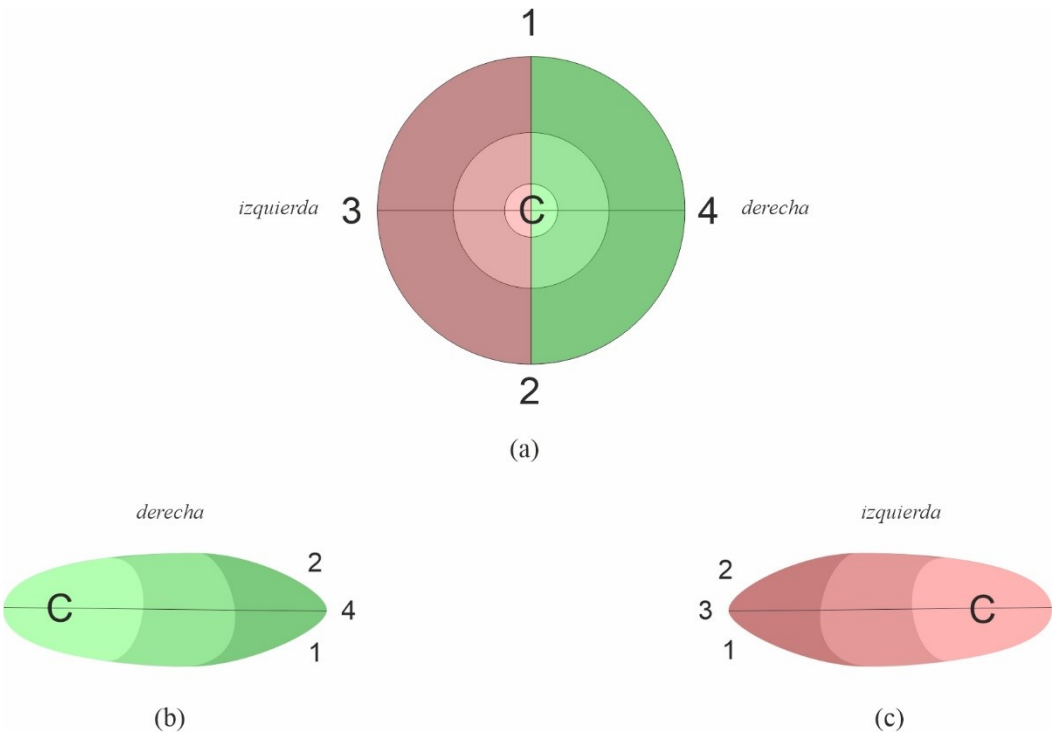


Figura 19 Mapas retinotópicos y sus relaciones topográficas con el espacio visual.
Nota: (a) espacio visual; (b) simplificación del mapa retinotópico de la región derecha, situada en el hemisferio izquierdo del cerebro; y (c) el mapa retinotópico correspondiente a la región izquierda, situado en el hemisferio derecho.

Como ejemplo, en la Figura 19, vemos un esquema simplificado de los mapas retinotópicos del área V1. La Figura 19.a muestra el espacio visual definido con coordenadas polares —la numeración indica las regiones de arriba, abajo, izquierda y derecha. La Figura 19.b muestra el mapa retinotópico de la región izquierda en el hemisferio derecho del cerebro y, la Figura 19.c indica el de la región derecha en el izquierdo. En ambos mapas, la región central ocupa cerca del 75%, mientras que los extremos (arriba, abajo, izquierda o derecha) se concentran en el 25% restante.

2.2.5.1 Campos receptivos

Cada neurona del mapa retinotópico representa una región del espacio visual a través de los campos receptivos, con un centro estimulador y un entorno inhibidor. El tamaño varía dependiendo de la distancia con el centro del mapa retinotópico, siendo más pequeño en la región central y mayor cuanto más se aleja. Existe una relación, por consiguiente, entre la excentricidad y el tamaño de los campos receptivos, tal y como Hubel lo describe (Hubel, 1988). Además, los campos receptivos varían en cuanto a su forma, por ejemplo, en el área V1 suelen ser alargados en vez de circulares. Se ha estudiado en los últimos años las diversas variedades existentes de campos receptivos para relacionar la funcionalidad con su forma, aunque en muchos casos se desconoce la finalidad final. Por ejemplo, los campos receptivos extraclásicos (Solomon y otros, 2002), en los que, a diferencia de los campos receptivos denominados clásicos, al centro estimulador y al entorno inhibidor, se añade un área mayor con una finalidad estimuladora. En la Figura 20, podemos ver unos ejemplos extraídos del artículo de Jeffries y colaboradores (Jeffries y otros, 2014), donde se ve la diferencia de tamaño entre un campo receptivo clásico, como el de las células ganglionares, y el denominado como extraclási-

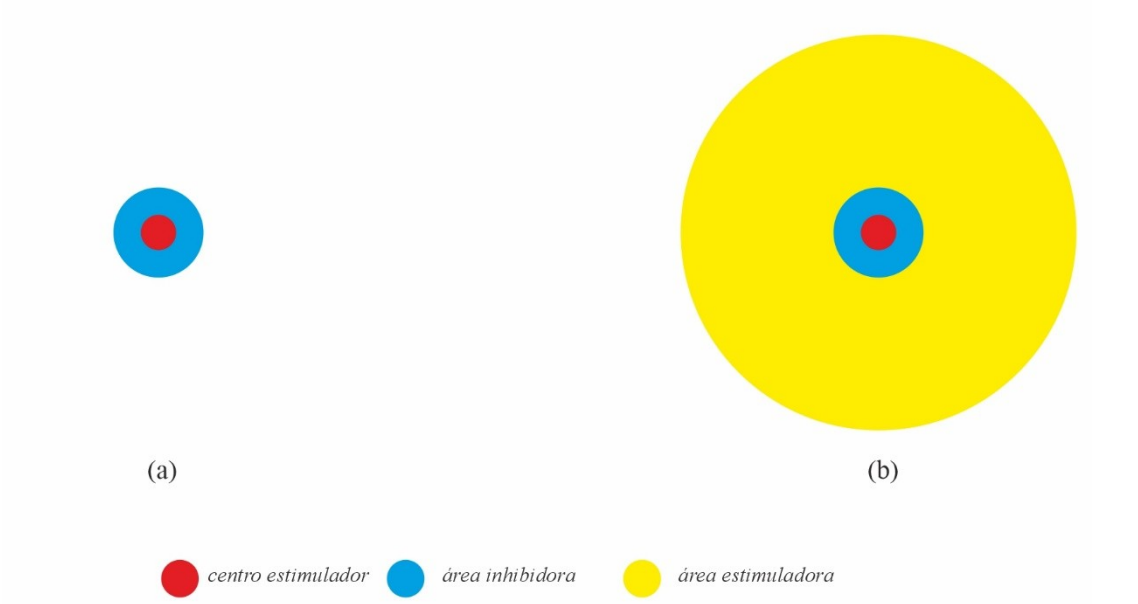


Figura 20 Campos receptivos clásicos y extraclásicos.
Nota: (a) campo receptivo clásico con un centro de estimulación y un área de inhibición; y (b) campo receptivo extraclásico con un centro de estimulación, un área de inhibición y un área mayor de estimulación.

co. Para comparar la dimensión de cada área, la región central estimuladora está en rojo, el entorno inhibidor en azul y el entorno estimulador del campo extraclásico en amarillo.

2.2.5.2 Magnificación y anisotropía

La arquitectura del mapa retinotópico de la retina representa el espacio visual con la función de excentricidad, donde la región central tiene una representación mayor en el mapa, sin embargo, en los mapas del NGL, y también en la corteza visual, está representación es incluso más grande. Inicialmente, en el trabajo de Holmes (Holmes, 1945), se establecía una arquitectura de la representación visual basada sólo en la excentricidad, pero como indicó Holton y Hoyt en 1991 (Horton & Hoyt, 1991), existía una «magnificación» que era inversamente proporcional a la excentricidad, y que se había constatado en muchos de los mapas retinotópicos. Este efecto ha sido estudiado con el fin de determinar dónde se produce y en qué grado, como en el trabajo de Scheneider y colaboradores en el que, a partir del análisis de varios sujetos por neuroimagen se obtuvo una fórmula (Schneider y otros, 2004) que con alguna variación se muestra también el estudio de Qiu y colaboradores (Qiu y otros, 2006), aunque los resultados son parecidos.

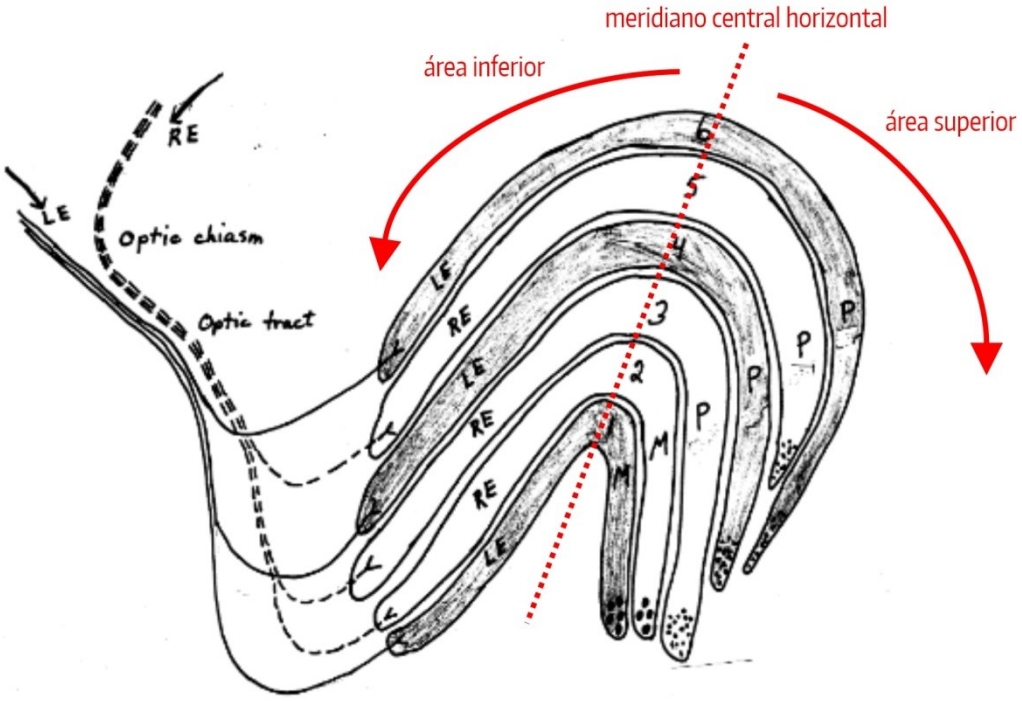


Figura 21 Eje meridiano central horizontal del NGL.
Nota: corte transversal que muestra las capas correspondientes a las vías separadas parvocelulares y magnocelulares. El meridiano central horizontal esta superpuesto a la ilustración original y se indica la distribución del área inferior y superior. *Adaptación del original de Teller* (Teller, 2014)

Existe además un efecto denominado anisotropía horizontal-vertical, izquierda-derecha e inferior-superior (Van Essen y otros, 1984) que se encuentra presente en muchos de los mapas retinotópicos, también en el NGL. Este efecto provoca que existan en los mapas una mayor cantidad de neuronas y, por lo tanto, una mayor representación del espacio visual, en el eje horizontal que el vertical, en la región izquierda que en la derecha y en la inferior que en la superior.

Otro aspecto importante en las arquitecturas de los mapas retinotópicos del NGL es que el meridiano horizontal central se convierte en un «centro» para la arquitectura. La Figura 21 muestra una reproducción del mapa izquierdo del NGL, donde es perceptible con claridad la organización acorde al eje horizontal que actúa como «centro», separando el área que representa la región superior de la inferior. Cada uno de los canales están representados (P, parvocélulas, y M, magnocélulas, y entre cada P, las koniocélulas, que no aparecen definidas en la figura porque fue un descubrimiento posterior), así como las áreas que representan la región derecha del ojo izquierdo y del derecho en el ojo derecho (LE y LR en la figura).

2.2.6 Esquemas de referencia del área parietal

En la zona parietal inferior se localizan los mapas visuales, espaciales y visio-motores descritos en el trabajo de Cohen y Andersen (Cohen & Andersen, 2002) los cuales son conocidos como esquemas de referencia por la funcionalidad de relacionar los distintos elementos presentes en el espacio visual con el sujeto y partes de su cuerpo. Estos esquemas de referencia han sido estudiados por distintos autores, como (Pouget & Sejnowski, 1997), (Galati y otros, 2010), (Pertzov y otros, 2011) o (Chen y otros, 2012). Colby indica que actúan como una representación topográfica del espacio físico que nos rodea (Colby, 1998), usualmente en tres dimensiones, donde existe un centro y relaciones con su entorno.

Al igual que los mapas retinotópicos, existe una excentricidad, aunque menor en cuanto a la magnificación del centro y se pueden representar con coordenadas polares, tal y como los han analizado Fattori y Pitzalis en sus estudios de los cerebros de los macacos y humanos (Fattori & Pitzalis, 2009) o como mapas topográficos por Lehky y Sereno. (Lehky & Sereno, 2010)

Los esquemas de referencia no sólo establecen relaciones con el espacio visual, sino también con distintas partes del cuerpo, conectándose a su vez con todo el aparato motor, como es descrito por Hadjidimitrakis et al para relacionar la información visual con las acciones a realizar (Hadjidimitrakis y otros, 2012). Los esquemas de referencia permiten la coordinación espacial, facilitando la relación espacial para localizar objetos, o interactuar con ellos. En la Figura 22, el ejemplo muestra un esquema de referencia espacial centrado en el sujeto y que establece la distancia y posición de varios objetos del entorno. En la Figura 22.a, el espectador detecta, dentro de su campo visual, un objeto azul y otro naranja. En la Figura 22.b, girando 180 grados y dejando por detrás al azul y al naranja, que permanecerán en sus posiciones en el esquema de referencia, de-

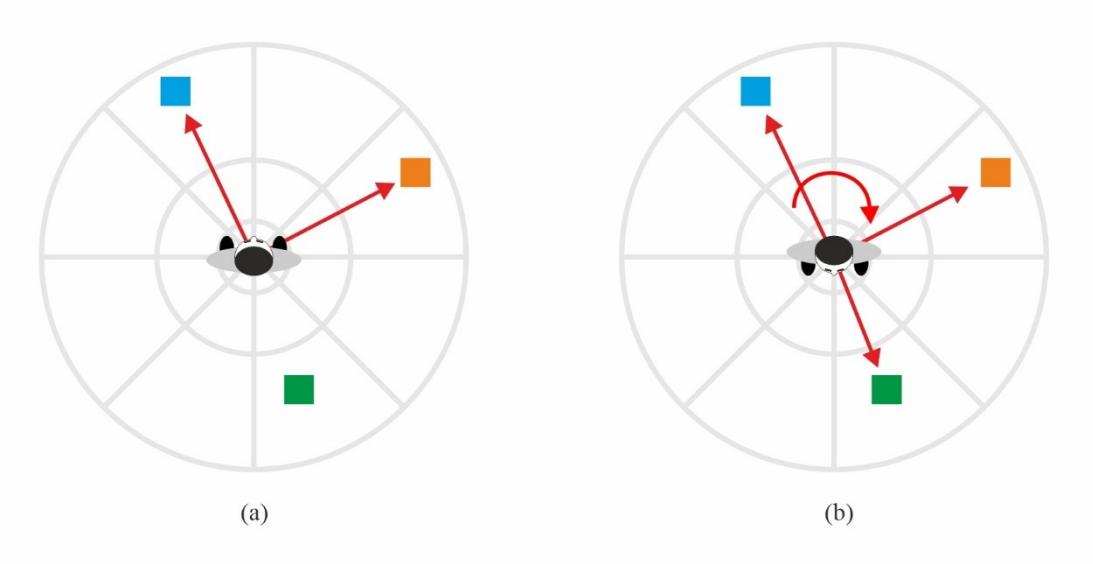


Figura 22 Ejemplo de esquema de referencia egocéntrico.
Nota: (a) se localizan el objeto azul y naranja; y b) se gira 180º y se localiza el objeto verde.

tecta un nuevo objeto verde. El esquema de referencia, con el centro en la persona, tendrá relacionado los tres cuadros.

Existen varios tipos de esquema de referencia, y aunque mantienen muchas de sus características como la excentricidad, varían en la cantidad de neuronas. La literatura determina dos tipos principales (Klatzky, 1998):

- Egocéntrico: centrado en el propio sujeto (los ojos, la cabeza, el brazo, el cuerpo, etc.).
- Alocéntrico (también conocido como geocéntrico o exocéntrico): centrado en algún elemento del espacio visual.

Las relaciones que se establecen entre ellos hacen posible la interacción con el espacio, el movimiento, coger objetos, etc. como analiza Meilinger en su síntesis de mapas gráficos y cognitivos en relación con la red neuronal de los esquemas de referencia (Meilinger, 2008).

Se puede deducir que los esquemas son fijos en el tiempo y sus neuronas van modificado sus valores dependiendo de los mapas retinotópicos que se generan en cada momento en el área visual y que, obviamente, amplían información con detalles concretos. Por consiguiente, estos esquemas actúan como una memoria visual —de hecho, se relacionan con parte del cerebro destinada para este fin— para facilitar la interacción con el mundo (Pertzov y otros, 2011).

2.2.7 Procesamiento de la percepción visual

Durante la década de los noventa, los avances en el estudio del cerebro fueron arrojando luz sobre su anatomía y la especialización funcional de sus áreas o el procesamiento

no secuencial (Zeki, 1993). Existen áreas en la corteza visual destinadas a tareas muy concretas, como la detección de formas y del color en el área V4, la extracción de bordes en el área V1 o la visión espacial en el área V2, así como tareas mucho más especializadas como el reconocimiento facial, aunque en trabajos recientes se ha descubierto que el procesamiento no es tan independiente y que, si bien hay una especialización funcional, se establecen conexiones entre las áreas durante el procesamiento (Zeki, 2005). Por otro lado, cada área está conectada con el resto del sistema visual y de otras áreas del cerebro, de tal manera que existen procesos de realimentación e interconexión que no siguen un proceso secuencial. En 1994, Ferrera y colaboradores (Ferrera y otros, 1994) verificaron que el área V4, especializado en la detección de formas y el color, recibía información de las áreas V1, V2 y V3 —donde el color está entremezclado con otras características como los bordes, contornos, texturas, etc. Además, almacenaba la información con una estructura similar a la existente en el NGL Los autores concluyeron que esto se debía a que esta estructura se mantenía en las distintas áreas visuales de la corteza, de donde en teoría recibía la información y no preveían que hubiera una conexión directa, ya que pensaban que esa área precortical sólo tenía una conexión directa con la corteza visual a través del área V1 y sólo recibía información retroalimentada del resto de las áreas visuales. Recientemente, se han mapeado las conexiones de las áreas precorticales con la corteza visual (Arrigo y otros, 2016), y se ha concluido que sí existe una conexión directa del NGL en los humanos con el área visual V4, lo cual explica por qué esa estructura de información de la visión temprana fue localizada en el V4 anteriormente por (Ferrera y otros, 1994).

La especialización en el tiempo y el espacio es uno de los principales aspectos del área visual del cerebro que Zeki denomina «cronoarquitectura» (Zeki, 2005). El procesamiento de la información no es sincronizado o en paralelo, sino que, por ejemplo, tareas como el reconocimiento del color o de movimiento se realizan de una manera separada y en centésimas de segundo distintas. Se desconoce si esta diferencia tiene un valor inhibitorio o estimulador, o simplemente actúan en tiempos distintos sin aparente relación funcional.

Otro aspecto importante es la denominada por Zeki como «microconsciencia», que deriva de la cronoarquitectura, y le da sentido a la modularidad y a la especialización funcional y temporal. En el estudio de un caso de daños en la corteza visual (Zeki, 2005), Zeki comprobó que un paciente con lesiones en el área V4, donde se procesa el color, y en las áreas V1 y V2, situados en la corteza visual primaria y en donde se realizan tareas como la segmentación o detección de bordes, era incapaz de detectar objetos y colores, pero sí era capaz de procesar el movimiento. A priori, y siguiendo un planteamiento secuencial, esto no podía suceder, ya que sin el funcionamiento del área V1 y el V2, no sería posible que pudiera realizar tareas más complejas y, por lo tanto, ser consciente del movimiento. Zeki planteó que no podía existir una única consciencia que controlara los procesos y analizara la información final sino, más bien, existían varias consciencias dependientes de cada área según su funcionalidad que denominó como «microconsciencias».

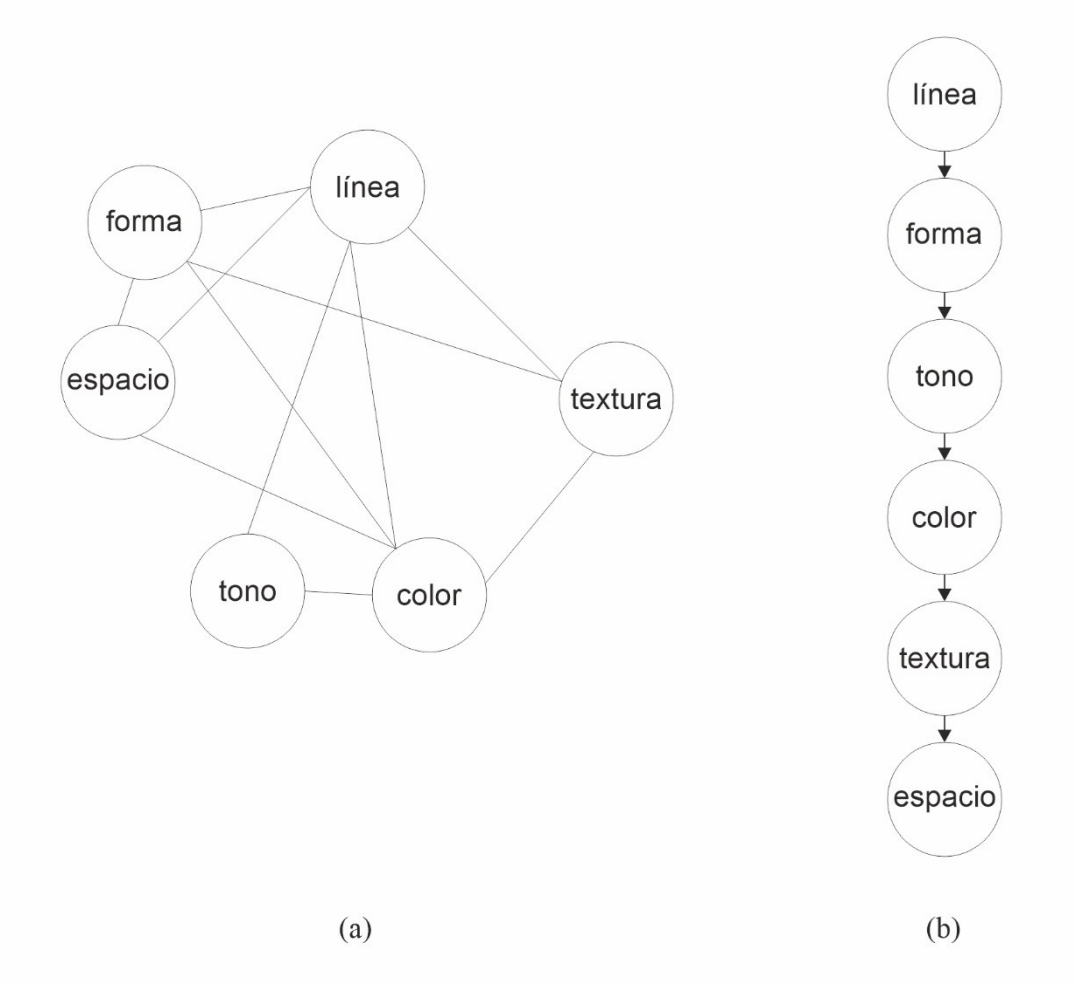


Figura 23 Tipos de procesamiento de la información visual.
Nota: (a) procesamiento por módulos visuales, la consciencia emerge en cada módulo como microconsciencia; y (b) procesamiento secuencial, la consciencia emerge al final del proceso.

La relación entre la especialización funcional, la cronoarquitectura y las microconsciencias implica un modelo de procesamiento menos organizado y secuencial de lo esperado. Cada módulo, especializado tanto funcional como anatómicamente, actúa de una manera asíncrona, por lo cual, la consciencia visual se haya distribuida. No existe en ese sentido, ni un centro específico, ni siempre es el mismo, ya que, según cada momento, un área u otra elevan la consciencia de una característica visual concreta. La Figura 23.a muestra el procesamiento por módulos con la cronoarquitectura y la microconsciencias, y la Figura 23.b muestra un procesamiento secuencial con una única consciencia al final del proceso.

2.3 Bases de la visión artificial

En las décadas de 1950 y 1960, la ciencia de la computación tenía como objetivo construir máquinas con capacidades motoras y cognitivas como las de los humanos, lo que llevó a la creencia de que tareas como caminar, ver o hablar podrían lograrse en un corto plazo. Sin embargo, los avances posteriores demostraron que estos objetivos habían

sido subestimados y se consideraron pronto inalcanzables. Esto condujo a que, en la década de 1970, se redirigieran todos los esfuerzos hacia la consecución de objetivos más modestos en el campo de la visión temprana. Los ingenieros coincidieron en que la percepción visual involucraba niveles de extracción de características, como contornos, formas y colores, así como niveles de asociación y reconocimiento de patrones. Los primeros éxitos, especialmente en la detección de bordes y la construcción de contornos, despertaron el interés de la neurociencia, que en ese momento se centraba en la relación anatómica del sistema de percepción visual y su funcionalidad.

En la década de 1970, las investigaciones de David Marr (Marr, 1982) sentaron las bases tanto para la visión artificial como para la psicología de la percepción visual y la neurociencia. Esto se debió no solo al avance que representaron muchos de sus algoritmos, la creación de un paradigma para la implementación de la visión artificial y el establecimiento de bases teóricas seguidas por muchos otros investigadores, sino también porque Marr fue el primero en tratar la visión artificial como un proceso de computación completo, con una teoría de la computación, una representación y una implementación en hardware.

Posteriormente, la relación entre la visión artificial y la neurociencia ha sido y sigue siendo altamente productiva, como indican, por ejemplo, Brown en su trabajo sobre las restricciones naturales de la visión artificial (Brown, 1984). En la actualidad, los avances en neuroimagen están permitiendo un conocimiento más profundo de cada área del cerebro involucrada en la percepción visual. Esto proporciona un valioso conocimiento de la computación natural que es crucial para el desarrollo de la visión artificial, como señalan Cox y Dean en su estudio sobre la bioinspiración (Cox & Dean, 2014), o Kruger y colaboradores en su investigación sobre lo que la visión artificial puede aprender del área primaria de la corteza visual (Kruger, y otros, 2013).

En esta sección, se resumen las principales teorías e implementaciones de la visión artificial que se han estudiado y analizado en el contexto de este proyecto. En primer lugar, se abordan los conceptos fundamentales propuestos por David Marr para un enfoque neurocientífico en la creación de modelos de computación. En segundo lugar, se examina el problema del salto semántico en el procesamiento de imágenes y los sistemas de recuperación de imágenes basados en el contenido (CBIR). En tercer lugar, se consideran los flujos de procesamiento y su relación con el salto semántico. En cuarto lugar, se exploran los mapas de prominencia y su relación con la atención y el análisis del espacio visual como representación de las áreas más relevantes para el procesamiento. En quinto lugar, se examina el impacto de las redes neuronales convolucionales en la última década y su relación con el salto semántico. Finalmente, se revisan las implementaciones relacionadas con la composición de imágenes artísticas.

2.3.1 El procesamiento de la información visual

En el modelo planteado por Marr, el objetivo era obtener la comprensión de la escena visual que representa la imagen a partir de la matriz de píxeles. Cada fase avanza sin más información que la de la fase anterior, en un proceso de abajo hacia arriba: después del

reconocimiento de bordes y detección de posibles formas (blobs), se reconoce los objetos para recuperar la estructura tridimensional y alcanzar una descripción semántica de la escena. Después de Marr, y hasta finales de los años noventa, todos los esfuerzos se centraron en mejorar los diversos algoritmos en cada una de las fases, aunque principalmente se centraron en la fase de extracción de características.

Sin embargo, el avance de trabajos como el de Andrej W. Przybyszewski (Przybyszewski, 1998) mostraron que áreas relacionadas con la extracción de características en un nivel bajo de procesamiento tenían conexiones directas con áreas de nivel alto, donde se asocian y se obtienen patrones con un procesamiento de arriba hacia abajo (*top-down*). Esto indica que la tarea de extracción de características está guiada por un conocimiento de nivel superior, como sucede, por ejemplo, en el reconocimiento de bordes que se lleva a cabo con información proveniente de la escena visual y, además, con un conocimiento de alto nivel sobre el tipo de borde a reconocer.

Con el cambio de siglo, los avances en neurociencia, sobre todo en el análisis de la relación entre las estructuras neuronales de las áreas cerebrales y su funcionalidad, bioinspiraron nuevas soluciones. Por ejemplo, la computación laminar, común en la zona estriada de la corteza visual, ayudó a comprender cómo el área de la corteza visual primaria procesaba la información de una manera más eficaz para detectar contornos y formas usando conexiones laterales, como revela el trabajo de Raizada y Grossberg (Raizada & Grossberg, 2003).

2.3.2 El salto semántico

El salto semántico es una interrupción en el procesamiento de la información al pasar de una realidad física (píxel) a otra semántica (conceptos), o lo que es lo mismo de una realidad computacional matemática basada en números a otra semántica basada en conceptos y estructuras de conocimiento. Este problema no es exclusivo de la visión artificial, sino que está presente en otras disciplinas como la lingüística, la teoría de señales y, en general, en la semiótica. La Figura 24 muestra un ejemplo con la pintura de Velázquez, La Fragua del Vulcano, donde el procesamiento de nivel bajo extrae las características visuales de la matriz de píxeles de la imagen, por ejemplo, el tono, el color, la categoría de color, la línea, las formas o el movimiento. El salto semántico se establece entre las características visuales extraídas y la descripción del contenido de lo que la escena representa.

Figura 24 Salto semántico.

Figura 25 Esquema funcional de un mapa de prominencia.

De una manera general, muchos de los investigadores percibieron que era un problema de la percepción y no de la propia imagen. La solución dependía de la capacidad de encontrar un proceso o conjunto de procesos que permitieran trazar un puente, tal y como el problema es presentado a finales del siglo XX en el trabajo de Smeulders y colaboradores (Smeulders y otros, 2000). A principios del siglo XXI, muchos investigadores analizaron el problema desde un nuevo enfoque en visión artificial que se comenzó a denominar como *Image understanding* (Shah, 2002).

Para salvar este salto, se comenzaron a «tender puentes» entre una y otra orilla, como por ejemplo, la aplicación de etiquetas semánticas que relacionaban ciertas características visuales con un concepto, como se ve en los distintos trabajos (Hare y otros, 2006), (Dorai & Venkatesh, 2003) o (Zhao & Grosky, 2002), entre otros. El resurgimiento de las redes neuronales convolcuionales a principios de la segunda década del siglo XXI reduciría cada vez más este salto con la creación de modelos con capas dedicadas a la extracción de características seguidas de capas destinadas a la asociación de patrones con una capa de salida con etiquetas semánticas.

2.3.3 Los mapas de prominencia

En los años 80, Koch y Ullman elaboraron una teoría de la atención relacionada con la percepción para describir cómo se producía el escaneo del espacio visual, y, de una manera inconsciente, los ojos se dirigían a las regiones más relevantes de ese espacio. En «Shifts in selective visual attention: towards the underlying neural circuity» (Koch & Ullman, 1987), se describen estructuras neuronales simples que son capaces de desarrollar procesos de atención, y con esta bioinspiración crearon un modelo dentro de la visión artificial que denominaron «mapa de prominencia» (*saliency maps*).

Este modelo, aplicado al análisis de imágenes digitales, establece dos fases secuenciales: en la primera, se extraen los mapas de características visuales, por ejemplo, de contornos, del color, de la orientación, de la dirección del movimiento, etc. En la segunda, se compila el mapa de prominencia a partir de las características visuales más relevantes en cada región (ver Figura 25).

El modelo fue mejorado en los siguientes años (Itti y otros, 1998) variando los procedimientos, aunque los procesos principales tanto de extracción como de compilación se mantenían. A partir de este trabajo, la selección de regiones prominentes y su relación con la atención ha facilitado el surgimiento de variantes del modelo original y desarro-

llos nuevos para generar un campo propio dentro de la visión artificial, que es descrito ampliamente en el estudio « State-of-the-art in visual attention modeling» (Borji & Itti, 2013). En el estudio, se realiza una comparativa entre los modelos existentes a partir varios criterios, como el tipo de análisis espacial que utilizan, la relación espaciotemporal o el tipo de flujo que usan para procesar y controlar los datos (de arriba hacia abajo o de abajo hacia arriba).

Borji e Itti (Borji & Itti, 2013) utilizan una clasificación para los modelos de mapas de prominencia dependiendo del proceso usado para la determinación de las regiones prominentes: modelos cognitivos, modelos bayesianos, modelos de decisión, modelos de información, modelos de análisis espectral y modelos de clasificación de patrones. Dentro de los modelos cognitivos, el de Zhaoping Li incluye una implementación neuronal basándose en el área V1 de la corteza visual de mamíferos (Lin & Fang, 2010), o el modelo basado en el sistema de visual humano de Le Meur (Le Meur y otros, 2006), más bioinspirado, donde se establecen funciones de sensibilidad, descomposición perceptual, enmascaramiento visual o interacciones centro-entorno, llegando a extenderlo con controles espacio-temporales. En este modelo, cada característica visual es gestionada de una forma modular, y el modelo de color realiza una transformación de la información RGB tricromática a un sistema de color de procesos de colores opuestos.

2.3.4 Las redes neuronales convolucionales

En los últimos años, las redes neuronales convolucionales han tenido una importante evolución en el campo de la visión artificial, obteniendo éxitos en tareas como la clasificación, reconocimiento o localización de objetos. Gran parte de este éxito se debe a la evolución del hardware, sobre todo con el uso de las GPU, pero también al amplio desarrollo de software con diversos *frameworks* y librerías que simplifican la implementación. Otro aspecto clave es la bioinspiración, que ha permitido la introducción de conceptos como los campos receptivos o las células simples y complejas que desarrollan la convolución y los filtros de estas redes.

En 1980, Kunihiko Fukushima desarrolló el neocognitron (Fukushima, 1980), un modelo de red neuronal artificial basado en los trabajos de Hubel y Wiesel (Hubel & Wiesel, 1965) relacionados con las neuronas simples y complejas del área V1 y su fun-

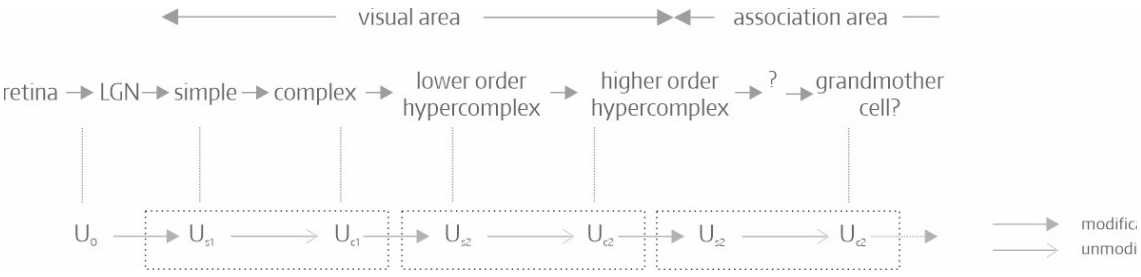


Figura 26 Arquitectura del Neocognitron
Nota: relación de la arquitectura de la red neuronal artificial con el sistema de percepción visual de la corteza visual (*adaptación del original de Fukushima*).

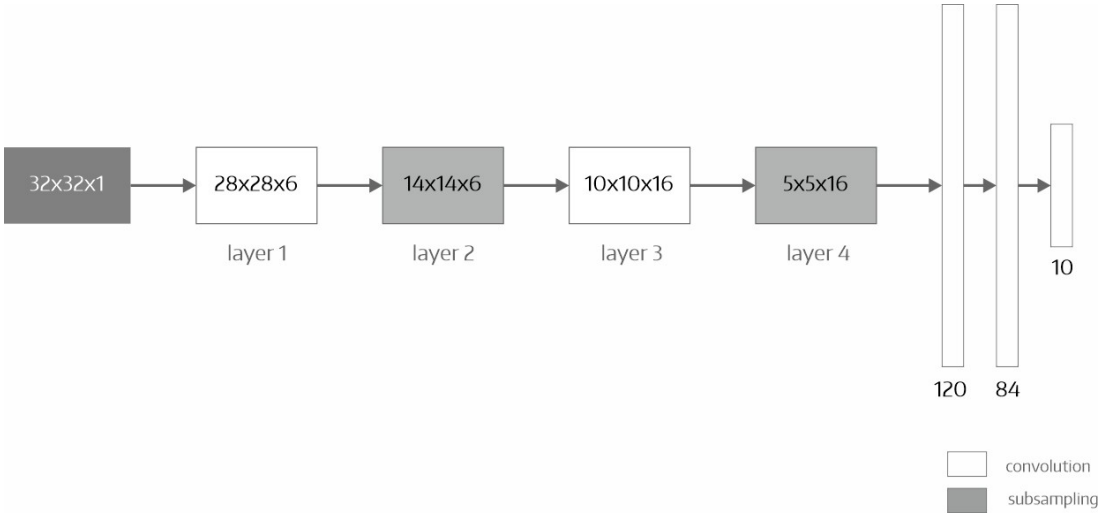


Figura 27 Arquitectura de LeNet 5.

cionalidad para la detección de bordes. Esta nueva red neuronal tenía dos áreas separadas: la visual y la de asociación. Cada una se componía de varias etapas donde se aplicaban los campos receptivos transformados en kernels de convolución y se reducían las dimensiones de la imagen con operaciones de subsampling, permitiendo detectar características con independencia de su situación en la imagen y su tamaño.

La Figura 26 muestra la relación entre la arquitectura del Neocognitron y el sistema de percepción visual según el trabajo de Hubel y Wiesel, donde la capa de entrada es la retina y el NGL, el primer bloque corresponde a las neuronas simples y complejas del área visual V1, el segundo a las neuronas hipercomplejas en un orden que Fukushima denomina de orden bajo y, posteriormente, alto y el tercero, a las neuronas de asociación. Este modelo tuvo éxito en el reconocimiento de caracteres, pero durante un tiempo no tuvo continuidad hasta que en 1989 Lecun y colaboradores. (LeCun y otros, 1989) presentaron la arquitectura mejorada LeNet para el reconocimiento de códigos ZIP del servicio postal de EE.UU. LeNet simplifica la estructura y su arquitectura: establece las capas de convolución y sus kernels e introduce las dos últimas capas conectadas completamente con la salida correspondiente a cada clase, así como los procesos de entrenamiento y validación (ver arquitectura LeNet 5 en Figura 27). Este modelo es mejorado a finales de los años 90, con la denominada LeNet 5 (LeCun y otros, 1998), que es la base de las siguientes arquitecturas en la primera década del siglo XXI, y sobre todo de Alexnet (Krizhevsky y otros, 2012). A partir de 2011, aparecen modelos con una mayor cantidad de capas y neuronas que incluyen mejoras tanto en las arquitecturas como en las técnicas de entrenamiento y procesamiento de los conjuntos de datos. En el año 2020 surge una nueva arquitectura, los *vision transformers*, donde se utilizan capas de atención sin convolución. La evolución de esta área ha sido tan fructífera que ha modificado muchos de los objetivos de la visión artificial.

2.3.5 La composición de las imágenes en la visión artificial

Los modelos que se han desarrollado en relación con la composición se han enfocado mayoritariamente en la evaluación estética, dependiendo de la tecnología dominante en cada momento. Uno de los más avanzados es el de Galanter (Galanter, 2012), que aplica técnicas y métodos de la psicología del arte con la finalidad de evaluar la estética en imágenes artísticas. Destaca la implementación de las teorías de la Gestalt, de reglas de composición como la regla de los tercios o la sección de oro, como en (Datta y otros, 2006) en el que se extraen características visuales como el color o la profundidad de campo, y se aplican métodos de composición como la regla de los tercios en una base de datos de fotografías.

El interés por analizar la composición empieza a tener sentido cuando los avances permiten obtener modelos más sofisticados que facilitan análisis formales más complejos, como localización de la línea de horizonte, la aplicación de la regla de los tercios, el uso de modelos autorregresivos para obtener ejes de composición, entre otros métodos aplicados. Es en el campo CBIR donde se aplican más este tipo de soluciones, como se puede ver en el trabajo de (Maeda y otros, 1999) donde se mezclan la extracción de características visuales y las estructuras de composición como la regla de los tercios para obtener un criterio de búsqueda con el que realizar la búsqueda en la base de datos.

Por otro lado, algunos proyectos extraen las estructuras de la composición a partir de patrones preestablecidos, como, por ejemplo, el eje vertical, el eje horizontal, diagonal, ejes paralelos, etc. tiene un avance importante en los últimos años, por ejemplo, en (Zhang y otros, 2021). También, otros proyectos aplican técnica de aprendizaje automático por su capacidad para obtener patrones complejos, como en el caso de (Lee J.-T. y otros, 2018). La Figura 28 muestra un esquema general para la obtención del criterio de búsqueda usado en un modelo CBIR, donde tanto las características visuales como los patrones de la composición son extraídos de la imagen para esta finalidad.

Otro campo donde se ha aplicado la composición es el de generación de imágenes. (Sims, 1991) desarrollaron un software que a través de variaciones en criterios compositivos obtenía imágenes nuevas. Con el avance de Internet, en la primera década del siglo XX, algunas investigaciones se centraron en la generación y valoración estética y de composición de los resultados por parte de los internautas. Esto posibilitaba una realimentación del generador y su posterior autoajuste, como en el modelo híbrido de

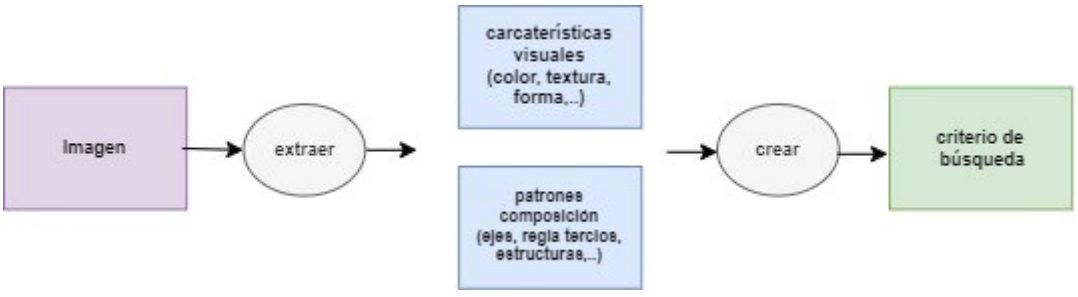


Figura 28 Obtención del criterio de búsqueda de un modelo CBIR.

(Draves, 2005) donde, por un lado, extraía las características visuales y patrones de la composición y, por el otro, incluía la valoración de humanos asociada.

Aunque todos estos enfoques se basan en la mejora de los métodos para extraer los patrones que se relacionan con la estructura de la composición, la complejidad para etiquetar las composiciones en imágenes, tanto por la subjetividad como por la dificultad técnica para realizar la tarea por parte de expertos, no ha facilitado tener unos *datasets* con los que entrenar los modelos (Lee J.-T. y otros, 2018). Esto dificulta la integración de la composición dentro de los modelos que usan *dataset* grandes, que son los usados por las nuevas tecnología de aprendizaje profundo, por ejemplo, para la generación de imágenes a partir de textos.

La Figura 29 resume en un esquema general la obtención de la composición a partir de la extracción de características visuales y patrones de composición independientemente con métodos distintos, como por ejemplo la detección de contornos con algoritmos de procesamiento de imagen o el uso de CNN (redes neuronales convolucionales) para extraer patrones de textura. De esta manera, a partir de un proceso de decisión que se puede implementar desde diversos enfoques (reglas semánticas, razonamiento bayesiano, redes neuronales artificiales, etc.) se obtiene una representación final de la composición. En (Lee J.-T. y otros, 2018) se llegan a plantear la extracción de hasta nueve criterios de composición como: la región de interés, centro, línea de horizonte, simetría, diagonal, curva, vertical, triángulo y patrón. Posteriormente, mediante una CNN se clasifica el tipo de composición.

Los distintos avances en visión artificial han mejorado la capacidad de los modelos que representan la composición de la imagen, sobre todo con la aplicación de las CNN (Santos y otros, 2021), tanto para la extracción de características visuales como para la de patrones de composición, pero no en conseguir un consenso sobre un modelo común. En este sentido, después de más de treinta años, sólo hay un esquema básico

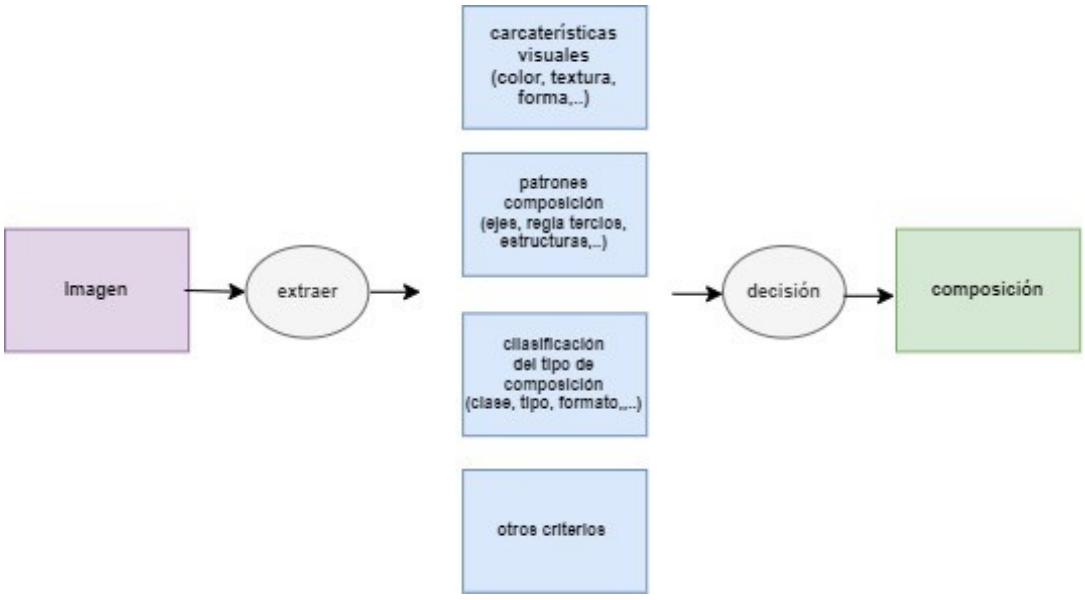


Figura 29 Sistema para la obtención de la composición de una imagen.

común a los distintos proyectos (Figura 29), pero variando la cantidad de características o patrones, la forma de extraerlos de la imagen y de procesarlos, y la representación de la composición final. A pesar de todas las dificultades, todos los investigadores están de acuerdo en la importancia de la composición para el procesamiento de imágenes artísticas y con criterios estéticos.

3 Metodología para la bioinspiración en visión artificial

A finales de la década de los setenta, David Marr era un neurocientífico interesado desde la inteligencia artificial en el sistema de percepción visual. Para Marr, la percepción visual era un conjunto de subprocesos con metas muy específicas y, por lo tanto, con teorías, representaciones e implementaciones diferentes entre sí (Marr, 1982). El reconocimiento facial, el seguimiento de objetos en movimiento, la detección de formas, texturas, colores, etc. son procesos específicos, diferenciados, tanto en la teoría como en la implementación, de un proceso global único que, como tal, dejó de tener importancia a partir de Marr, pero que recientemente ha vuelto a tener interés con el resurgimiento de la inteligencia artificial general. En los años 70, se sabía que el sistema de percepción visual de los mamíferos estaba altamente especializado, de tal manera que existían áreas de la corteza visual dedicadas a tareas muy concretas como plantea Crick (Crick, 1994). Parecía obvio, por otro lado, que no pudiendo generar un sistema de percepción similar al del ser humano, sí se podían dedicar esfuerzos a conseguir resolver tareas más concretas. Para poder avanzar en esa línea, era necesario tener una metodología más amplia que una descripción formal, como indican Jain y Binford (Jain & Binford, 1991), por lo que Marr definió una metodología establecida en tres niveles:

- Nivel de teoría de computación, centrado en la investigación y análisis del objetivo de la computación y la lógica para alcanzarlo.
- Nivel de representación y del algoritmo, enfocado en la implementación de la teoría con la representación de las entradas y salidas y el tipo de transformaciones que se llevan a cabo.
- Nivel de implementación física, que construye el hardware donde la representación y algoritmo se implementan.

Yiannis Aloimonos y David Shulman en 1990 en el libro «Integration of Visual Modules. An extension of the Marr Paradigm» (Aloimonos & Shulman, 1990) ahondarían aún más en la especialización funcional al introducir los módulos visuales. Este nuevo enfoque incluía un nivel intermedio entre la representación y el algoritmo y la implementación física, con el fin de dar robustez al sistema. En la reseña de la obra de Haim Levkowiz (Levkwoitz, 1990) podemos encontrar algunos problemas de este nuevo enfoque, sobre todo en la puesta en marcha del nivel intermedio propuesto, ya que Aloimonos y Shulman no describieron en su trabajo cómo implementarlo.

La bioinspiración desde Marr ha tenido siempre un papel relevante en la visión artificial, aunque no se han desarrollado metodologías más detalladas que faciliten el trabajo.

3.1 Descripción

Recientemente, el estudio de Medathati y colaboradores (Medathati y otros, 2016) ha analizado las principales cuestiones para tener en cuenta en modelos bioinspirados en visión artificial, siendo las siguientes:

- Análisis del problema en los tres niveles de David Marr, que ahonda en una teoría y en su implementación.
- Estudio de los circuitos neuronales relacionados con los comportamientos que se pretenden bioinspirar.
- Descubrimiento de los «trucos» que utiliza el cerebro para resolver tareas concretas, que suelen ser fruto de la evolución.
- Asociación de la conectividad neuronal con problemas computacionales.
- Comprobación de los modelos creados bioinspirados en relación con modelos naturales y artificiales.
- Comparación de los modelos basados en tareas versus los sistemas con propósitos globales.

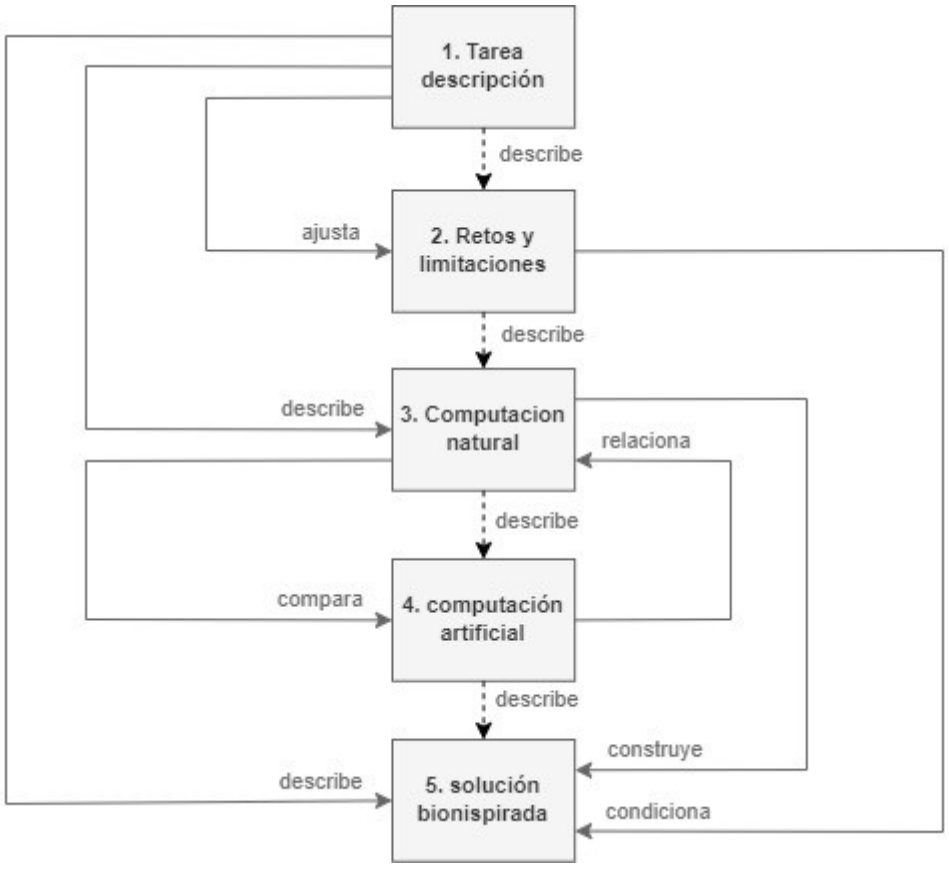


Figura 30 Esquema de la metodología para la bioinspiración.

Medathati y colaboradores plantean una metodología para la aplicación de la bioinspiración en visión artificial que dividen en cinco fases inicialmente secuenciales:

1. La definición de la tarea del procesamiento visual que se intenta computar.
2. La identificación de los retos principales que sintetizan las limitaciones físicas, algorítmicas o temporales y cómo repercuten en el tratamiento que debe realizarse de las imágenes o secuencias de imágenes.
3. La descripción de la solución biológica del problema para analizar la dinámica neuronal y los circuitos subyacentes a las soluciones biológicas haciendo hincapié en los elementos de computación canónicos que se están implementando en algunos modelos computacionales recientes.
4. La comparación con otras soluciones de visión artificial, para analizar algunos de los enfoques actuales de la visión artificial con la finalidad de esbozar sus límites y retos. Contrastar estos retos con los mecanismos conocidos en la visión biológica que permita prever qué aspectos son esenciales y cuáles no.
5. La determinación de las soluciones bioinspiradas a partir del análisis comparativo entre la visión artificial y la biología, para discutir los enfoques recientes de modelado en visión biológica y destacar las ideas novedosas que se consideren prometedoras para futuras investigaciones.

Entre las fases, existen relaciones no secuenciales que se representan en la Figura 30 y no se establecen en la metodología, pero se pueden inferir por los objetivos y la información previa que por lógica necesita. El ajuste de la tarea establece los retos y las limitaciones. La computacional natural parte de la descripción de la tarea, ya que su objetivo es localizar una tarea similar y analizar su funcionamiento. Además, la computacional natural localizada en relación con la tarea es comparada con las funcionalidades similares existentes en computación artificial (diferencias en cómo realizan la tarea). A su vez, se relacionan las computaciones artificiales que resuelvan alguna parte de la tarea o la tarea completa con las que se han localizado en la computación natural (puede que existan soluciones artificiales parciales o totales que puedan ser utilizadas junto a la computación natural o para resolver partes de la tarea). Por último, desde la computación natural se construye la solución con los retos y limitaciones identificados y con la descripción de la tarea.

3.2 Adaptación de la metodología

En esta tesis, no sólo se plantea una bioinspiración del sistema visual, sino que, además, se introducen las teorías de la psicología del arte, que se centran en el estudio del comportamiento del espectador ante las imágenes y cómo estas deben ser compuestas para conseguir un efecto concreto. Desde este enfoque, la metodología debe tratar las cuestiones neuronales y su relación con los hechos psicológicos. Esto debe ser más evidente en los puntos 3 y 5, ya que, por un lado, hay que analizar ambas disciplinas por separado y, por el otro, localizar la solución común. Por consiguiente, se plantean los

siguientes cambios en la metodología teniendo en cuenta que el objetivo está en las composiciones de las imágenes y no sólo en las imágenes:

1. Descripción de la tarea, del procesamiento visual y de la respuesta esperada en el espectador.
2. Los retos principales en relación con las limitaciones físicas, algorítmicas, temporales o psicológicas.
3. Solución biológica del problema en relación con el comportamiento psicológico.
4. Comparación con otras soluciones de visión artificial
5. Implementación de la solución a partir del análisis comparativo entre la visión artificial con las soluciones biológicas y psicológicas.

La principales diferencias están la fase primera y tercera, ya que, además de analizar la tarea desde el punto de vista de la computación natural y cómo esta tarea es resuelta, se relaciona con la respuesta psicológica esperada. Esta relación es una de las innovaciones que esta tesis plantea en la bioinspiración por las características computacionales del tejido interno de las composiciones. La Figura 31 muestra las relaciones entre las fases a partir del cambio introducido en la metodología para incluir la respuesta psicológica. En

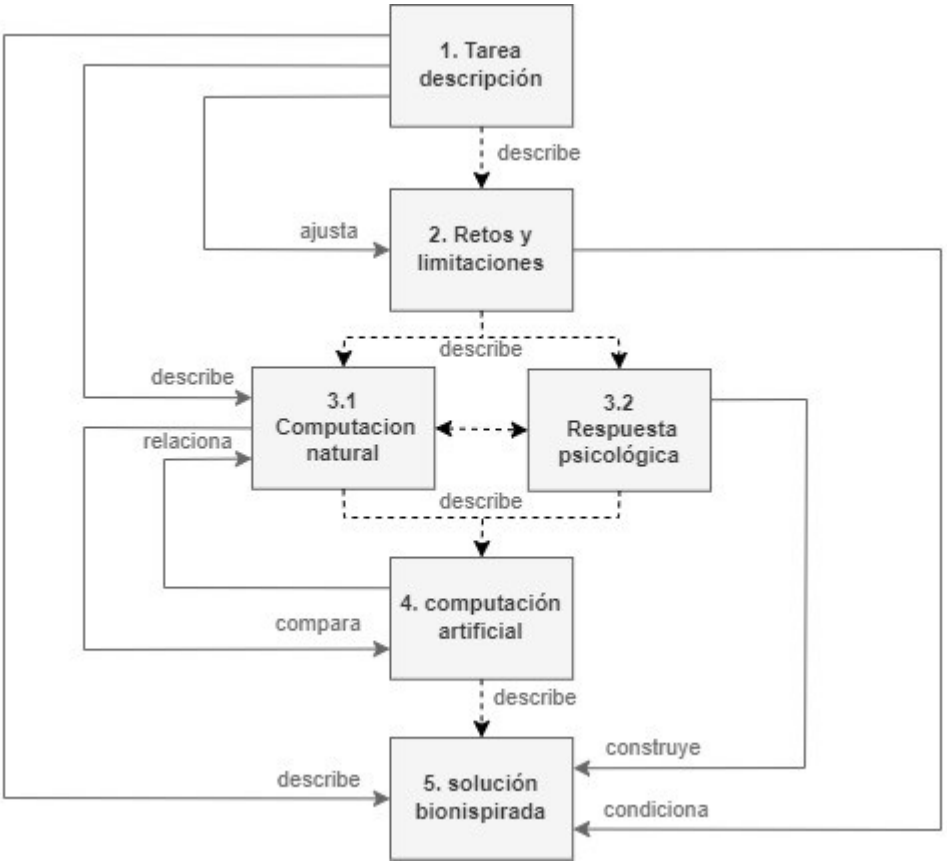


Figura 31 Esquema del ajuste propuesto en la metodología de bioinspiración.

el esquema, las principales diferencias son la relación bidireccional entre la computación natural y la respuesta psicológica en la fase tercera. Además, la salida de esta relación es desde la respuesta psicológica para la construcción de la solución bioinspirada.

4 Modelo Inner Fabric

La visión artificial tiene una conexión directa y bidireccional con la neurociencia y, en menor medida, con la psicología del arte. Esto no quiere decir que no exista, sólo que ha sido menos explorada. Cuando lo ha sido, se ha centrado en obtener resultados muy específicos como, por ejemplo, la evaluación de la estética de fotografías a través de algoritmos, como en el trabajo de Joshi y colaboradores (Joshi y otros, 2011) en relación con la estética y las emociones en las imágenes.

El tejido interno es una estructura subyacente a la composición, y el objetivo de esta tesis es describirlo e implementarlo con un modelo de visión artificial. Este es el principal objetivo, y en relación con él, se han recopilado las investigaciones y descubrimientos de la psicología de la percepción y del arte, de la neurociencia y de la visión artificial, que se pueden leer en el capítulo segundo. Antes de avanzar en el modelo, vamos a resumir las características más relevantes de la visión en relación con el tejido interno que se desprenden de estos tres enfoques.

En primer lugar, el concepto que se tiene sobre la imagen ha variado a lo largo de la historia de la humanidad: desde su utilidad como representación (retratos, escenas, simbología religiosa, etc.) hasta llegar a su autonomía como objeto visual (imágenes artísticas o estéticas por sí mismas). La relación entre la imagen y el sistema de percepción visual es estrecha, ya que la realidad visual en la que se encuentra la imagen condiciona su percepción. Como ejemplo, se ha descrito el sistema de percepción visual de la rana, condicionado por las funciones de cazar y huir en caso de peligro. El tipo de imágenes que una rana podría percibir está relacionado con esta funcionalidad, de ahí que la representación de su realidad visual sea un fondo blanco con los trazos en negro de los objetos en movimiento que se interponen entre la rana y la fuente de luz. Es decir, la realidad visual que es capaz de representar en su cerebro depende del hecho de que solo percibe el movimiento y no la quietud en su entorno. Para un ser humano, la realidad visual es tricromática debido a que su sistema de percepción se ha adaptado a objetivos más complejos. La imagen como objeto de una realidad visual es un concepto que condiciona el procesamiento que vamos a realizar, de ahí que se haya profundizado en las teorías sobre la imagen desde otras disciplinas para entender que la imagen es un objeto visual que se representa neuronalmente en el cerebro y se describe con características visuales a las cuales se asocian patrones y etiquetas semánticas.

En segundo lugar, la composición de la imagen, desde el punto de vista de la psicología del arte, ha avanzado hacia una sintaxis visual que permite tanto el análisis como la creación de imágenes. Esta sintaxis tiene aspectos para tener en cuenta en relación con las imágenes, su composición y su percepción. Por un lado, las regiones no son homogéneas; algunas producen mayor atracción en el espectador que otras. Por otro lado, existen características visuales que se extraen en la percepción y que se pueden procesar en módulos autónomos, como los contornos, formas, colores, texturas, etc.

El cerebro representa la información del espacio visual a través de los mapas retinotópicos, donde tampoco hay homogeneidad con la magnificación y la excentricidad. En neurociencia y psicología del arte coinciden las teorías y descubrimientos en el hecho de que existe una diferencia en la atracción entre la región central y la externa, entre la región inferior y la superior o entre la región izquierda y la derecha. Por consiguiente, los mapas retinotópicos son coincidentes con las teorías de la psicología del arte en la existencia de una estructura donde el centro de la imagen es el principal foco de atención.

Las características visuales y su procesamiento por separado de la psicología del arte tienen una relación con la modularidad del área visual del cerebro, que, junto con la cronoarquitectura propuesta por Zeki, profundiza aún más en la independencia de cada característica visual y su relación con el resto. Otro punto interesante es la interconexión entre las áreas de la corteza visual con el NGL, en especial la que existe con el área V4, especializada en formas y en el color. Esta conexión indica que la estructura de información del NGL se mantiene en el área V4, donde el color es procesado junto con la forma. Esto es relevante desde el punto de vista de esta tesis, ya que relaciona el procesamiento complejo de contornos, formas, texturas, colores o profundidad, que es relacionable con la percepción vertical de la composición, con uno más simple sin procesar, que es relacionable con la percepción horizontal del tejido interno.

En tercer lugar, el problema del salto semántico es de vital importancia en visión artificial, ya que el objetivo final del procesamiento de una imagen es obtener su contenido semántico, ya sea para clasificar, detectar objetos, patrones o describir el contenido de la imagen. De hecho, la idea de «tender un puente» para solucionarlo es bastante metafórica. En las CNN, que han tenido tanto éxito en la última década, también se emplea un «puente» para superar el salto semántico. Esto implica una fase de extracción de características seguida de otra de asociación para obtener patrones complejos conectados a la salida, donde se asignan etiquetas semánticas. El tejido interno, como estructura subyacente de la composición y más cercano a la estructura de píxeles de la imagen, también puede ser un «puente» para abordar el salto semántico, sobre todo en la sintaxis visual. Aunque apenas está definido y descrito como un problema, la visión artificial tiene mucho que aportar a la psicología del arte en este sentido.

En cuarto lugar, los mapas de prominencia son un modelo para la representación de la información visual en el procesamiento de imágenes relacionado con la atención. Las teorías sobre los movimientos oculares muestran su relación con la atención, tanto en la neurociencia como en la visión artificial. Los movimientos sacádicos, dirigidos por procesos tanto de abajo hacia arriba como de arriba hacia abajo, así como las fijaciones, son un tema relevante que se relaciona con algunos sistemas de escaneo de imágenes y su implementación para resolver problemas de visión artificial. En este sentido, los movimientos de ojos no son tan aleatorios como podría parecer, sino que dependen tanto de la imagen (procesamiento de abajo hacia arriba) como de los objetivos del sistema de percepción (procesamiento de arriba hacia abajo).

Finalmente, la estructura subyacente en la composición, denominada tejido interno, se relaciona con una percepción horizontal, sin detalles y desenfocada, de tipo inconsciente, en contraposición con una percepción vertical, detallada y enfocada en regiones

concretas, de tipo consciente. Esto plantea un importante desafío para la visualización por parte del ser humano, ya que se trata de una percepción de tipo inconsciente. En este sentido, la visión artificial se presenta como una herramienta ideal para representar el tejido interno como un mapa de prominencia que ayude al ser humano a visualizarlo. Y simultáneamente, teniendo en cuenta la capacidad del tejido interno de describir las regiones de interés que captan la atención del espectador de manera inconsciente, éste puede usarse en visión artificial para el procesamiento de imágenes.

La sintaxis de la imagen constituye actualmente la descripción formal más avanzada para representar la composición de las imágenes creadas por humanos, y aunque se centra principalmente en las artísticas, se puede extender a cualquier otro tipo de imagen. La metodología para la bioinspiración en visión artificial presentada en el capítulo anterior (ver Figura 31) facilita la descripción de soluciones tanto para tareas generales como para las más específicas. En este sentido, se aplicará primero a la tarea principal, que es la obtención del mapa del tejido interno de la composición a partir de la imagen, y luego a las distintas subtareas necesarias.

4.1 Representación del tejido interno como mapa de prominencia

A. Descripción de la tarea, del procesamiento visual y de la respuesta esperada en el espectador

El tejido interno, descrito por Ehrenzweig (Ehrenzweig, 1967), representa la estructura subyacente de la composición de una imagen artística, aunque esta descripción es aplicable a otros tipos de imágenes creadas por seres humanos, así como a imágenes generadas artificialmente gracias al avance de los sistemas inteligentes. La tarea consiste en generar un mapa de prominencia que represente este "tejido interno" a partir de una imagen, lo que permite su análisis por parte de un experto y también su aplicación en el procesamiento de imágenes.

B. Los retos principales en relación con las limitaciones físicas, algorítmicas, temporales o psicológicas e las que dependen las composiciones de las imágenes

El principal desafío consiste en construir un modelo de visión artificial que automatice un proceso natural, cuyo efecto psicológico conocemos, pero del cual desconocemos su proceso biológico subyacente. En este sentido, la limitación principal radica en la identificación de la localización específica en el cerebro, o áreas, relacionadas con el tejido interno y su representación neuronal.

La relación entre la forma y la función es uno de los desafíos al diseñar el modelo, ya que muchas de las características de la composición dependen de relaciones topográficas, como la existencia de dos puntos de atención opuestos (el centro geométrico de la imagen y el espacio exterior que engloba las cuatro esquinas) o la preponderancia de

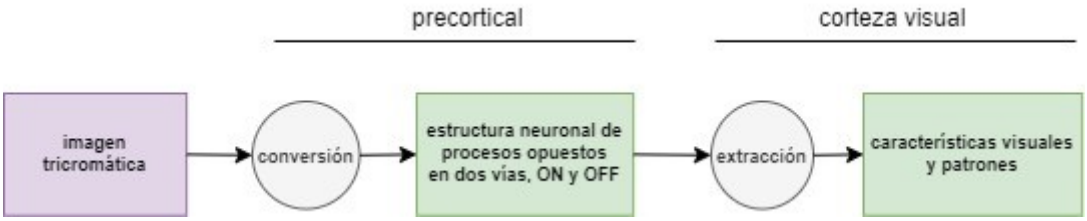


Figura 32 Áreas del procesamiento de la información visual en el cerebro.

ciertas regiones de la imagen sobre otras. Estos condicionantes, que son analizados tanto desde la psicología del arte como presentes en los mapas retinotópicos que representan la información visual en el cerebro, exigen la introducción de algoritmos que procesen y representen simultáneamente, estableciendo así una conexión entre arquitectura y función.

C. Solución biológica del problema en relación con el comportamiento psicológico

En el sistema de percepción visual, existen dos fases secuenciales en el procesamiento por su arquitectura y por su funcionalidad (ver Figura 32). La primera, en el área precortical, está relacionada con el proceso de captar y transformar la señal de las ondas lumínicas por la retina en una estructura neuronal de vías separadas y paralelas. La segunda, en la corteza visual, está relacionada con las operaciones de extracción de características visuales y patrones, donde se detectan contornos, formas, colores, texturas, etc. La estructura subyacente, analizada desde la psicología del arte y tal como la plantea Ehrenzweig (Ehrenzweig, 1967), es una representación inconsciente y global de la composición, donde se relacionan las regiones de la imagen entre sí sin detalles. Si esta estructura subyacente debe estar procesada por una percepción visual horizontal, descrita como global, desenfocada, polifónica, conjuntiva o intuitiva, sólo podría emerger en la primera fase, ya que, en la segunda, las operaciones de abstracción para extraer características visuales se relacionarían con una percepción vertical: local, monofónica, focal o disyuntiva.

El NGL es un área cuyo estudio y análisis ha evolucionado en el tiempo, de tal manera que la neurociencia ha pasado de pensar que era un órgano sin papel relevante en el procesamiento a tratarlo como un controlador de los datos que acceden a la corteza visual (Teller, 2014). La principal pregunta es si esta área podría ser la primera, dentro del proceso de percepción visual, en representar una estructura subyacente en relación con la composición de la imagen. Tanto las microconsciencias como la cronoarquitectura (Zeki, 2005) (ver el apartado 2.2.2, Anatomía y fisiología del sistema de percepción visual) llevan a concluir, por la modularidad y la ausencia de una única consciencia, que lo más evidente es que esta estructura subyacente esté distribuida en las distintas áreas visuales y no sea exclusiva de una en concreto. Pero si el NGL es una puerta de entrada, debe contener, en esta fase temprana, toda la información disponible y, por consiguiente, su representación de la información debería contener la estructura subyacente sin estar distribuida como sucede en la corteza visual (en sus distintas áreas).

Sin embargo, el NGL genera, en cada momento, una representación global de la imagen desde una región concreta del espacio visual donde se fija la mirada. Es decir, el conjunto de todas las representaciones obtenidas por el NGL durante un proceso completo de escaneo de la imagen facilitaría, siguiendo la metodología de los mapas de prominencia, la obtención de un mapa compilatorio final. Sin embargo, la estructura subyacente no puede ser un lote de «fotos fijas» de cada una de las regiones de la imagen como si fuera un mosaico, sino que debe tener una estructura global que represente a todo el espacio visual integrado y relacionado entre sí.

Por otro lado, los esquemas de referencia son utilizados para representar las relaciones topográficas tanto de los objetos que se encuentran en el espacio visual como del propio sujeto o partes de él (brazos, piernas, la cabeza, etc.). La imagen es un objeto en el espacio visual y, por consiguiente, se encuentra representada en un esquema de referencia en relación con el sujeto, pero también existe un esquema de referencia de la propia imagen donde sus regiones están relacionadas. Por consiguiente, este esquema representa las relaciones entre las regiones de una imagen obtenidas en cada escaneo visual y, por lo tanto, podemos considerar que es una representación global de la imagen.

D. Comparación con otras soluciones de visión artificial

Históricamente, el análisis de la composición de una imagen ha sido abordado desde dos puntos de vista a veces contradictorios: desde la imagen misma o desde la perspectiva del espectador. La psicología del arte, basándose en los principios de la Gestalt (Kofka, 1955) y especialmente con los avances de Arnheim (Arnheim, 1956) y Dondis (Dondis, 1974), se ha centrado principalmente en el primer punto de vista. En este enfoque, el análisis se centra en los elementos presentes en la imagen, en sus características visuales y en cómo se relacionan entre sí. En este contexto, se representa la composición mediante esquemas que utilizan ejes o regiones para indicar los elementos más significativos y las fuerzas que generan debido a su posición o a sus relaciones con otros elementos. A partir de esta información, es posible determinar una estructura subyacente, aunque no necesariamente siguiendo la línea propuesta por el tejido interno, ya que este enfoque se centra en resaltar los elementos prominentes en lugar de explorar una percepción horizontal.

La neuroestética (Berlyne, 1973), por otro lado, se ha centrado en el segundo punto de vista, es decir, en el espectador y el creador, mostrando un mayor interés por los efectos que una imagen provoca en el sistema de percepción. Esto incluye el análisis de qué áreas del cerebro se activan y qué regiones de los mapas retinotópicos se estimulan en respuesta a la visualización de una imagen. Este enfoque pone más énfasis en el impacto de la imagen en el observador que en la propia imagen y sus características visuales, lo que refleja el interés de los investigadores en comprender la representación neuronal de la imagen para comprender su composición.

En el campo de la visión artificial, es poco común investigar problemas relacionados con la visión temprana que preceden a la detección de contornos, formas, colores, texturas, etc. Incluso en las etapas de preprocesamiento, rara vez se abordan tareas que puedan acercarse a la obtención de una estructura subyacente.

Figura 33 Esquema modelo del mapa de prominencia de Koch y Ullman.

E. Diseño de la solución.

Los mapas de prominencia representan la atracción de la atención del espectador a través de los elementos visuales (Koch & Ullman, 1987). La estructura subyacente representa asimismo la atracción de la atención del espectador a través de una percepción horizontal. En ambos casos, existe una relación entre la capacidad de atraer (agudización) y la prominencia, donde las regiones de la imagen «compiten» por la atención del espectador, y donde, desde el punto de vista de Dondis en su sintaxis visual (Dondis, 1974), el espectador busca el equilibrio constantemente (nivelación). Por otro lado, el movimiento sacádico de los ojos en el espacio visual no deja de ser una búsqueda del equilibrio a través de las regiones que «compiten» por conseguir la atención. Es fácil ver en la estructura subyacente, por un lado, la competición por la atención y, por el otro, la relación entre la «búsqueda del equilibrio» y un proceso de escaneo de la imagen, es decir, la agudización y la nivelación que plantea Dondis. En este sentido, el tejido interno es viable definirlo como un mapa de prominencia de la estructura subyacente a partir del escaneo del espacio visual en un proceso de agudización y nivelación.

Los mapas de prominencia han tenido una importante evolución en el análisis de las imágenes a partir la teoría de la atención de Koch y Ullman (Koch & Ullman, 1987) desde la cual distintos investigadores han desarrollado modelos bioinspirados. Estos modelos son cada vez más sofisticados, como los descritos y analizados por Boeji y Itti (Borji & Itti, 2013) variando el método para obtener las prominencias. Por lo tanto, implementar un mapa de prominencia del tejido interno de la imagen, a través de un simulador bioinspirado del movimiento de ojos, es una solución que se ajusta a la percepción horizontal.

Los mapas de prominencia se construyen en dos fases secuenciales: la primera para generar los mapas de características y la segunda para compilarlos en el mapa de prominencia final (ver Figura 33). Con el enfoque propuesto para el tejido interno y su representación en el área precortical de la visión, el mapa de prominencia que lo representa se obtendría a partir del escaneo del espacio visual, implementando los mapas retinotópicos y funcionalidades de la retina y NGL. El escaneo analizaría ordenadamente cada región de la imagen con un proceso de agudización y nivelación y no a través de la compilación de mapas de características de la imagen completa (ver Figura 34). De esta manera, los mapas de características pasan a ser mapas retinotópicos de las regiones de la imagen (retina y NGL) y el mapa de prominencia a un esquema de referencia que se construye paso a paso y no con la compilación de los mapas de características. La agudización se relaciona con la selección de la siguiente región a escanear a partir del mapa de pesos visuales, y la nivelación calcula la prominencia de la región al equilibrar

su valor de actividad neuronal con el resto de las regiones. Por lo tanto, el proceso itera escaneado las regiones que se van agudizando y los valores de prominencia se cargan a partir de la nivelación.

Para construir el modelo es necesario describir las siguientes subtareas en relación con el esquema de la Figura 34:

- Representar el espacio visual de cada región como mapa retinotópico.
- Crear el mapa de prominencia como un esquema de referencia.
- Calcular los pesos visuales.
- Agudizar y nivelar.
- Escanear el espacio visual.

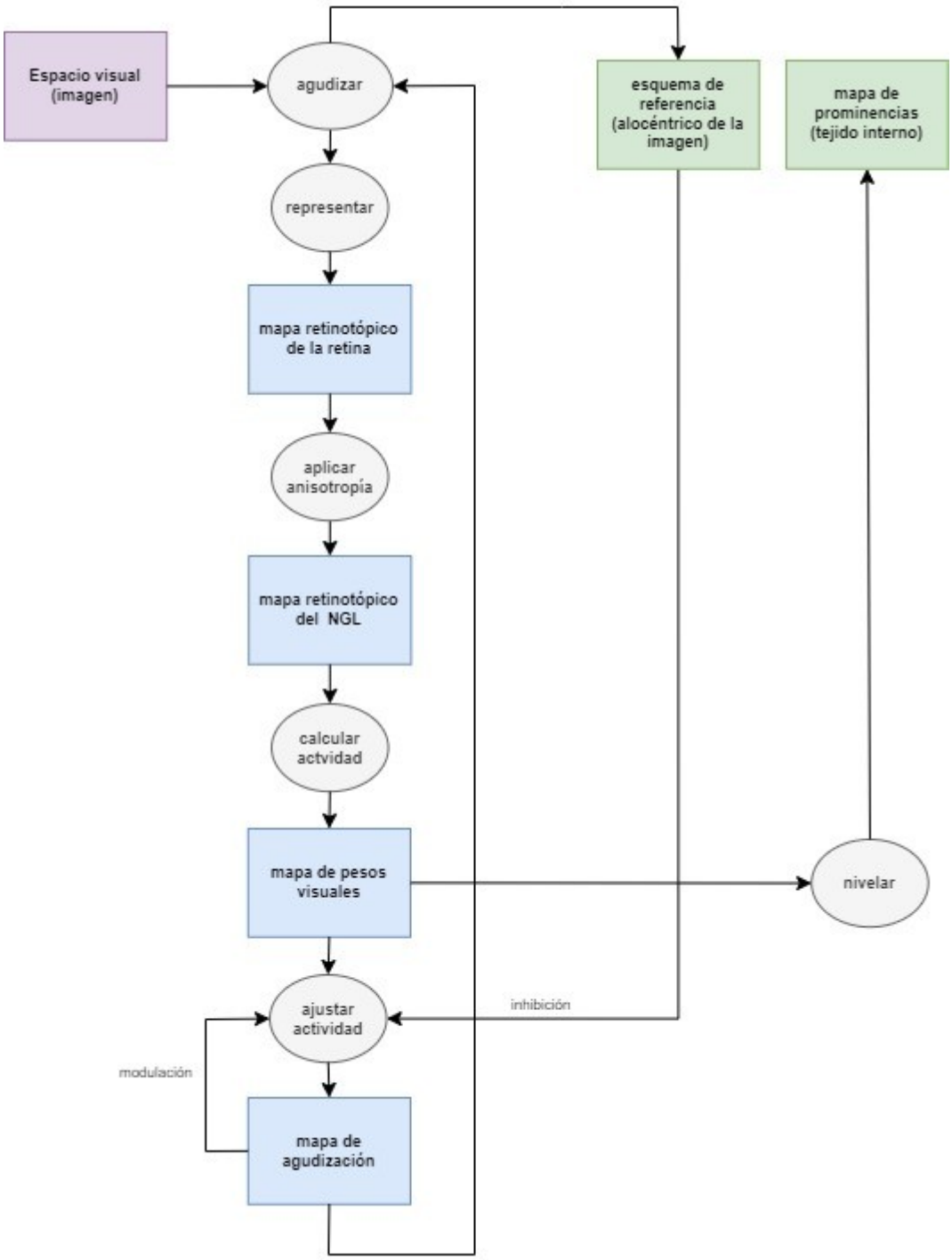


Figura 34 Esquema funcional del modelo Inner Fabric.

4.2 Subtareas del modelo

4.2.1 Representación del espacio visual al focalizar la atención en una región de la imagen

A. Descripción de la tarea, del procesamiento visual y de la respuesta esperada en el espectador

La tarea tiene como objetivo la transformación del espacio visual, donde está la imagen, en una representación retinotópica de la retina y NGL. Al agudizar, se focaliza la atención en una región concreta de la imagen y se deforma la percepción del espacio visual, prestando más atención a la región cercana. El mapa retinotópico es una manera elegante y sencilla de producir esa deformación e implementar el poder de atracción del centro, así como las relaciones anisotrópicas centro-exterior, abajo-arriba, izquierda-derecha y horizontal-vertical, descritas desde la psicología del arte y la neurociencia.

B. Los retos principales en relación con las limitaciones físicas, algorítmicas, temporales o psicológicas e las que dependen las composiciones de las imágenes

El principal reto es representar la imagen compuesta por píxeles de un sistema de coordenadas cartesianas en uno con excentricidad y campos receptivos. Una limitación importante para tener en cuenta es la diferencia de resolución entre las regiones de la imagen dependiendo de su excentricidad y también del coste de computación de estas diferencias de tratamiento o procesamiento. Es decir, la región central necesita una resolución mucho mayor que el borde de la imagen, pero variar la resolución es un proceso con un alto coste computacional pero que a la larga se amortiza.

Por otro lado, una limitación importante es que la preponderancia de la región inferior izquierda implica que la parte superior derecha tenga una menor cantidad de campos receptivos, pero que esta cantidad sea menor en el exterior que el centro, y que, incluso, la región central superior derecha tenga más cantidad de campos receptivos que la región exterior inferior izquierda.

C. Solución biológica del problema en relación con el comportamiento psicológico

Rudlof Arnheim introdujo la idea de la existencia de una estructura topográfica de la imagen, denominada marco estructural, a partir de la aplicación de la Teoría de la Gestalt para analizar las composiciones de obras artísticas (Arnheim, 1983). En esta estructura topográfica existen dos focos de atracción: por un lado, el centro geométrico, como inicio y origen de cualquier relación dentro de la composición y, por otro lado, las cuatro esquinas de la imagen que actúan como contrapeso para la generación del dinamismo de la composición.

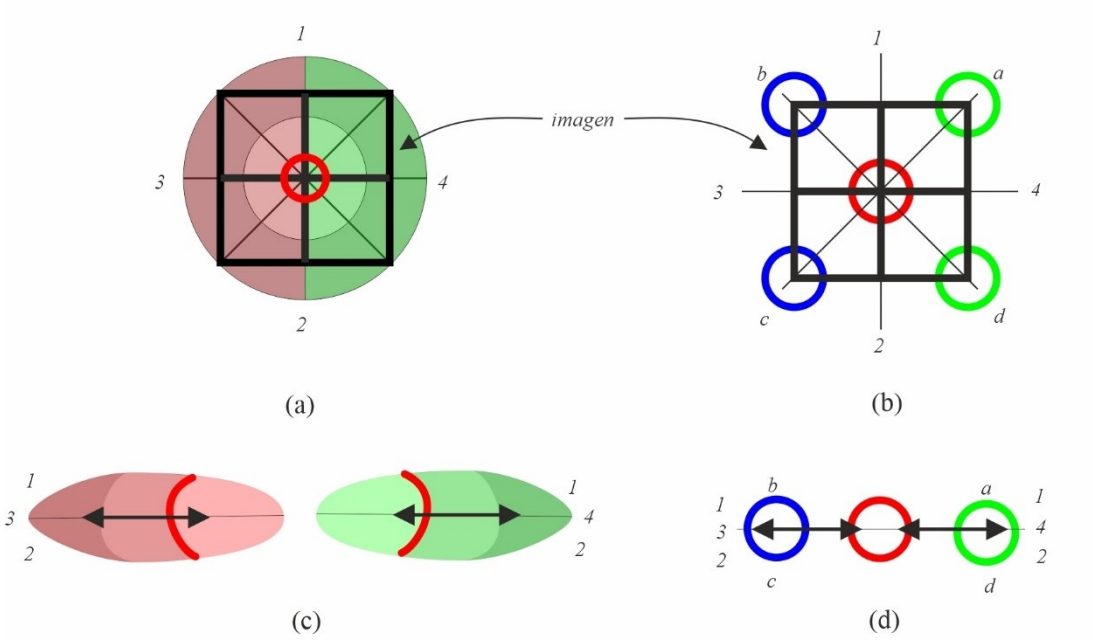


Figura 35 Marco estructural de la composición y los mapas retinotópicos.
Nota: (a) espacio visual; b) el marco estructural en una imagen (c) la dinámica centro-exterior de los mapas retinotópicos.; y (d) la dinámica de fuerzas centro-exterior del marco.

Desde la neurociencia, se han descrito los mapas retinotópicos como arquitecturas neuronales que representan al espacio visual utilizando campos receptivos a partir de una función de excentricidad (Van Essen y otros, 1984) (Wandell & Winawer, 2011). La neurociencia ha estudiado estas arquitecturas, concluyendo que son comunes en las distintas áreas del cerebro implicadas en la percepción visual, pero también en otras que no lo están. Y, además, han comprobado que la excentricidad no es igual en todas las áreas, en la retina es menor, pero en el NGL o en el área V1 es mayor, determinando un proceso que se denomina magnificación central.

Existe una relación entre la topografía de la imagen planteada por Arnheim (Arnheim, 1983) y la arquitectura del mapa retinotópico descubierta por Holmes (Holmes, 1945). El dinamismo descrito por Arnheim, en relación con dos fuerzas de atracción (una central y otra externa), es posible relacionarlo con la excentricidad de los mapas retinotópicos. Es decir, la propia estructura excéntrica de los mapas retinotópicos establece los dos focos de atracción: uno central y otro externo. Para ilustrar esta idea, la Figura 35 muestra la relación centro-exterior del espacio visual (Figura 35.a) y del marco estructural propuesto por Arnheim (Figura 35.b) y a su vez izquierda-derecha y horizontal-vertical. En los mapas retinotópicos, existe un eje centro-exterior, tanto hacia la izquierda como hacia la derecha (Figura 35.c). En el marco estructural existen dos ejes diagonales del centro hacia las cuatro esquinas, que, si son plegados en el eje horizontal como los mapas retinotópicos, se simplifican las fuerzas del centro-exterior de una manera similar.

En el NGL, además, se produce una magnificación que es inversamente proporcional a la excentricidad según los análisis realizados (Horton & Hoyt, 1991) (Schneider y otros,

2004). Existe además un efecto denominado anisotropía horizontal-vertical (Corbett J. E., 2011). Este efecto provoca que existan en los mapas una mayor cantidad de campos receptivos en la dirección del eje horizontal que en el vertical, provocando una reducción mayor del espacio visual representado.

Por otro lado, desde la sintaxis de la imagen de Dondis (Dondis, 1974), a la relación centro-exterior se añade una preponderancia por la región inferior izquierda. Los estudios realizados sobre la cantidad de campos receptivos que representan en los mapas retinotópicos del NGL las áreas del espacio visual demostraron que había diferencias en cuanto a la cantidad según la región, siendo mayor en el mapa derecho, que representa a la región izquierda del espacio visual, y en la región inferior en relación con la superior. Por consiguiente, esta preponderancia por la región inferior izquierda es posible definirla funcionalmente a través de una relación entre una mayor cantidad de campos receptivos que facilitan una mayor atracción en los ejes izquierda-derecha e inferior-superior.

En relación con la información que los campos receptivos procesan, existen varias transformaciones en la retina y NGL. La información visual es captada en la retina en un sistema tricromático y es convertida a un sistema de procesos de colores opuestos a través de una arquitectura neuronal de vías paralelas y canales separados (Hubel, 1995). La arquitectura de las células ganglionares se compone de dos vías paralelas ON y OFF (Schiller y otros, 1986) (Hubel, 1988), la primera representando la intensidad de la señal, y la segunda, la ausencia de intensidad. Cada canal representa la señal captada por un campo receptivo de un rango de longitud de onda diferente, en concreto cuatro correspondientes a tres tipos de neurona (descrito en detalle en 2.2.3 Retina).

Desde la psicología del arte, se han creado sistemas para relacionar los colores y describirlos como el de Munsell (Munsell, 1915) basado en la esfera de color de Runge (Runge, 2010), donde cada color se define a través de sus propiedades de matiz, luminosidad y saturación. Una de las cuestiones importantes analizadas es que existen relaciones opuestas y complementarias entre los colores (Pridmore, 2008), (Manzotti, 2017), (Zeki y otros, 2017) que sólo son posibles con una estructura donde los colores primarios se relacionan en los ejes: rojo-verde, y amarillo-azul. Esta estructura de procesos de colores opuestos está relacionada con la arquitectura de vías paralelas y canales segregados, donde onda larga equivale al rojo, onda media equivale al verde, onda corta equivale al azul y onda larga y media equivale al amarillo.

D. Comparación con otras soluciones de visión artificial

Algunos proyectos de visión artificial han usado mapas retinotópicos para representar la información. En 1993, Blackburn (Blackburn, 1993) desarrolló un modelo para robótica que usaba un mapa logarítmico (log-map) para construir los campos receptivos de la retina. Más recientemente, los RESOM (Retinopic Self-organization Maps) son variantes de los SOM (Self-Organizing Maps) utilizados por Ramirez-Quintana y colaboradores (Ramirez-Quintana y otros, 2018) para simular áreas de la corteza visual.

Por otro lado, la convolución ha sido ampliamente utilizada en visión artificial para representar campos receptivos. Existen varias formas de construir modelos matemáticos de los campos receptivos como los planteados por Soodak (Soodak, 1986) incluso modelando las unidades de la retina de Dacey y colaboradores (Dacey y otros, 2000), donde la «diferencia de gaussianas» (DoG) (Einevoll, 2003) es la más utilizada para representar la funcionalidad de los campos receptivos.

El espacio RGB es el más utilizado en visión artificial, pero existen investigaciones y proyectos que han implementado sistemas basados en procesos de colores opuestos como el modelo oRGB de Bratkova y colaboradores (Bratkova y otros, 2009) que genera tres ejes: rojo-verde, amarillo-azul y blanco y negro, en un espacio cercano a CIE LAB o en el sistema para localizar señales de tráfico de Mathibela y colaboradores (Mathibela y otros, 2013). En todos estos casos, se realizan conversiones desde RGB aplicando pesos para generar los ejes, con un sistema de conversión parecido al de CIE.

E. Diseño de la solución

Los mapas retinotópicos representan el espacio visual donde se encuentra la imagen, de tal manera que alrededor de ella hay un espacio que también hay que representar para facilitar el procesamiento de los bordes de la imagen con su entorno. Por consiguiente, la imagen se encuentra en el espacio visual representado en coordenadas polares como $V_{r,\theta,c}$ donde r y θ son las coordenadas polares y $c \in \{r, g, b\}$ son los canales RGB de los píxeles. La posición en coordenadas cartesianas del espacio visual $V_{x,y,c}$ se relaciona con las coordenadas polares con las siguientes ecuaciones:

$$x = (W/2) - \mathrm{E(j)} * \sin\theta$$
$$y = (W/2) + \mathrm{E(j)} * cos\theta \quad (1)$$

siendo W el r máximo del espacio visual, que equivale a un ángulo de 90º del campo visual (45 º para cada ojo), y $\mathrm{E(j)}$ la función de excentricidad que veremos más adelante. El campo de visión es de unos 180º pero, en la contemplación de una imagen, 90º es el campo visual donde podemos detectar formas y es el que usamos como límite para el espacio visual. El ángulo del campo visual donde la visión detecta el color es aproximadamente de 60º, y usamos ese ángulo para situar la imagen (ver Figura 36). Por consiguiente, W_{imagen} es el ancho en una imagen apaisada o el alto en una vertical correspondiente a un ángulo visual de 60º en un espacio visual de 90º donde:

$$W_{imagen} = \frac{2}{3}W \quad (2)$$

El espacio visual del entorno de la imagen debe tener valores neutros, por ejemplo, la media de intensidad global de la imagen en su histograma u otras funciones existentes en la literatura para esta cuestión. Este espacio no se escanea, pero es importante para que los mapas retinotópicos tengan valores en las regiones de los bordes.

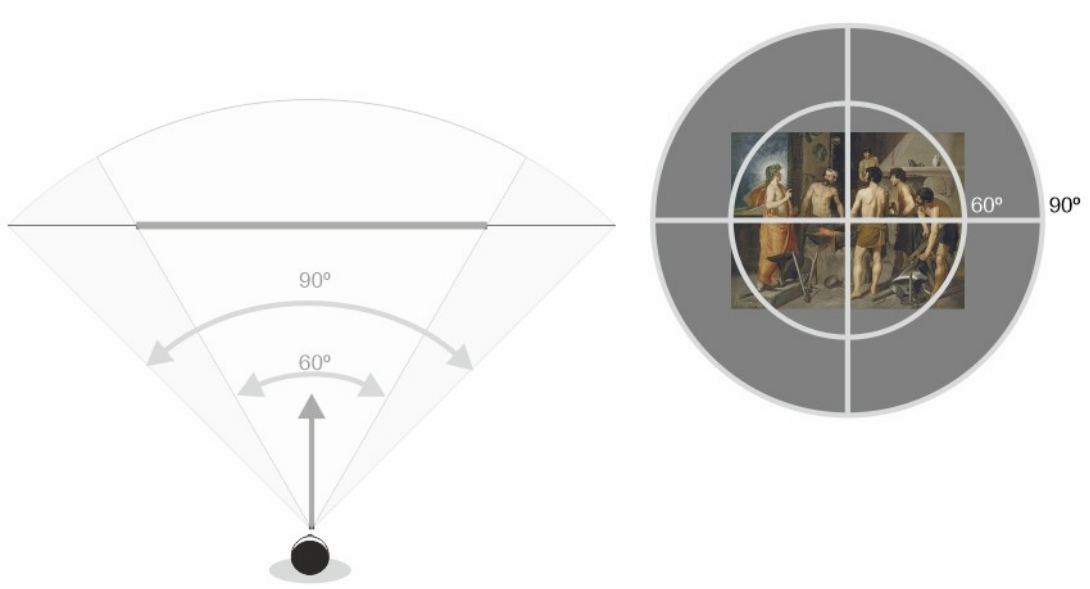


Figura 36 Campo visual.
Nota: El ángulo del campo visual es de 90º y la imagen se sitúa en un ángulo de 60º que es donde se discrimina el color.

La arquitectura del mapa retinotópico de la retina está formada por campos receptivos que representan regiones del espacio visual según la excentricidad, mientras que la del NGL lo hace, además, aplicando la magnificación y anisotropía sobre el mapa de la retina. Es decir, la entrada de la retina es el espacio visual, mientras que la del NGL es la de la retina. Además de la representación retinotópica se produce una transformación del espacio visual en píxeles RGB a dos vías, ON y OFF, y cuatro canales por vía en un sistema de colores opuestos. Esto indica que hay ocho campos receptivos por cada posición en el mapa retinotópico.

Retina

El modelo mapa retinotópico usado por Blackburn (Blackburn, 1993) utiliza un mapa logarítmico (log-map) desde el espacio visual $V_{r,\theta,c}$ donde las columnas j del mapa retinotópico son las distancias r seleccionadas por la función de excentricidad y el ángulo θ las filas i en una representación log-polar. Por lo tanto, el mapa logarítmico de la retina viene defino como:

$$R^{h}_{i,j,v,o,t}, \tag{3}$$

donde $h \in \{izquierda, derecha\}$ —mapa del hemisferio derecho o el del izquierdo— siendo el total de j columnas N_{retina} y $\frac{N_{retina}}{2}$ el de i filas en cada h, las vías paralelas $v \in \{ON, OFF\}$, y los canales $o \in \{L, M, S, LM\}$ para $v = ON$ y $o \in \{-L, -M, -S, -LM\}$ para $v = OFF$ (los canales se definen en la sección de diseño de la solución de 4.2.3 Calcular los pesos visuales). Por último, t es el intervalo

de escaneo. Aunque hay dos mapas (izquierda y derecha) nos referiremos a ellos conjuntamente. Para localizar la posición en el espacio visual de un campo receptivo de la columna j del mapa retinotópico, aplicamos la función de excentricidad $E(j)$ utilizada por (Blackburn, 1993):

$$\mathrm{r} = \mathrm{E(j)} = e^{log\left(\frac{W}{2}\right)*\left(\frac{j}{N_{retina}}\right)} \tag{4}$$

La posición de la fila i se obtiene con el ángulo θ_i:

$$\theta_i = \frac{i * 2\pi}{N_{retina}} \tag{5}$$

y cada h en relación con θ_i es:

$$h = \begin{cases} dcha, & \theta_i \leq \frac{\pi}{2} \quad o \quad \frac{3\pi}{4} < \theta_i \leq 2\pi \\ izq, & \frac{\pi}{2} < \theta_i \leq \frac{3\pi}{4} \end{cases} \tag{6}$$

La Figura 37 muestra la relación entre el espacio visual en coordenadas polares y los mapas retinotópicos como representación de mapas logarítmicos. Cada mapa, independiente, representa a una mitad del espacio visual (derecha e izquierda), usando las filas para representar las posiciones de θ_i y las columnas las posiciones en la distancia r del espacio visual según la función de excentricidad. Los mapas logarítmicos se suelen representar en vertical, pero en este trabajo, para facilitar la visualización y con la existencia de dos mapas representando a la mitad izquierda y derecha, aplicaremos una representación horizontal. La representación de la Figura 37 aplica $N_{retina} = 8$ para que el ejemplo permita ver el efecto de excentricidad con claridad. La representación de la región central ocupa tres cuartos de los campos receptivos disponibles (las columnas

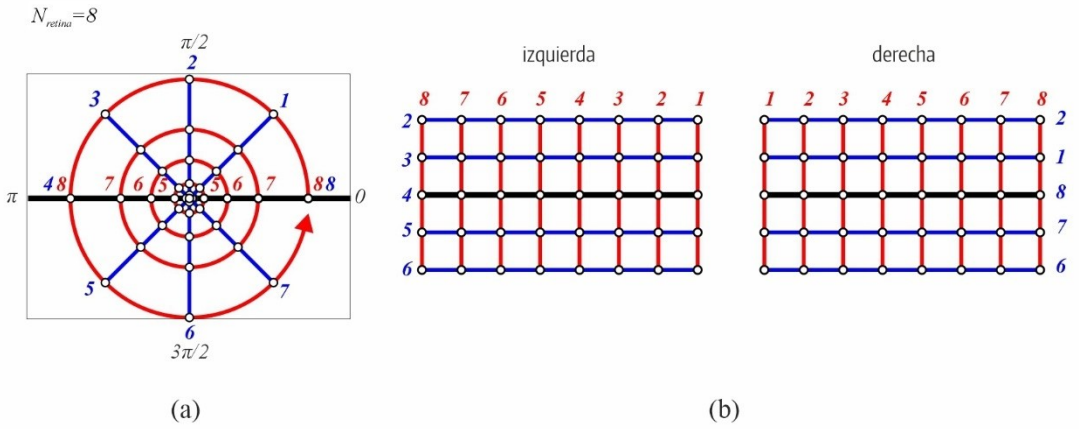


Figura 37 Coordenadas polares del espacio visual y mapas retinotópicos logarítmicos.
Nota: (a) espacio visual en coordenadas polares; y (b) mapas retinotópicos logarítmicos (log-map).

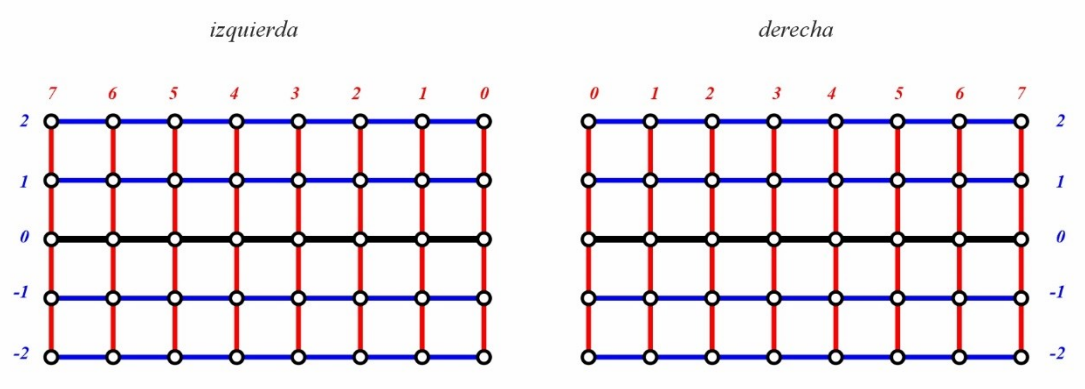


Figura 38 Modificación de las coordenadas de los mapas logarítmicos.
Nota: la finalidad es facilitar las futuras tareas entre las regiones usando coordenadas simétricas. El eje vertical y el horizontal son 0, con el punto central en [0,0], la parte superior se representa en positivo y la inferior en negativo.

de los mapas desde $j = 1$ hasta $j = 5$). El eje horizontal del centro del espacio visual está representado con la fila de la mitad de ambos mapas, mientras que el eje vertical se representa con la primera fila y la última, ambos según la mitad del espacio visual de cada mapa.

Para facilitar las operaciones futuras con ambos mapas, se modifica el índice i para que se encuentre en ambos mapas en el intervalo $\left[\frac{-N_{retina}}{4}, \frac{N_{retina}}{4}\right]$, siendo la región inferior negativa, el eje central horizontal (meridiano horizontal central) igual a 0, y la superior positiva. Las columnas se ordenan desde el 0, para que ambos mapas mantengan una estructura simétrica y homogénea con el fin de que los ejes tengan índices igual a 0 en el punto central. Este criterio también se aplica a las columnas, comenzado en 0. La Figura 37, muestra está reconfiguración de los índices en los mapas logarítmicos de la Figura 38.b.

La Figura 39 muestra cómo se relaciona el espacio visual en coordenadas polares (Figura 39.a) con la arquitectura de la representación logarítmica de los mapas retinotópicos (izquierda y derecha) (Figura 39.b). Para visualizar mejor la relación, se ha incluido en la Figura 39.c una imagen (círculo cían) situado en el espacio visual (circulo blanco) donde el círculo verde delimita aproximadamente la región central y el rojo su entorno más cercano. Además de la representación en el mapa logarítmico, donde se visualiza con claridad el efecto de excentricidad, se incluye una representación en coordenadas polares de ambos mapas (izquierdo y derecho) sin separación. La fragua y los dos brazos de ambos hombres ocupan la región central, y los tres cuartos de ambos mapas. Sin embargo, el resto de los círculos, rojos y cian, aparecen comprimidos en un espacio menor.

Cada campo receptivo representa una región del espacio visual. El radio del campo receptivo depende de la función de excentricidad y se obtiene para cada columna j siendo igual para todas las filas i de esa columna. La ecuación es la siguiente:

$$CR_j = 2 * \left(1 - cos\left(\frac{2 * \pi}{N_{retina}}\right)\right) * E(j) \quad (7)$$

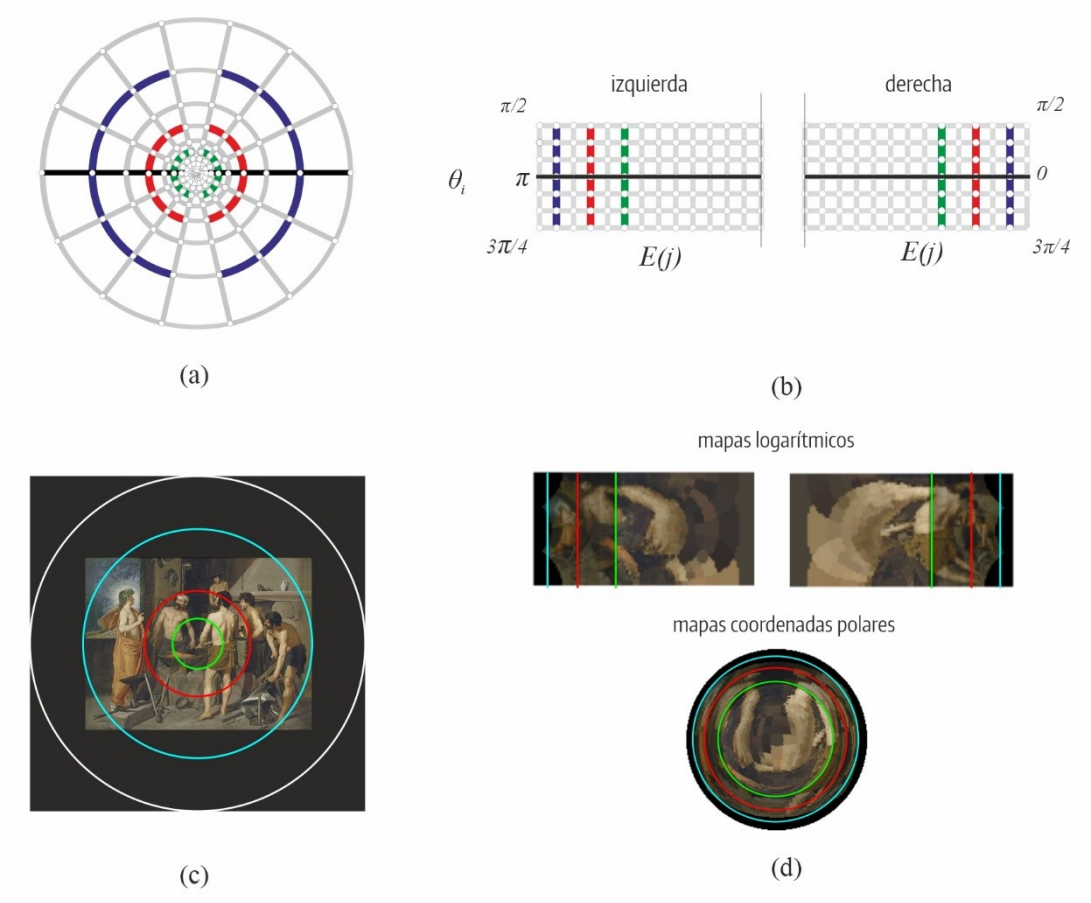
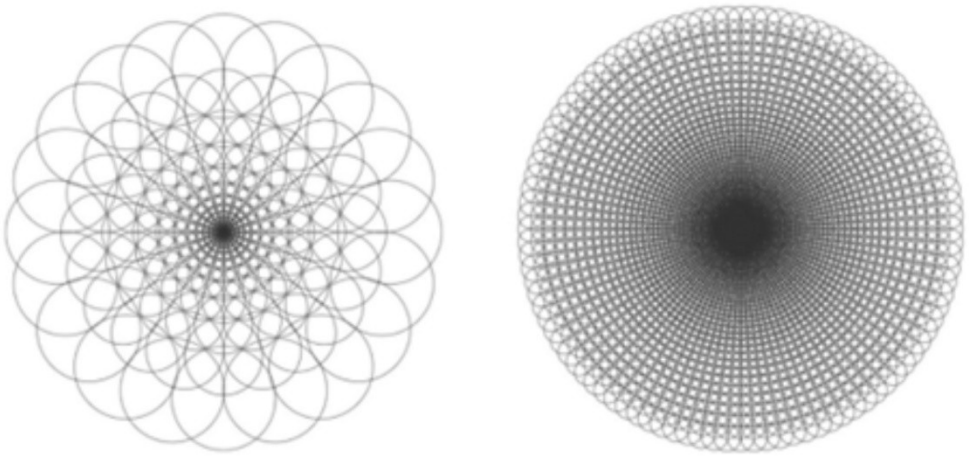


Figura 39 Ejemplo de la arquitectura del mapa retinotópico como mapa logarítmico.
Nota:.(a) coordenadas polares en el espacio visual.; (b) mapas retinotópicos representados como mapas logarítmicos; (c) ejemplo de coordenadas polares en una imagen; (d) ejemplo de mapas logarítmicos con la imagen y con las coordenadas polares. Como referencia de las distancias, la verde es límite de la región central, línea roja es una región intermedia, la azul es la región externa y la blanca el límite del campo visual.

En la Figura 40, hay dos ejemplos, uno para $N_{retina} = 20$ y otro para $N_{retina} = 100$ donde se aprecia claramente la relación del tamaño del campo receptivo con la excentricidad.



(a) (b)

Figura 40 Ejemplo de la posición y tamaño de los campos receptivos.
Nota: (a) $N_{retina} = 20$ y (b) $N_{retina} = 100$..

NGL

Los mapas retinotópicos de la retina aplican la excentricidad con una cantidad de campos receptivos similar en las coordenadas i y j. Sin embargo, en el NGL, la cantidad de campos receptivos varía por la magnificación central y la anisotropía dependiendo si representan la región izquierda o la derecha, la inferior o superior, tal y como nos muestran diversos estudios como el de Van Essen y colaboradores. (Van Essen y otros, 1984) y el de Wandell y Winawer, (Wandell & Winawer, 2011) o, en concreto, sobre las diferencias entre la región izquierda y la derecha de Amunts y colaboradores (Amunts y otros, 2007). La arquitectura de los mapas retinotópicos del NGL tiene la siguiente representación:

$$NGL^{h}_{i',j,v,o,t} \tag{8}$$

donde $h,\ j$, v, o y t son coincidentes con $R^{h}_{i,j,v,o,t}$, e i' son las filas reducidas por la aplicación de $\nabla(g)$ debido a la magnificación central:

$$\nabla(g) = 1 - \left(\frac{M(g)}{M(0)}\right) \tag{9}$$

donde $\theta_{i'} = \theta_i$ si i' corresponde a i. La magnificación se calcula a partir del ángulo del campo visual que representa cada columna j. Existen varios estudios que han analizado esta relación y aplicaremos la ecuación de Schneider y colaboradores (Schneider y otros, 2004):

$$M(g) = 46.6(g + 0{,}52)^{-2.43} \tag{10}$$

siendo g el grado del ángulo del campo visual. $M(g)$ que nos indica cuánto aumenta según la cercanía a la región central, donde $M(0)$ sería el valor máximo. La ecuación se

obtiene de la cantidad de neuronas que representa cada región del espacio visual en el NGL, es decir, los campos receptivos existentes para representar una región del espacio visual concreta y, por lo tanto, una mayor cantidad de campos receptivos implica una mayor cantidad de información. Aumentar la cantidad de campos receptivos no aportaría ninguna ventaja ni computacional, ni representativa, ya que el aumento implicaría una redundancia de la información ya existente. Por lo tanto, aplicamos una función inversa, que es la reducción de campos receptivos donde la mayor cantidad de campos receptivos están en la región central, $M(0)$, normalizando el resultado en el intervalo $[0,1]$. Con la función $\nabla(g)$ se obtiene la cantidad de las filas i' en cada columna j. Para obtener g en cada columna j del mapa retinotópico NGL, aplicamos la siguiente ecuación:

$$g = \frac{r * 45}{W} \tag{11}$$

siendo r la distancia en $V_{r,\theta,c}$ de cada columna j del mapa retinotópico de la retina., y 45, los grados en el mapa de la izquierda y en el de la derecha del espacio visual (90 grados es el campo visual del espacio visual).

Además de la aplicación de la magnificación, es necesario implementar la anisotropía y para esta finalidad se incluyen dos pesos en la ecuación (9) para la relación horizontal-vertical y la anisotropía izquierda-derecha e inferior-superior:

$$\nabla(g) = \frac{\left(1 - \left(\frac{M(g)}{M(0)}\right)\right)}{3.5} * w \tag{12}$$

donde 3.5 es el valor aproximado de la relación entre el eje horizontal y el vertical según el estudio de Schneider y colaboradores (Schneider y otros, 2004), y w es el peso configurable que depende de la región según $\theta_{i'}$. Aunque no existe un coeficiente para la diferencia en la cantidad de campos receptivos entre el mapa de la derecha y el de la izquierda y, tampoco, entre las regiones superior e inferior, a partir de varios estudios como el de Van Essen y colaboradores (Van Essen y otros, 1984) o el de Corbett y Carrasco (Corbett J. E., 2011), se utilizarán los siguientes criterios para configurar w:

- El mapa que representa la región derecha del espacio visual contiene menos campos receptivos, es algo que se repite en otras áreas del cerebro teniendo en cuenta que ambos hemisferios en el cerebro no son iguales.
- Hay una tendencia a tener más campos receptivos en la región inferior que en la superior.

En la arquitectura del NGL, el eje meridional horizontal actúa como «centro» de la región superior e inferior. Aunque no existe una solución clara de cómo aplicar la reducción de los campos receptivos, el eje meridional horizontal, como centro, facilita una distribución excéntrica para ambas áreas, con lo que un planteamiento viable es aplicar una reducción exponencial desde este eje tanto hacia la parte inferior como a la

superior variando la cantidad de campos receptivos según el peso w. Por coherencia con el modelo y la función de excentricidad, la propuesta es aplicar la ecuación de la función de excentricidad de la ecuación (4) para la reducción de filas de los mapas de la retina a los del NGL:

$$E(i') = e^{log\left(\left(\frac{N_{retina}}{4}\right)*\left(\frac{i'}{\nabla(g)}\right)\right)} \tag{13}$$

La función $E(i')$ nos indica la posición i del mapa $R^{h}_{i,j,v,o,t}$ del campo receptivo. El radio del campo receptivo se obtiene con la adaptación de la ecuación (7) para un solo cuadrante de $\frac{3\pi}{4}$, ya que varía la cantidad según cada cuadrante por $\nabla(g)$:

$$CR_{i'} = 2 * \left(1 - cos\left(\frac{\frac{3\pi}{4}}{\nabla(g)}\right)\right) * E(i') \tag{14}$$

La Figura 41 muestra la representación de los mapas retinotópicos de la retina y NGL en mapa logarítmico (izquierda) y con coordenadas polares (derecha). En los mapas de la retina se visualiza la excentricidad, mientras que en los del NGL se aprecia la magnificación central con la anisotropía horizontal-vertical, izquierda-derecha e inferior-superior. En la representación polar, vemos cómo el eje vertical se contrae (los hombros y cabezas se acercan en la región superior). En los mapas logarítmicos, el efecto de magnificación central es más evidente si comparamos la mano en la región central del mapa izquierdo del NGL con la del mapa de la retina, y con su entorno (brazo, cuerpo, otras figuras, etc.) que es reducido.

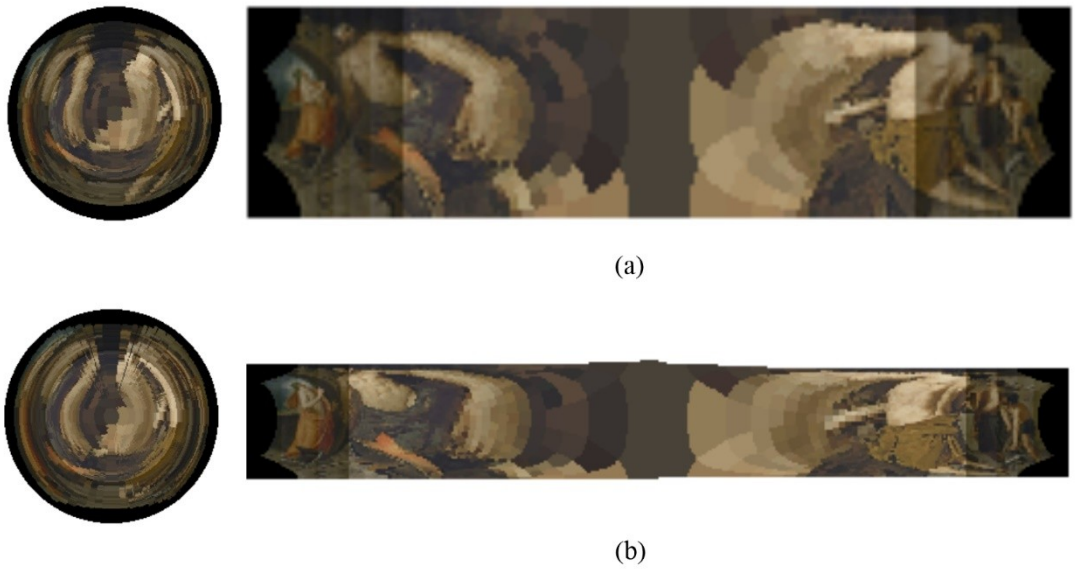


Figura 41 Ejemplo de mapas retinotópicos de la retina y NGL.
Nota: (a) retina y (b) y NGL: La representación en coordenadas polares (izquierda) y mapas logarítmicos (derecha).

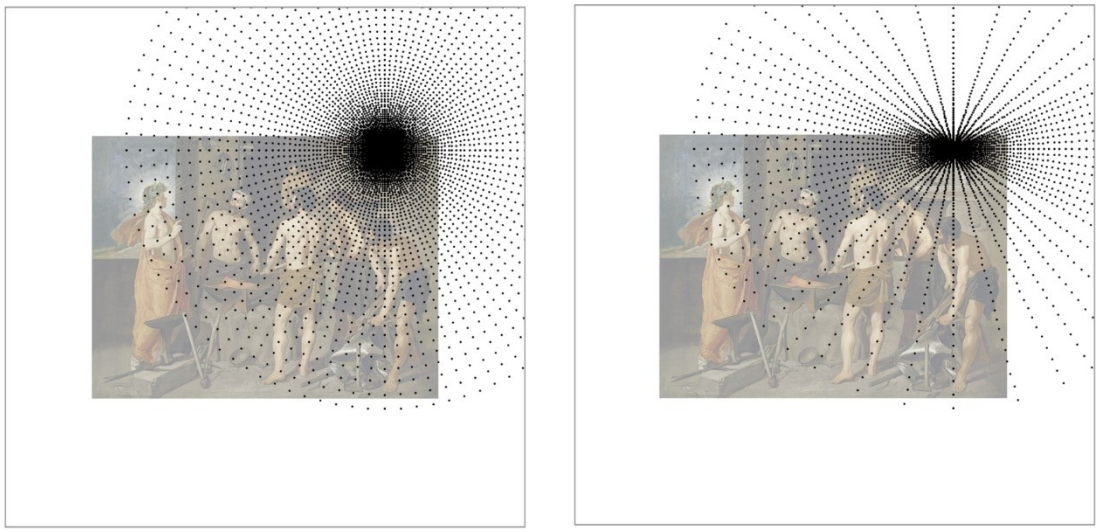


Figura 42 Campos receptivos de mapas en la región superior derecha.
.*Nota:* (a) retina y (b) NGL (b).

Por último, como ejemplo, la Figura 42 muestra la posición de los campos receptivos de los mapas de la retina y NGL para la región superior derecha. Esto permite comprobar cómo se representa la relación entre una región y el resto de las regiones en tanto en la distancia como en su posición en un caso donde la parte extrema más lejana no está representada y cómo las regiones cercanas tienen una mayor representación.

Parámetros de los mapas de la retina y NGL

El valor de N_{retina} de los mapas de la retina es un coeficiente importante a la hora de construir la arquitectura de ambos mapas. La Figura 43 muestra dos ejemplos para $N_{retina} = 100$ y $N_{retina} = 50$ que nos permite comprobar las diferencias y también la reducción de campos receptivos de los mapas de NGL. La concentración de campos en la región central de la retina es un círculo, mientras que en el NGL es una elipse con el diámetro de mayor tamaño en el eje horizontal. El resto de los campos receptivos son reducidos en la dirección vertical del espacio siendo la reducción mayor en la región superior derecha, seguido de la superior izquierda, inferior derecha y finalmente la inferior izquierda, que es la que mayor cantidad de campos receptivos tiene. El resultado final de las arquitecturas de los mapas NGL es que se representan las cuestiones tanto de la preponderancia de la región inferior izquierda de Dondis como el marco estructural de Arnheim en relación con la cantidad de campos receptivos y su posición en el espacio visual.

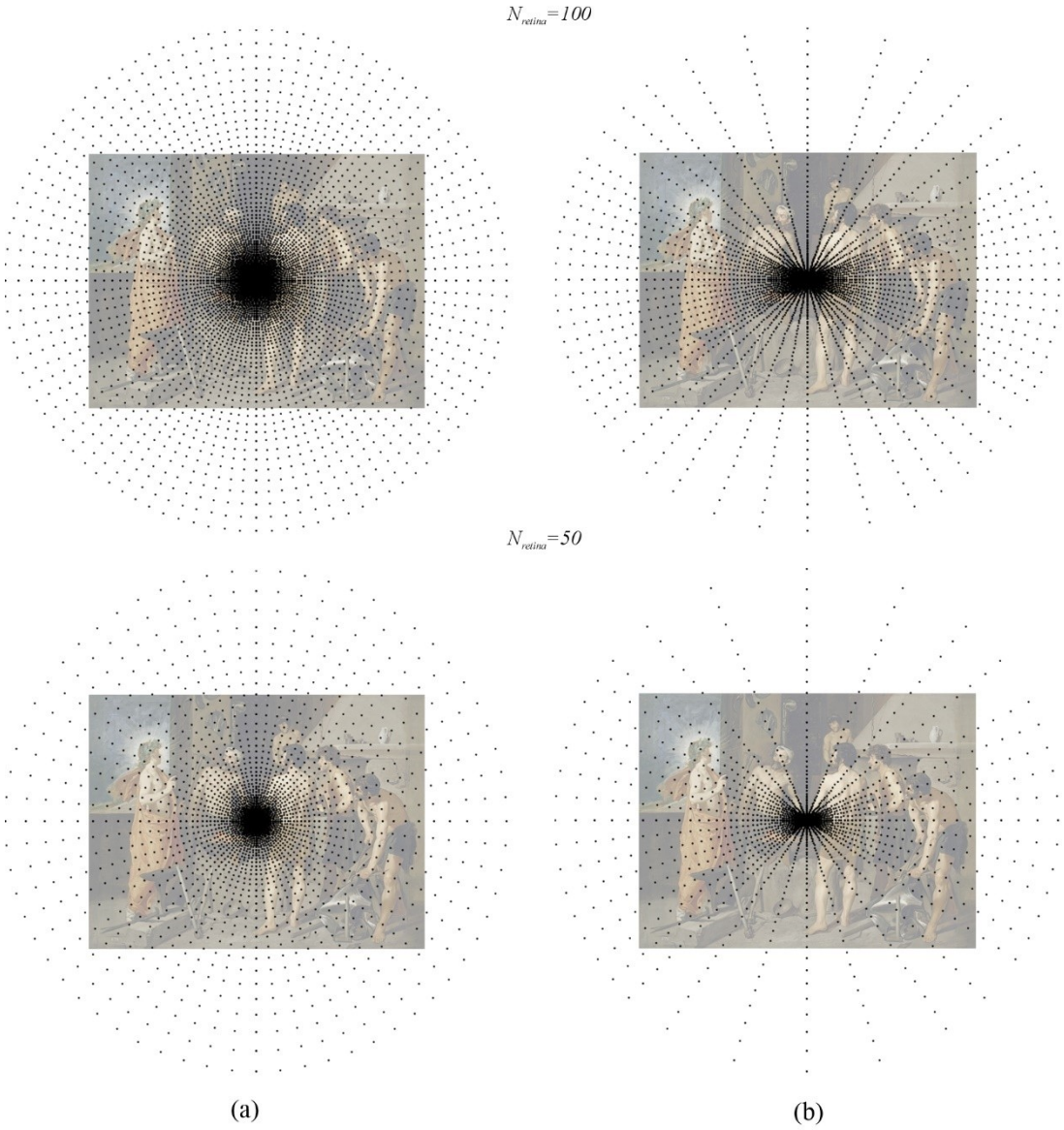


Figura 43 Campos receptivos de mapas en el centro geométrico..
.*Nota:* (a) retina y (b) NGL (b).

4.2.2 Creación del mapa de prominencia como un esquema de referencia

A. Descripción de la tarea, del procesamiento visual y de la respuesta esperada en el espectador

La tarea es diseñar la arquitectura del mapa de prominencia como esquema de referencia. La finalidad es que satisfaga las características funcionales descritas desde la psicología del arte y la neurociencia, ajustándose con la descripción de la sintaxis visual

de Dondis y en el «marco estructural» de Arnheim. La arquitectura se debe articular a partir de dos focos estructurales: el centro y el exterior, y la preponderancia de la región inferior izquierda.

B. Los retos principales en relación con las limitaciones físicas, algorítmicas, temporales o psicológicas e las que dependen las composiciones de las imágenes

Los retos y limitaciones son similares a los de las arquitecturas de los mapas retinotópicos de la retina y NGL, por la implementación tanto de los dos focos de atracción como de la preponderancia de la región inferior izquierda. Además, la principal limitación está en la selección de la cantidad de campos receptivos, ya que una gran cantidad implicaría un proceso de escaneado largo y una cantidad reducida, una representación de información insuficiente. El equilibrio de ambas cuestiones es más complejo para esta arquitectura que para las de los mapas de la retina y NGL.

C. Solución biológica del problema en relación con el comportamiento psicológico

Cuando se contempla una pintura, en el cerebro del espectador se establece una relación espacial en un esquema de referencia egocéntrico (el espectador como centro) que representa todo el entorno físico que le rodea y, a su vez, la pintura es representada en uno alocéntrico (la pintura como centro). Los esquemas de referencia se encuentran en el área parietal inferior de la corteza y son mapas visuales, espaciales y visio-motores (Cohen & Andersen, 2002) con la funcionalidad de relacionar las distintas regiones del espacio visual (Pouget & Sejnowski, 1997) (Galati y otros, 2010) (Pertzov y otros, 2011) (Chen y otros, 2012). El esquema egocéntrico se encarga de localizar las posiciones de los ojos en la imagen, la localización de la imagen y también de los movimientos a realizar. El esquema alocéntrico de la imagen representa las regiones en relación con su centro geométrico.

La Figura 44 muestra la relación entre el esquema egocéntrico (verde), donde la imagen y sus regiones están referenciadas con el espectador, en concreto con los ojos ya que para la contemplación del cuadro no es necesario, a priori, mover la cabeza o el cuerpo. Por otro lado, la imagen es representada por un esquema alocéntrico (magenta) donde existe una relación interna entre el centro geométrico y los ejes horizontal y vertical. Esta simplificación de ambos esquemas ayuda a visualizar cómo la imagen es representada en dos esquemas de referencia con distintos centros —uno en el espectador y otro en la imagen— y cómo se establece, por un lado, la representación de la imagen (magenta) y, por el otro, su escaneo (verde) moviendo los ojos por la imagen.

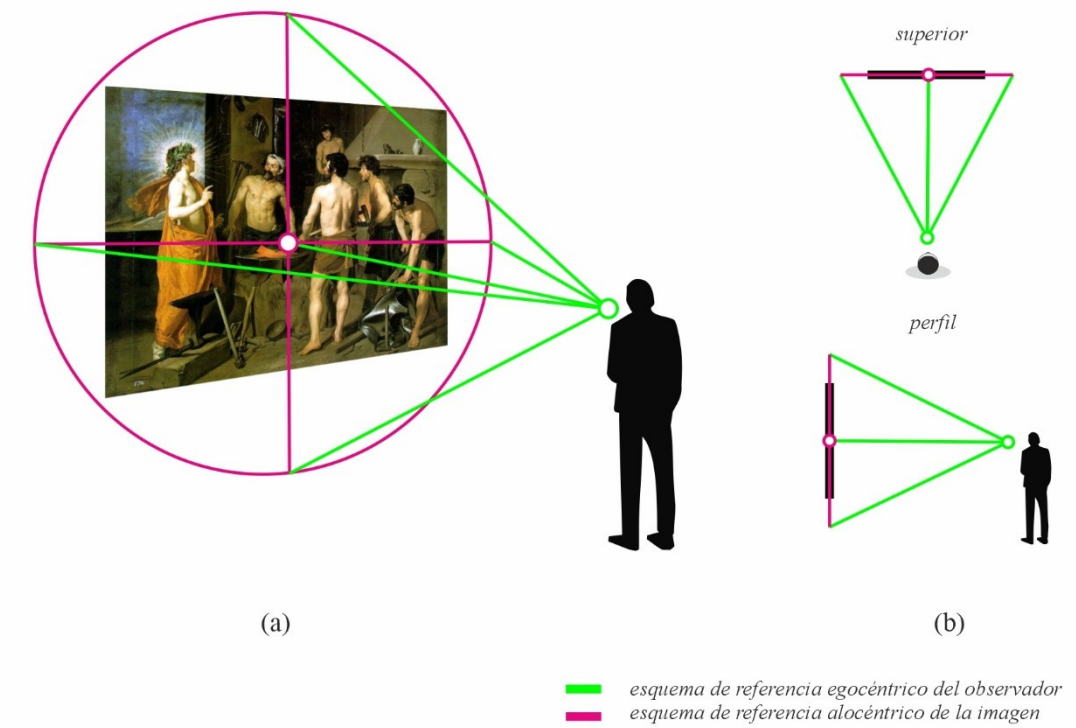


Figura 44 Interacción entre el esquema de referencia egocéntrico y el alocéntrico
Nota: (a) representación de ambos esquemas en relación al espectador y la imagen; y (b) representación superior y en perfil.

Las relaciones que se establecen entre ellos hacen posible la interacción con el espacio, el movimiento, coger objetos, etc. (Meilinger, 2008). Estos esquemas son fijos en el tiempo y sus neuronas van modificado sus valores dependiendo de los mapas retinotópicos que se generan en cada momento y que, obviamente, amplían información con nuevos detalles. Estos esquemas actúan como una memoria visual —de hecho, se relacionan con parte del cerebro destinada para este fin— para facilitar la interacción con el mundo.

D. Comparación con otras soluciones de visión artificial

No existen representaciones del mapa de prominencia usando esquemas de referencia, tanto para determinar el tamaño de los campos receptivos, como de la anisotropía izquierda-derecha e inferior-superior. Sin embargo, sí existe la representación de las imágenes por regiones, habitualmente homogéneas en el tamaño a través de rejillas, que se aplican habitualmente para evitar trabajar con píxeles por los costes computacionales. Desde las Bag of Words (o Bag of features), usando rejillas regulares, como en el trabajo de Fergus y colaboradores (Fergus y otros, 2005) o por regiones con puntos clave, como en Csurka y colaboradores (Csurka y otros, 2004) hasta los *Vision Transformers*, como en el trabajo de Chen y colaboradores. (Chen y otros, 2021).

E. Diseño de la solución

Los esquemas de referencia mantienen rasgos comunes con los mapas retinotópicos como la excentricidad, aunque la magnificación central es menor. La arquitectura utiliza coordenadas polares como describe Fattori y Pitzalis en su estudio del área V6 de los macacos (Fattori & Pitzalis, 2009) y se construye a partir del espacio visual $V_{r,\theta,c}$ como:

$$ER_{p,q}^{h} \tag{15}$$

donde $h \in \{izquierda, derecha\}$ —mapa del hemisferio derecho o el del izquierdo—, p la posición de θ y q la posición del r de $V_{r,\theta,c}$, que se obtiene con la función de excentricidad:

$$\mathrm{r} = \mathrm{E}(q) = e^{log\left(\frac{W}{2}\right)*\left(\frac{q}{N_{esquema}}\right)} \tag{16}$$

siendo $N_{esquema}$ el total de columnas, W el r máximo del espacio visual.

A diferencia del mapa NGL, que aplica una anisotropía entre sus regiones a partir del mapa de la retina, el esquema de referencia lo realiza desde el espacio visual directamente. Para poder aplicarlo, el esquema se divide en cuatro cuadrantes, donde el eje horizontal es el índice 0, y a partir de ahí, cada cuadrante tiene una cantidad distinta de campos receptivos para las filas según cada columna. En este sentido, la posición de la fila p se obtiene con el ángulo θ_p:

$$\theta_p = \frac{\theta_0 * \Upsilon(p * \frac{\pi}{4})}{\mu * \frac{N_{esquema}}{4}} \tag{17}$$

siendo Υ =1 una constante cuando hay incremento a partir de θ_0 (en la región superior-derecha e inferior-izquierda), y Υ =-1 cuando hay decremento (en la región superior-izquierda e inferior-derecha). De esta manera, la posición de cada p representa a un ángulo θ_p a partir del eje horizontal, bien ascendiendo con índices positivos o bien descendiendo con índices negativos. El θ_0 de cada cuadrante, que está en el eje horizontal, sería:

$$\theta_0 = \begin{cases} 0, & 0 \le \theta_p < \frac{\pi}{4} \\ \pi, & \frac{\pi}{4} \le \theta_p < \pi \\ \pi, & \pi \le \theta_p < \frac{3\pi}{2} \\ 0, & \frac{3\pi}{2} \le \theta_p < 2\pi \end{cases}$$

el peso μ depende del cuadrante y es configurable para la anisotropía. Aunque no existen estudios que indiquen la cantidad o las diferencias con exactitud, sí que, al igual que vimos con los mapas retinotópicos del NGL, hay una preponderancia de la región izquierda sobre la derecha y de la inferior sobre la superior. Este peso se configura dependiendo de la implementación que se haga del modelo.

Al igual que en los mapas retinotópicos del NGL, el campo receptivo se obtiene con la ecuación (14) pero con la siguiente adaptación:

$$CR_p = 2 * \left(1 - cos\left(\frac{\frac{3\pi}{4}}{\mu * \frac{N_{esquema}}{4}}\right)\right) * E(p) \tag{18}$$

El radio de los campos receptivos depende de la cantidad de campos en cada cuadrante, determinada con el peso μ , con lo que la distancia entre los ángulos se obtiene para cada cuadrante según θ_p. El tamaño del radio es similar en todos los campos receptivos q en cada columna p.

La Figura 45 muestra, como ejemplo, un esquema de referencia con $N_{esquema} = 20$, donde se representa la arquitectura con los campos receptivos en una representación polar del espacio visual y los dos mapas de los esquemas de referencia como mapas logarítmicos. La reducción de filas, en este ejemplo, según el cuadrante, obtiene: cinco para la región inferior-izquierda, cuatro para la superior izquierda e inferior derecha y

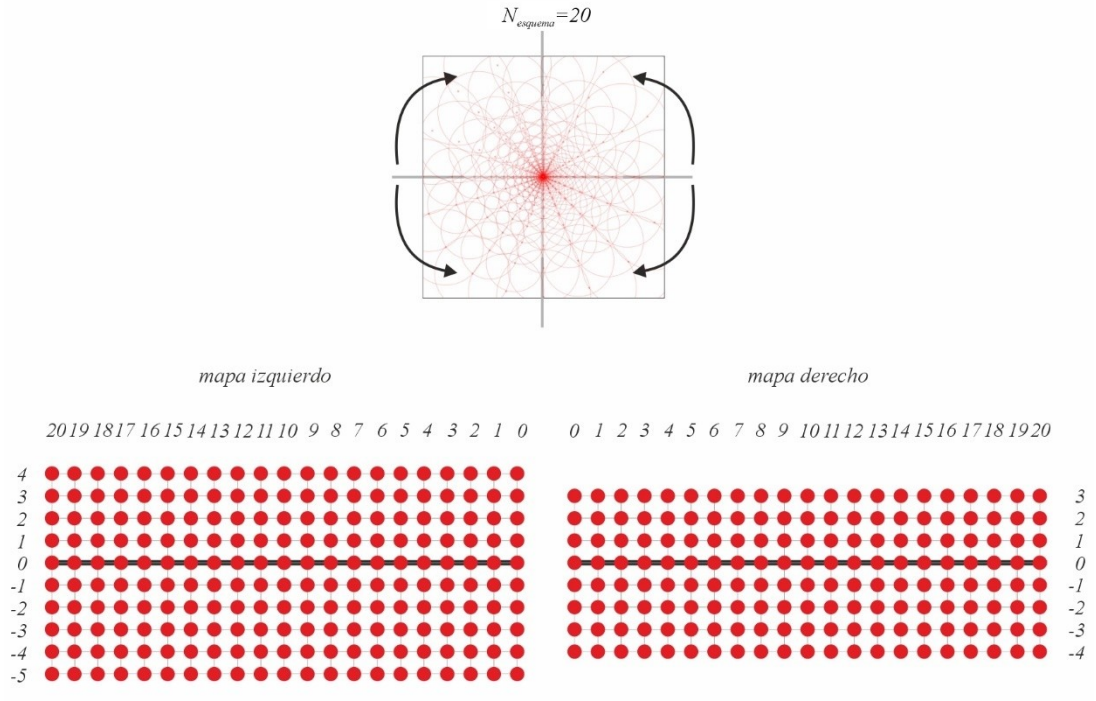


Figura 45 Arquitectura del esquema de referencia como mapa logarítmico.
Nota: en la representación, cada punto rojo es un campo receptivo y los índices positivos equivalen a la región superior y los negativos a la inferior.

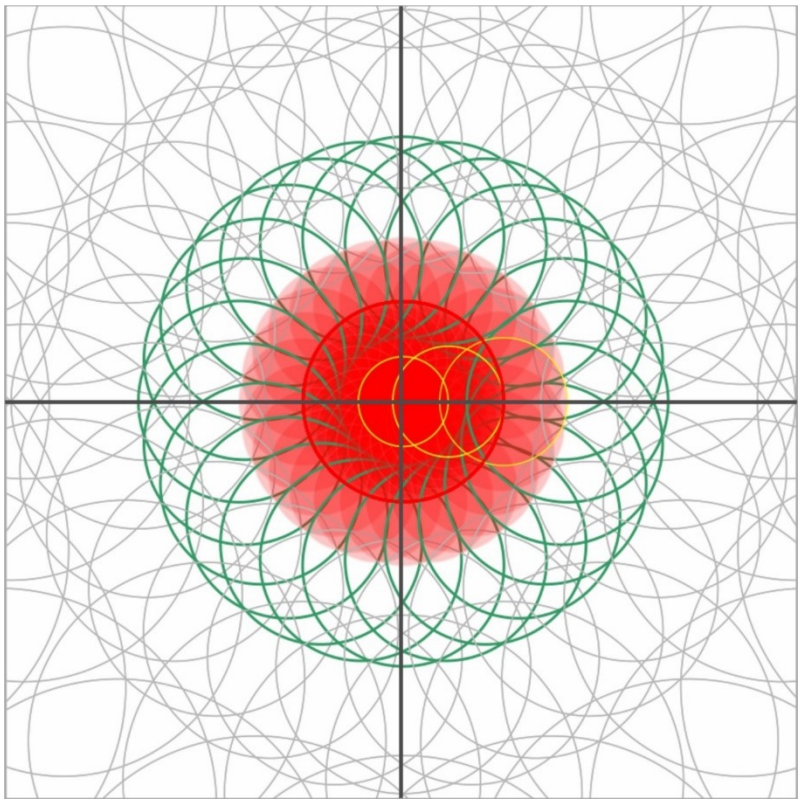


Figura 46 La región central del esquema de referencia.
Nota: el límite de la región central son los campos receptivos en verde, y los campos que quedan dentro de la región central están rellenos de rojo.

tres para la superior derecha. Al igual que con los mapas retinotópicos, se hará referencia siempre al esquema de referencia en singular, incluyendo el esquema de la izquierda y el de la derecha, salvo que se especifique lo contrario.

La función de excentricidad genera campos receptivos muy pequeños en la región central, que para el mapa de prominencia del tejido interno son innecesarios. Al reducir el nivel de detalle y establecer unos valores pequeños de $N_{esquema}$, también debemos establecer un campo receptivo para la región central que agrupe a todos campos receptivos que estén por debajo de un tamaño mínimo. Para eso, el tamaño de este centro deberá ocupar un espacio porcentual de $N_{esquema}$:

$$r_{centro} = N_{esquema} * \alpha \tag{19}$$

estando α en un intervalo $[0,1]$. Este radio indica los campos receptivos que pertenecen a la región central si su campo receptivo es inferior a r_{centro}:

$$ER_{p,q}^{h} \in centro\ si\ CR_p \leq r_{centro} \tag{20}$$

siendo $\alpha = 0.125$ el valor más adecuado en las pruebas realizadas, aunque, este coeficiente se podría ajustar para ampliar o reducir la región central si la tarea fuera distinta a la del objeto de esta tesis. La Figura 46 presenta un ejemplo en el que se muestran relle-

nos de rojo los campos receptivos que tiene un radio inferior al de la región central marcado con una línea roja. Los campos receptivos marcados con la línea verde serían los más cercanos a la región central y los grises, el resto. Se presentan tres marcados con línea amarilla, pertenecientes a la región derecha y en el eje horizontal, para visualizar con mayor claridad que con este ajuste se incluirían las regiones de todos estos campos receptivos demasiado pequeños en el campo receptivo de la región central (círculo rojo).

4.2.3 Calcular los pesos visuales

A. Descripción de la tarea, del procesamiento visual y de la respuesta esperada en el espectador

El objetivo de la tarea es calcular el peso visual de cada campo receptivo del NGL. El peso visual cuantifica la capacidad de una región para atraer la atención del espectador y, para el tejido interno, representa la prominencia de una región local por sí misma sin la relación con el resto de las regiones. Además, la estructura de datos de los píxeles del espacio visual en RGB debe transformarse en un sistema de colores opuestos en dos vías, ON y OFF tal y como se encuentran en la retina y NGL.

B. Los retos principales en relación con las limitaciones físicas, algorítmicas, temporales o psicológicas e las que dependen las composiciones de las imágenes

Los valores de cada campo receptivo son los canales de un sistema de colores opuestos en dos vías (ON y OFF). El principal reto es situar cada color en una escala jerárquica que permita obtener un valor cuantificable que represente la combinación de las propiedades de matiz, luminosidad y saturación. donde el blanco es el extremo superior y el negro, el inferior en la vía ON, y a la inversa en la OFF. La principal limitación es la pérdida de información en un proceso de reducción de varios canales a uno solo.

C. Solución biológica del problema en relación con el comportamiento psicológico

Rudolf Arnheim citó a Schopenhauer como precursor de la teoría de colores opuestos en su libro Arte y percepción (Arnheim, 1956). Indicaba que su escala de diferencias cuantitativas era de interés todavía para su investigación y que su concepción básica de las parejas complementarias en la funcionalidad de la retina anticipaba la teoría de color de Ewald Hering. Sin embargo, concluía con la incapacidad de Schopenhauer para poder describirla a nivel fisiológico. El sistema tricromático se basa en las propiedades físicas de la luz, en concreto en las longitudes de onda y como la retina es capaz de detectar tres rangos (largo, medio y corto, denominados como L, M y S). Goethe (Goethe, 1840) y, sobre todo, Schopenhauer (Schopenhauer, 1816) pusieron en duda el planteamiento de Newton de que los colores surgieran a partir de la descomposición de la luz. Goethe criticó esta teoría realizando una serie de experimentos de percepción visual, especialmente el estudio de la postimagen, que sólo es posible explicar si existe un sis-

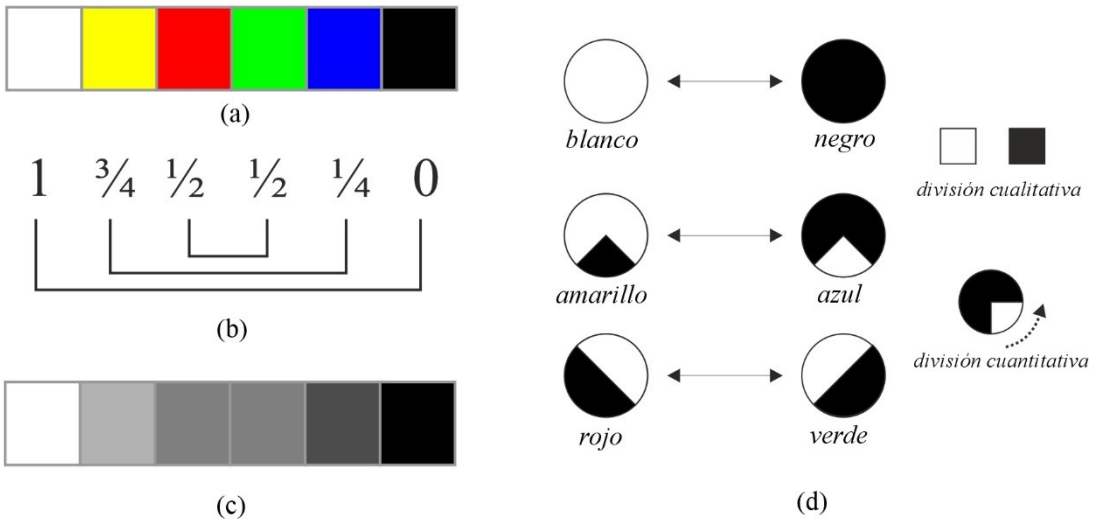


Figura 47 Escala jerárquica del color basada en la actividad dividida de la retina.
Nota: (a) colores primarios; (b) escala jerárquica basada en la actividad dividida de la retina; (c) escala jerárquica; y (d), relación entre la escala jerárquica y la actividad dividida. Las secciones blancas de los círculos representan la actividad y las negras la inactividad (división cualitativa), y el tamaño del área de cada uno el nivel de actividad e inactividad (división cuantitativa).

tema de colores opuestos a través de algún tipo de procesamiento cerebral y no como un efecto físico de la luz.

Schopenhauer elaboró su propia teoría a partir de varios conceptos, siendo los más evidentes y definibles:

- La actividad de la retina relacionada con el procesamiento del color. Hay que tener en cuenta que la teoría vigente en su época establecía que la «descomposición de la luz» generaba los distintos colores. Schopenhauer indica en su libro «Sobre la visión y el color» (Schopenhauer, 1816) que: «A la peculiar reacción del ojo al estimulo externo la denominaré su *actividad*, y es más en concreto, la actividad por sí misma, de forma inmediata y originaria, es la luz. Así que el ojo recibe el influjo de la luz manifiesta la *plena actividad de la retina.* En ausencia de la luz, o en *oscuridad*, aparece la *inactividad* de la retina».

- La división de la actividad, bien por cantidad o bien por cualidad. En su libro indica que «La acción de la luz y del blanco en la retina, y la consiguiente actividad de ésta, poseen grados en los cuales la luz se aproxima a la oscuridad y el blanco al negro en un tránsito continuado […] La indudable divisibilidad de la actividad de la retina en intensidad y en extensión que se ha expuesto hasta aquí, se puede sintetizar en el concepto común de una *divisibilidad cuantitativa de la actividad de la retina.* […] La diversidad de los colores es el resultado de las diversas mitades cualitativas en las que esa actividad se puede separar y de su relación recíproca. Esas mitades pueden ser *iguales* una sola vez, y entonces se presentan el rojo verdadero y el verde perfecto».

En la arquitectura de los mapas retinotópicos, cada campo receptivo representa la información tricromática captada en la retina y transformada a una estructura de dos vías opuestas (ON y OFF), y en uno de los cuatro canales: dos parvocélulas, (L y M), una koniocélula (S) y una magnocélula (LM). El blanco sería la máxima intensidad de todos los canales (L, M, S y LM), en la vía ON, y ninguno en la OFF, y el negro, lo opuesto, la máxima intensidad en la vía OFF (-L, -M-S y -LM), y ninguna en ON. La plena actividad de la retina y la degradación hacia la inactividad se relaciona fácilmente con la vía ON y, al contrario, con la OFF. Por consiguiente, según la teoría de Schopenhauer, cada color lleva asociado un grado de actividad de la retina, pero también de inactividad. Por ejemplo, el amarillo sería ¾ de actividad, y ¼ de inactividad. En la Figura 47, se muestra gráficamente la relación entre la escala de color y la actividad de la retina para los colores primarios planteada por Schopenhauer. Los colores primarios (Figura 47.a) establecen una relación de procesos de colores opuestos entre el amarillo y el azul, rojo y verde, y blanco y negro (Figura 47.b). La suma de los valores de todas las parejas sería la plena actividad (Figura 47.d). La jerarquía de esta graduación de color en relación con la actividad de la retina permite relaciones opuestas, por ejemplo, el amarillo ¾ en la vía ON y ¼ en la OFF, y su pareja azul, ¼ en la vía ON y ¾ en la OFF, establece una relación inversa con su graduación en ON y en OFF, y a su vez con su pareja opuesta (ver la relación entre Figura 47.b y Figura 47.c).

La función relé es uno de los aspectos más característicos de los campos receptivos del NGL permite, como principales funciones en relación con el procesamiento de la información, por un lado, inhibir regiones (Alitto & Usrey, 2003) y, por el otro, modular las entradas para que la respuesta se ajuste a estados anteriores (Agarwal & Sarma, 2011). Esta función se desarrolla en el NGL de diversas maneras, teniendo en cuenta, además, que cerca del 70% de conexiones que tiene con la corteza visual son de retroalimentación. En el enfoque de Einvoll (Einevoll, 2003), las funciones de modulación e inhibición provienen de diversas fuentes que simplifica como *brainstem reticular formaction (BRF)* y escalonadas en varias conexiónes. Por ejemplo, la inhibición es realizada por el núcleo retícular del tálamo, que controla, entre otras cuestiones, el movimiento de los ojos y la coordinación de ambos, que a su vez recibe flujos de estimulación de la corteza visual. Davila Teller (Teller, 2014) establece un modelo más simple, donde el campo receptivo recibe un flujo de datos desde las células ganglionales, de estimulación en la zona central y de inhibición desde la denominada intercélula que recoge los flujos de su entorno —ambos forman el campo receptivo—, y una segunda vía inhibitoria desde núcleo reticular del tálamo. En este segundo modelo, podemos ver, por un lado, la entrada de datos con el campo receptivo y, por el otro, la modificación por el valor anterior del campo receptivo, que determina el procesamiento de un señal temporal (Norheim y otros, 2012). Existen más posibilidades de control aparte de las de modulación e inhibición sobre la entrada recibida, pero ambas definen perfectamente la funcionalidad de relé del NGL.

D. Comparación con otras soluciones de visión artificial

En visión artificial y en el procesamiento de la imagen, la conversión de tres canales (RGB) a uno solo (escala de grises) se ha llevado a cabo a partir de varios enfoques. El

principal es el estudio de cómo el espacio de los colores es percibido por los seres humanos, con el fin de ajustar la señal tricromática aplicando pesos a cada canal RGB. Este tipo de procesamiento se ha utilizado en las pantallas en blanco y negro y en los programas de edición de imágenes, y destacamos principalmente el de la Recomendación BT.601 de la Unión Internacional de Telecomunicaciones (UIT) que establece los siguientes pesos para cada canal RGB con el fin de la obtención de un solo canal de escala de grises:

$$G = 0.29 * R + 0.587 * G + 0.114 * B$$

siendo G el valor del canal de escala de grises. Otros ejemplos de este procedimiento lo podemos encontrar en el conversor de Bala y Eschbach (Bala & Eschbach, Color to grayscale conversion method and apparatus, 2008), de Majewicz, y Smith (Majewicz & Smith, 2013) o de Ng (Ng, 2013)). En segundo lugar, otra forma de conversión utilizada es la del PCA (principal componente análisis) que realiza una reducción de dimensionalidad desde un conjunto grande de datos, en este caso desde tres canales RGB a un solo canal. Este sistema es aplicado, por ejemplo, en el trabajo de Seo y Kim (Seo & Kim, 2013), para realizar una conversión de una imagen en color a una en escala de grises manteniendo la información del color. En tercer lugar, hay procedimientos que utilizan el contraste de cada píxel con su vecindario, donde destaca color2gray de Gooch y colaboradores (Gooch y otros, 2005) en Kuk y colaboradores (Kuk y otros, 2011) o en el trabajo de Bala y Eschbach donde usaron consideraciones espaciales a su conversor (Bala & Eschbach, 2004).

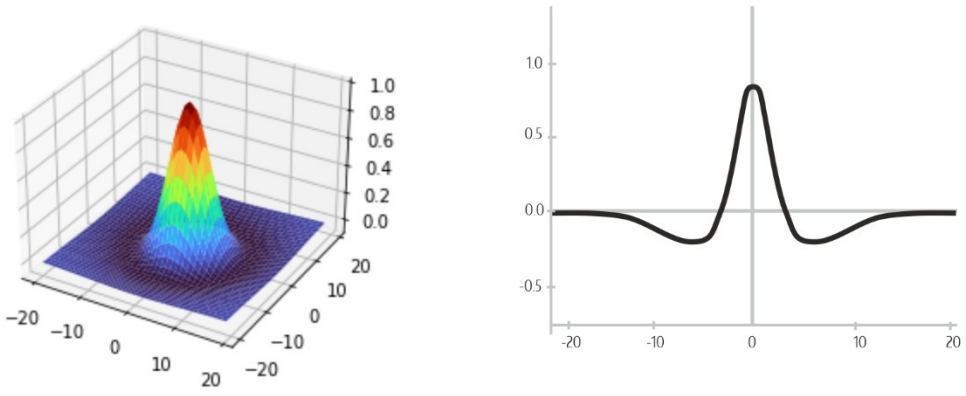
Sobre la cuestión de modulación e inhibición en arquitecturas retinotópicas, no existen aplicaciones conocidas, aunque tanto la modulación como la inhibición han sido usadas con otras finalidades a la subtarea que estamos describiendo en el procesamiento de imágenes.

E. Diseño de la solución

El peso visual para el tejido interno se relaciona con «la actividad» de los campos receptivos en dos vías (ON y OFF) y cuatro canales del mapa NGL modulada e inhibida. Para poder implementar esta funcionalidad se plantean tres operaciones secuenciales: la primera, la transformación de los valores RGB de los píxeles del espacio visual a través de los campos receptivos de la retina y de los de la retina a los campos receptivos del NGL en las vías ON y OFF; la segunda, la obtención de un valor único de «actividad» para cada vía ON y OFF que combine a los cuatro canales; y, la tercera, la aplicación de la modulación y la inhibición.

La transformación RGB a un sistema de colores opuestos en los campos receptivos de la retina y NGL

En la retina, el procesamiento de la información de los campos receptivos se ha implementado de varias maneras, bien como modelos matemáticos (Soodak, 1986) o bien por bioinspiración (Dacey y otros, 2000), pero es la «diferencia de gaussianas» (DoG) descrita por Einevoll, (Einevoll, 2003) la más utilizada en la actualidad (ver Figura 48).



(a) (b)

Figura 48 Gráfico de la función DoG para un espacio en dos dimensiones.
Nota: (a) representación en 3D; y (b) representación en 2D.

Al representar un espacio en dos dimensiones, el valor de una unidad dentro del campo receptivo se obtiene:

$$DoG(x', y') = \left(\frac{1}{2\pi\sigma_1^2} e^{\frac{-x'^2+y'^2}{2\sigma_1^2}}\right) - \left(\frac{1}{2\pi\sigma_2^2} e^{\frac{-x'^2+y'^2}{2\sigma_2^2}}\right) \quad (21)$$

donde los valores positivos corresponden al área de estimulación, y los negativos, al de inhibición. Los valores x' y y' son relativos en $V_{x,y,c}$ al centro del campo receptivo correspondiente a i y j. Para los coeficientes de σ_1 y σ_2 se propone la siguiente relación, que es la más utilizada, donde el valor σ_1 de la primera gaussiana es un 1/3 del de la segunda σ_2:

$$\sigma_1 = \frac{1}{3} * \sigma_2 \text{ siendo } \sigma_2 = CR_j \quad (22)$$

siendo CR_j es el radio del campo receptivo.

El valor final se obtiene a través de una convolución de todos los valores $V_{x',y',c}$ dentro del campo receptivo obtenidos con la función DoG:

$$R^h_{i,j,v,o,t} = \sum_{x'=0}^{CR_j} \sum_{y'=0}^{CR_j} V_{x',y',c} * DoG(x', y') \quad (23)$$

Los valores de cada canal o se obtienen teniendo en cuenta la transformación del espacio tricromático de los canales c de $V_{x,y,c}$ al de cuatro canales opuestos donde un valor

	L	M	S	LM	-L	-M	-S	-LM
centro	R	G	B	min(R+G,1)	1-R	1-G	1-B	1-min(R+G,1)
entorno	G	R	min(R+G,1)	B	1-G	1-R	1-min(R+G,1)	1-B

Tabla 1 Centro y entorno de los campos receptivos del mapa de la retina.
Nota: canales RGB.

positivo en $DoG(x', y')$ equivale al centro estimulador del campo receptivo y el negativo al entorno inhibitorio:

$$o = \begin{cases} centro, & DoG(x', y') \geq 0 \\ entorno, & DoG(x', y') < 0 \end{cases} \tag{24}$$

donde en $v = ON$ los canales $c \in \{L, M, S, LM\}$ y $= OFF$ los canales $c \in \{-L, -M, -S, -LM\}$. Siguiendo la relación entre la longitud de onda (L,M y S) con los canales RGB, donde R equivale a L, G equivale a M y S equivale a B, y a partir de los criterios de los campos receptivos de las magnocélulas, parvocélulas y koniocélulas, la correspondencia de los canales RGB para el centro y entorno se describen en la Tabla 1. Los canales RGB están normalizados en el intervalo $[0,1]$.

En el NGL, cada campo receptivo aplica la función de DoG al igual que los mapas de la retina, siendo el valor final para cada canal:

$$NGL^{h}_{i',j,v,c,t} = \sum_{p=0}^{CR_{i'}} \sum_{q=0}^{CR_{i'}} R^{h}_{E(i')+p,j+q,v,o,t} * DoG(E(i') + p, j + q) \tag{25}$$

Los valores positivos de la DoG corresponden el centro estimulador del campo receptivo y los negativos al entorno inhibitorio siguiendo los criterios de la ecuación (24). Al igual que los campos receptivos de la retina, $v = ON$ los canales $c \in \{L, M, S, LM\}$ y $= OFF$ los canales $c \in \{-L, -M, -S, -LM\}$. Los canales que estimulan el centro o inhiben el entorno son los de los campos receptivos de la retina y se describen en la Tabla 2

	L	M	S	LM	-L	-M	-S	-LM
centro	L	M	S	LM	-L	-M	-S	-LM
entorno	M	L	LM	S	-M	-L	-LM	-S

Tabla 2 Centro y entorno de los campos receptivos del mapa del LGN.
Nota: Canales OCC; los canales del título de cada columna son los del mapa retinotópico del NGL y los valores de las columnas son los del mapa retinotópico de la retina.

El peso visual de cada campo receptivo

A partir de la teoría de Schopenhauer, que relaciona el color con la actividad de la retina, se convierten los cuatro canales c (L, M, S, y LM) en la vía $v = ON$, y c (-L, -M, -S y-LM) en la vía $v = OFF$, a un valor de «actividad neuronal». Partiendo de los conceptos de actividad dividida de la retina de Schopenhauer, definimos las vías opuestas ON y OFF, que transportan los valores de actividad e inactividad, siendo su suma, «la actividad completa» (AC), constante. En lo que sigue, asumiremos que cada canal transporta una señal en el rango [0,1] y que la actividad plena es AC=1. La forma más simple de estimar la actividad e inactividad en las vías ON y OFF es considerar que los distintos canales son homogéneos y, por tanto, tienen el mismo peso en la transformación. Así, definimos los valores de actividad A_{ON} e inactividad A_{OFF} como sigue:

$$A_{ON} = \frac{L+M+S+LM}{4} \tag{26}$$

$$A_{OFF} = \frac{-L+-M+-S+-LM}{4} \tag{27}$$

De esta manera, los mapas de pesos visuales se obtienen como el valor de actividad para ambas vías desde los mapas NGL:

$$PV^{h}_{i',j,ON,t} = \frac{NGL^{h}_{i',j,ON,L,t}+NGL^{h}_{i',j,ON,M,t}+NGL^{h}_{i',j,ON,S,t}+NGL^{h}_{i',j,ON,LM,t}}{4} \tag{28}$$

$$PV^{h}_{i',j,OFF,t} = \frac{NGL^{h}_{i',j,OFF,-L,t}+NGL^{h}_{i',j,OFF,-M,t}+NGL^{h}_{i',j,OFF,-S,t}+NGL^{h}_{i',j,OFF,-LM,t}}{4} \tag{29}$$

manteniendo $h,\ i', j,$ y t de $NGL^{h}_{i',j,v,a,t}$.

Esta transformación es consistente con el concepto de actividad dividida definida por Schopenhauer tanto en su parte cualitativa como cuantitativa y se relaciona con el sistema de procesos de colores opuestos de Hering por las posibilidades que ofrecen las combinaciones de los canales L, M, S y LM. En Figura 50, mostramos los valores de actividad/inactividad para los colores primarios y secundarios propuestos por Schopenhauer (ver Figura 50.b), donde se evidencian las relaciones opuestas derivadas de las transformaciones (26) y (27). La teoría de colores opuestos establece las relaciones amarillo-azul y rojo-verde, además de blanco-amarillo, y se generalizan usando las vías de actividad e inactividad ON/OFF (ver Figura 50.c y Figura 50.d), es decir, el opuesto tendrá un valor $1 - A_{ON}$ y $1 - A_{OFF}$. Por otro lado, las relaciones amarillo-azul, rojocian y verde-magenta son complementarias porque la suma de los valores de sus canales L, M y S es la actividad completa, es decir, el blanco.

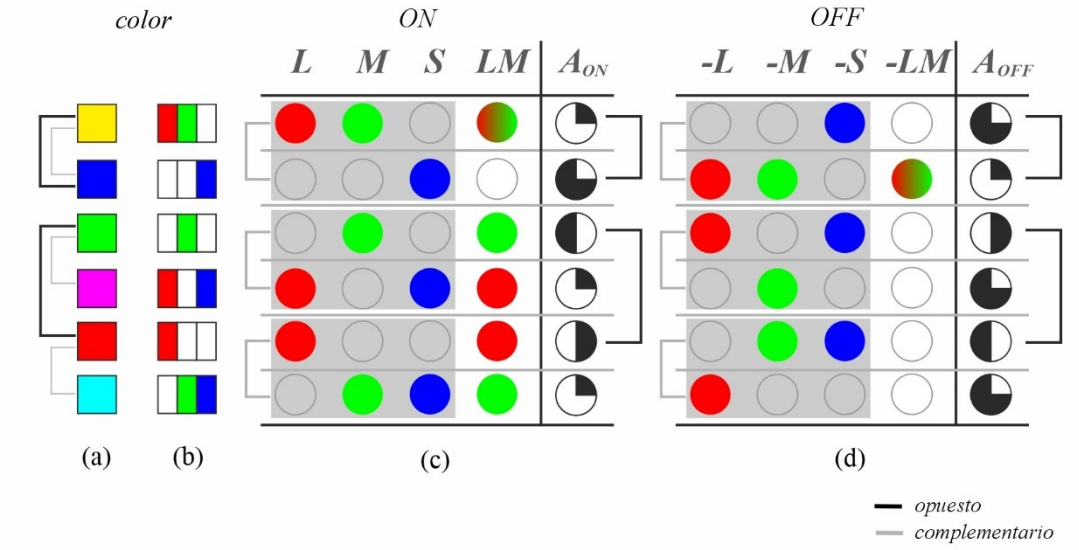


Figura 50 Relaciones opuestas y complementarias del sistema OCC.
Nota: (a) colores primarios; (b) canales RGB; (c) canales OCC en la vía ON; y (c) canales OCC en la vía OFF. Para representar los canales OCC, el canal L usa el rojo, para M, el verde, para S, el azul y para LM, el amarillo.

Así pues, por un lado, los procesos de colores opuestos se relacionan con la actividad e inactividad de la retina y, por otro, las relaciones complementarias se relacionan con la señal captada por los conos de la retina. Este sistema de color que se utiliza en esta tesis se denomina OCC (Opponent, Complementary Color) (España Patente nº 201831253, 2017) ya que permite describir ambas relaciones y obtener los valores en cada caso y,

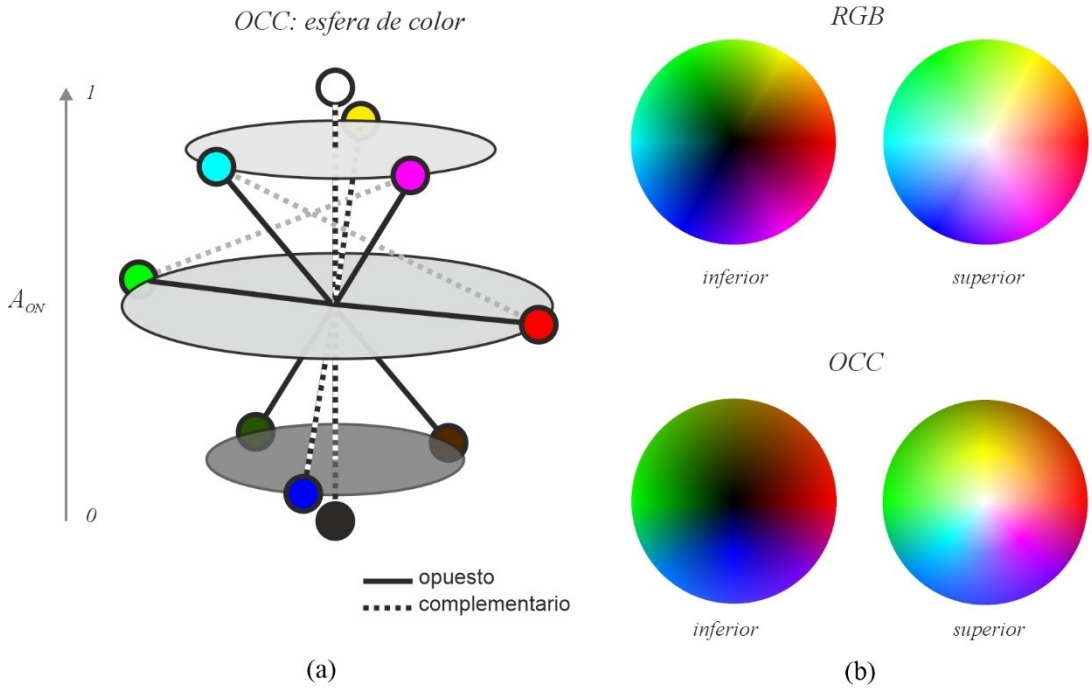


Figura 49 Esfera de color del sistema OCC.
Nota: (a) colores primarios y secundarios, con sus opuestos y complementarios en la esfera de color del sistema OCC; y (b) Comparación entre el sistema OCC y RGB.

por lo tanto, como sistema de color, posiciona cada color en una escala jerárquica que representa su actividad neuronal.

En la Figura 49.a, hemos situado cada color primario y secundario en la esfera de color incorporando su complementario y opuesto, tal y como los define la Figura 50. El verde y el rojo ocupan el valor de 0.5 (1/2), mientras que el amarillo, cian y magenta un 0.75 (3/4) y el azul, 0.25 (1/4), el resto de las relaciones de los rangos se pueden ver con mayor claridad en la Figura 49.b , donde se compara el espacio de color RGB y el OCC propuesto. Por ejemplo, en RGB, el amarillo, el cian y magenta se sitúan a la misma altura del rojo y el verde, junto con el azul, en el ecuador, con un valor de 0.5. En OCC, el amarillo, el cian y el magenta se encuentran en la parte superior con un valor de 0.75.

Modulación e inhibición

Otra funcionalidad importante en el NGL es la de la modulación y la inhibición. Aunque en el NGL estas funcionalidades se realizan en todos los canales de ambas vías, en el modelo sólo se va a aplicar en el mapa de pesos visuales con la finalidad de simplificar la computación, ya que no hay diferencias sustanciales de hacerlo por separado. El objetivo de la modulación es ajustar el valor actual con el valor anterior en el proceso de escaneado, es decir, que, si un campo receptivo x tenía un valor alto y en el nuevo estado tiene uno bajo, la funcionalidad es incrementar el estado actual hacia el anterior. Al contrario, si es mayor.

El resultado de aplicar la modulación e inhibición en el mapa de pesos visuales es el mapa de agudización:

$$AG^{h}_{i',j,a,t} = \left(PV^{h}_{i',j,a,t} - mod\left(PV^{h}_{i',j,a,t-1} - PV^{h}_{i',j,a,t}\right)\right) * I \tag{30}$$

siendo $mod(x)$ una función sigmoide en el intervalo $[-1,1]$ que permite un crecimiento y decrecimiento exponencial :

$$mod(x) = \tau * 2\left(\frac{1}{1+e^{-3x}} - 0.5\right) \tag{31}$$

donde $\tau \leq 1$ es un coeficiente positivo que indica la intensidad de aplicación de la modulación con x en el intervalo $[-1,1]$ La modulación es positiva cuando el valor anterior $PV^{h}_{i',j,a,t-1}$ es superior y negativa, al contrario, permitiendo que el valor actual se acerque en el intervalo $[0,\tau]$ al valor anterior.

Por último, I es el valor de inhibición, siendo 0 cuando hay inhibición y 1 cuando no la hay:

$$I = \begin{cases} 1, & ER^{h}_{p,q} = 0 \\ 0, & ER^{h}_{p,q} > 0 \end{cases} \tag{32}$$

estando la posición de $AG^{h}_{i',j,a,t}$ del espacio visual $V_{r,\theta,c}$ en $ER^{h}_{p,q}$. Se asume que un $ER^{h}_{p,q}$ sin valor de prominencia es igual a 0. La Figura 51 muestra un ejemplo con dos intervalos de escaneo de la imagen (Figura 51.a): el primer escaneo, donde no hay inhibición y modulación (Figura 51.b.), y el segundo escaneo con la inhibición y modulación producida por el escaneo anterior (Figura 51.c). Para este segundo, la figura muestra cinco mapas AG con valores de τ: 0, 0.25, 0.50, 0.75 y 1. La finalidad de esta figura es visualizar cómo varía el mapa AG con una intensidad de modulación concreta. En la intensidad máxima de τ, la modulación supera el estado actual por encima del anterior. Se puede visualizar comparando las regiones no inhibidas del AG del primer escaneo con el AG del segundo mapa (hay que tener en cuenta que la parte central de este mapa está inhibida). Esto facilita que la memoria de estados anteriores se mantenga con mucha intensidad en los siguientes. Por otro lado, con una modulación con τ=0, la memoria anterior queda eliminada. Si se compara el AG del primer escaneo con el segundo escaneo con τ=0 vemos que no hay ningún rastro del primer mapa. En los valores de τ bajos, como 0.25, el AG se mantiene una combinación entre el estado actual y el anterior más equilibrada, y a partir de 0.50, el estado actual es anulado por el anterior. En casos de mucho contraste entre el estado actual y el anterior, si τ es cercano a 1, en valor del anterior no sólo anula al actual, sino que incrementa su intensidad (ver en el mapa con $\tau = 1$ de la Figura 51.c, por ejemplo, el brazo de la parte superior izquierda de la región central).La Figura 52 muestra un esquema de la secuencia completa para la obtención del mapa de pesos visuales y el de agudización.

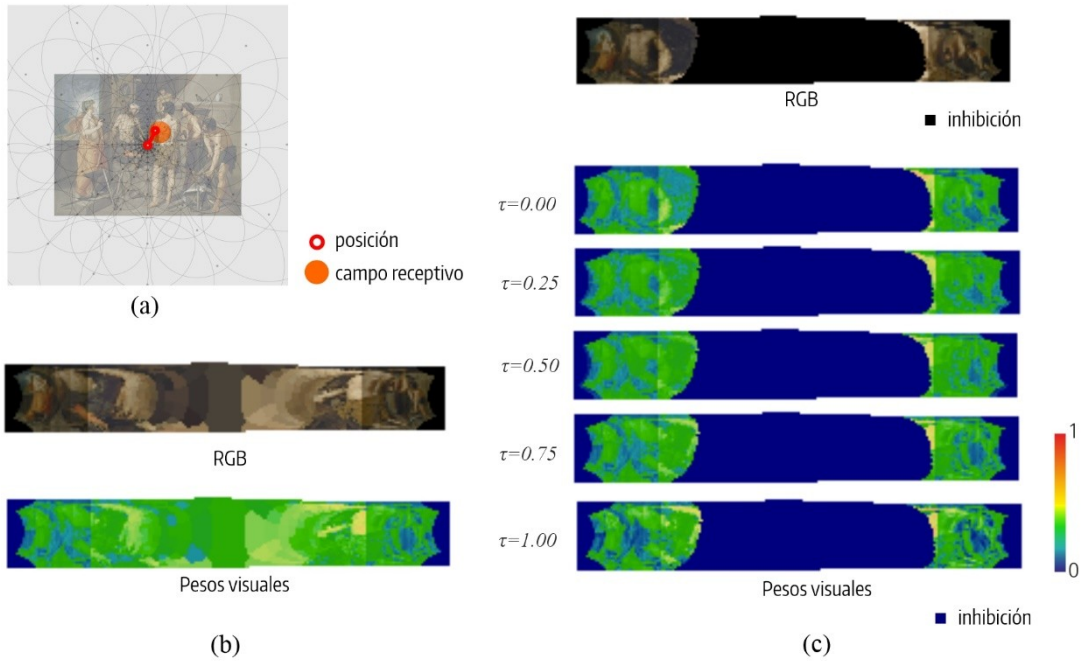


Figura 51 Ejemplos de la modulación en el mapa de pesos visuales.
Nota: (a) esquema de referencia del espacio visual; (b) mapa primero, donde no hay modulación e inhibición; y (c) mapa segundo, con la inhibición de la región central y cinco tipos de modulación.

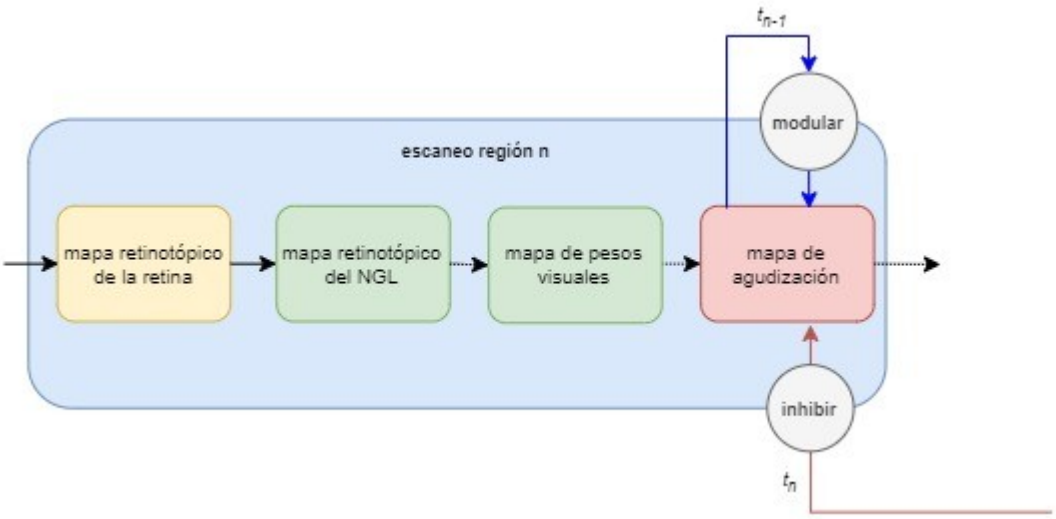


Figura 52 Secuencia de mapas y funciones en el escaneo de una región.

4.2.4 Agudizar y nivelar

A. Descripción de la tarea, del procesamiento visual y de la respuesta esperada en el espectador

La tarea tiene como finalidad la implementación de la agudización y la nivelación de la sintaxis visual. Por un lado, una región tiende a destacar del resto y, por el otro, a encontrar un equilibrio. Esta funcionalidad, aparentemente contradictoria, genera el dinamismo en la composición que el espectador explora en su escaneo y que el creador ha dispuesto al seleccionar los elementos visuales en cada región. En este sentido, una composición debe, por un lado, agudizar algunas regiones y, por el otro, equilibrarlas, como lo haría una balanza de pesos, moviendo el punto de equilibrio o modificando los pesos.

La relación entre la agudización y la nivelación con la atención es estrecha al igual que con el movimiento de ojos, lo que nos lleva a conectar este proceso con una tarea de localizar regiones que captan la atención (agudización) y equilibrar su atracción ajustando su peso visual con el resto de las regiones (nivelación).

B. Los retos principales en relación con las limitaciones físicas, algorítmicas, temporales o psicológicas e las que dependen las composiciones de las imágenes.

El principal reto es la selección de los campos receptivos más prominentes por su peso visual y cómo relacionarlos entre ellos para buscar un equilibrio. Esto plantea un reto computacional importante, ya que hay que implementar una función que, por un lado, destaque una región sobre el resto y posteriormente, otra, que encuentre un equilibrio. Otro aspecto, es la evaluación de valores en vías ON y OFF, que son opuestos, en una misma escala, ya que, por ejemplo, si una región es relevante por ser blanca, otra lo de-

be ser en igualdad de condiciones, por lo contrario, por ser negra. Es decir, la selección de qué vía es predominante en cada paso del escaneo y cómo combinar ambas es uno de los principales retos.

La principal limitación está en el proceso de escaneado donde hay que elegir el paso siguiente, que es la región en la que fijar la atención, a partir de un mapa que representa a una región previamente agudizada en un paso anterior. Esto implica seleccionar regiones que en la composición tienen un papel secundario o meramente circunstancial, pero que son necesarias de explorar para localizar las más relevantes.

C. Solución biológica del problema en relación con el comportamiento psicológico

La teoría de la atención es un área de estudio de la psicología y la neurociencia que investiga sobre los procesos cognitivos que permiten al cerebro enfocarse en ciertos estímulos ignorando otros. No hay un consenso para una definición única, aunque existen enfoques que tratan de explicar sus causas y efectos. Las corrientes más relevantes son: la atención selectiva de Posner (Posner, 1993) y Broadbent (Broadbent, 2013), la atención dividida de Bennett y Flach (Bennett & Flach, 1992) y la atención sostenida de Davies y Parasuraman (Davies & Parasuraman, 1982). Desde el punto de vista del procesamiento y control, la teoría de la atención establece dos tipos: el procesamiento de arriba hacia abajo, y de abajo hacia arriba. Lo habitual es que ambos convivan y que entre ellos exista inhibición y competencia. Para la percepción visual, existe una relación entre la atención y el lugar donde se mira que puede ser causa de los dos tipos de procesamientos: bien porque es el campo visual quien inicia el proceso de atención en el sujeto (abajo-arriba) o bien porque el sujeto fija su mirada con intencionalidad (arriba-abajo). Debido a esta relación entre mirada y atención, el estudio de los movimientos oculares ha tenido un importante desarrollo como indica Wade en su estudio histórico (Wade, 2010).

La agudización y la nivelación son dos conceptos opuestos, que tienen que ver con lo sorpresivo (agudización) y lo previsible (nivelación), es decir, entre lo que provoca tensión, el primer caso, o lo que se sitúa dentro de un equilibrio, el segundo. De hecho, esa relación de fuerzas no sólo depende de los elementos visuales y del lugar que ocupa cada región en la imagen, sino que el propio espacio visual establece diferencias entre el centro y el exterior de la composición, que Arnheim define como focos de atracción y en donde la composición debe crear, además, un equilibrio «centro-exterior» (Arnheim, 1983).

Otro aspecto relevante son las vías ON y OFF de la retina y el NGL y su relación con el contraste figura/fondo. La Figura 53 muestra un ejemplo de dos composiciones opuestas: en la primera aparece un cuadro blanco sobre fondo negro y en la segunda un cuadro negro sobre fondo blanco. La Figura 53.a establece el cuadro blanco como la figura en un fondo negro usando la vía ON y la Figura 53.b, al contrario, un cuadro negro como la figura en un fondo blanco usando la vía OFF. El sistema de percepción selecciona una vía, pero nunca ambas a la vez, lo cual facilita la detección de un cuadro blanco o uno negro.

La selección de qué vía (ON u OFF) es preponderante en cada escaneo del espacio visual —sólo puede haber una— permite obtener regiones prominentes en una vía o en la otra. Por ejemplo, detectar como prominente un cuadro blanco y uno negro sobre un fondo neutro en gris: si sólo se usara la vía ON, la escala jerárquica sólo permitiría obtener el cuadro blanco y no el negro, ya que, evidentemente, en esa escala tendría un valor de cero. En ambos casos, el ser prominentes depende del contraste de cada uno con su entorno, pero el proceso que utiliza el cerebro para seleccionar una vía u otra no está muy claro.

D. Comparación con otras soluciones de visión artificial

No existen algoritmos previos que seleccionen la vía ON u OFF «más conveniente en cada momento de un proceso de escaneado». Los denominados campos receptivos extraclásicos como describen Solomon y colaboradores (Solomon y otros, 2002) o Jeffries y colaboradores (Jeffries y otros, 2014) se han utilizado con anterioridad para la segregación de figura-fondo como en el trabajo de Ghosh y Pal (Ghosh & Pal, 2010), aunque usando una convolución en cada píxel de la imagen, y no como función de relación entre el valor local de una región y el resto de la imagen. Funcionalmente, para implementar los campos receptivos extraclásicos se utiliza DoG como en los campos receptivos clásicos, añadiendo una tercera gaussiana con un radio mayor tal y como indica Ghosh y Pal.

En los años 80, Koch y Ullman elaboraron una teoría de la atención relacionada con la percepción para describir cómo se producía el escaneo del espacio visual, y, de una manera inconsciente, los ojos se dirigían a las regiones más relevantes de ese espacio. En «Shifts in selective visual attention: towards the underlying neural circuity» (Koch & Ullman, 1987), se describen estructuras neuronales simples que son capaces de desarrollar procesos de atención, y con esta bioinspiración crearon un modelo dentro de la visión artificial que denominaron «mapa de prominencia».

En su modelo basado en la atención, para seleccionar en cada región la característica prominente, aplican el algoritmo «*The winner takes it all*» (el ganador se lo lleva todo) donde todas las características de las regiones compiten por «conseguir la atención», y sólo una región con unas características lo puede conseguir.

Figura 53 Las vías ON y OFF en la relación con la figura y el fondo.
Nota: (a) el cuadro es una figura en la vía ON; y (b) el cuadro es una figura en la vía OFF.

E. Diseño de la solución

Selección de la vía predominante ON u OFF

A partir de los planteamientos de la psicología del arte en relación con la figura/fondo, la propuesta para resolver esta funcionalidad, sin un conocimiento claro de cómo se realiza en el cerebro, es evaluar si en la vía actual (ON u OFF) la región central del mapa NGL es inhibida o no por su entorno, es decir, si es figura o es parte del fondo. Si no es inhibida en la vía, por ejemplo, ON, es que existe un contraste figura-fondo, y entonces el centro actúa como figura y el entorno como fondo, si eso no sucede, es que el centro es figura en la vía OFF. Para poder evaluar si el centro es figura o no, necesitamos dos valores para comparar: el del centro sin inhibición del entorno, que denominamos como $C_{a,t}$, siendo $a \in \{ON, OFF\}$, y el del centro con la inhibición del entorno que denominamos $E_{a,t}$. La diferencia entre el centro sin inhibición del entorno y con inhibición sería:

$$D = |C_{a,t} - E_{a,t}| \tag{33}$$

donde el valor de la diferencia se encuentra en el intervalo [0,1] y se tiene que confirmar que:

$$D < f \tag{34}$$

siendo f un umbral mínimo en el que el centro es figura. Cuando se supera este umbral, el centro en la vía a que se compara actúa como figura, y si es inferior, no, y, por lo tanto, predomina la vía opuesta, donde sí es figura. El valor del centro $C_{a,t}$ se obtiene con la convolución de los valores de $AG^{h}_{i',j,a,t}$ usando una función gaussiana:

$$C_{a,t} = \sum_{i'=0}^{N}\sum_{j=0}^{M} AG^{dcha}_{i',j,a,t} * Gauss(x', y') + \sum_{i'=0}^{N}\sum_{j=0}^{M} AG^{izq}_{i',j,a,t} * Gauss(x', y') \tag{35}$$

siendo N el total de las filas i' y M de las columnas j, donde x' e y' son las posiciones cartesianas equivalentes de j e i', respectivamente, y a la vía ON y OFF. Gauss es una función gaussiana en dos dimensiones:

$$Gauss(x, y) = \left(\frac{1}{2\pi\sigma_1^2} e^{\frac{-x^2+y^2}{2\sigma_1^2}}\right) \tag{36}$$

El valor del centro $E_{a,t}$ se obtiene con la convolución de los valores de $AG^{h}_{i',j,a,t}$ usando una función de campo receptivo extraclásico:

$$E_{a,t} = \sum_{i'=0}^{N}\sum_{j=0}^{M} PV^{dcha}_{i',j,a,t} * exDoG(x',y') + \sum_{i'=0}^{N}\sum_{j=0}^{M} PV^{izq}_{i',j,a,t} * exDoG(x',y') \quad (37)$$

siendo N el total de las filas i' y M las unidades en cada columna j, donde x' e y' son las posiciones cartesianas equivalentes de j e i', respectivamente, y a la vía ON y OFF. La función extraclasica (exDoG) añade una tercera gaussiana a la función DoG para ampliar el entorno lo cual facilita la detección del contraste entre la región central y su entorno más cercano:

$$exDoG(x,y) = \left(\frac{1}{2\pi\sigma_1^2}e^{\frac{-x^2+y^2}{2\sigma_1^2}}\right) - \left(\frac{1}{2\pi\sigma_2^2}e^{\frac{-x^2+y^2}{2\sigma_2^2}}\right) + \left(\frac{1}{2\pi\sigma_3^2}e^{\frac{-x^2+y^2}{2\sigma_3^2}}\right) \quad (38)$$

donde $\sigma_1 = \frac{1}{3} * \sigma_2$ y $\sigma_3 = \frac{M}{2}$ y el valor σ_2:

$$\sigma_2 = \rho * \sigma_3 \quad (39)$$

siendo ρ un parámetro configurable en el intervalo [0,1] que controla el tamaño del centro y entorno. Se usa $\frac{1}{3}$ en la relación de σ_1 con σ_2, de tal manera que el valor máximo de $\rho = 1,\ \sigma_1 = 0.33$. Se debe cumplir que $\sigma_1 < \sigma_2 < \sigma_3$.

Agudización: el ganador se lo lleva todo (WTA)

En el trabajo de Koch y Ullman (Koch & Ullman, 1987) se aplica el algoritmo del «ganador se lo lleva todo» (WTA) para seleccionar la región más prominente. En el mapa de pesos visuales, $AG^{h}_{i',j,a,t}$, la selección del campo receptivo más agudizado (CRA) se resuelve con la aplicación del algoritmo WTA:

$$CRA_t = \text{argmax}\left(AG^{h}_{i',j,a,t}\right) \quad (40)$$

donde CRA_t sería el campo receptivo ganador en el intervalo t de $AG^{h}_{i',j,a,t}$.

Nivelación: la prominencia de la región

La nivelación es un proceso donde una región agudizada se equilibra con el resto de las regiones. Por lo tanto, la prominencia de la región del espacio visual donde se encuentra el CRA depende del valor de actividad de este campo en relación con el resto de las regiones. En este sentido, para calcular la nivelación del CRA_t hay que usar el mapa de

pesos visuales que se genera en el siguiente paso de escaneo $t+1$: $PV^{h}_{i',j,a,t+1}$, ya que, en este mapa, el centro es el CRA_t.

Para relacionar el centro del mapa con el resto de las regiones, el planteamiento es aplicar una gaussiana que incluya a todo el mapa para que el valor de cada región se vea condicionado por la arquitectura excéntrica y anisotrópica del mapa retinotópico. Este valor de actividad nivelado será el valor de prominencia del campo receptivo del esquema de referencia $ER^{h}_{p,q}$ donde se encuentra el CRA_t. Por lo tanto, el valor de nivelación se calcula aplicando la convolución a partir de los valores de $PV^{h}_{i',j,a,t}$:

$$ER^{h}_{p,q} = \sum_{i'=0}^{N}\sum_{j=0}^{M} PV^{dcha}_{i',j,a,t} * Gauss(x',y') + \sum_{i'=0}^{N}\sum_{j=0}^{M} PV^{izq}_{i',j,a,t} * Gauss(x',y') \quad (41)$$

siendo N el total de las filas i' y M las unidades en cada columna j, donde x' e y' son las posiciones cartesianas equivalentes de j e i', respectivamente, y a es la vía preponderante en la que agudizó CRA_t. Por consiguiente, el valor final depende de los valores de todos los demás campos receptivos del mapa según su cercanía o lejanía en la estructura del mapa retinotópico.

Estructura final

La Figura 54 amplía el esquema de la Figura 52, incluyendo la agudización y la nivelación. Desde el esquema se visualiza los dos principales procesos: agudización y nivelación, la relación entre ellos y los objetivos: obtener la nueva de región a escanear desde la situación actual y el valor de prominencia del CRA anterior a partir del mapa que lo representa como centro.

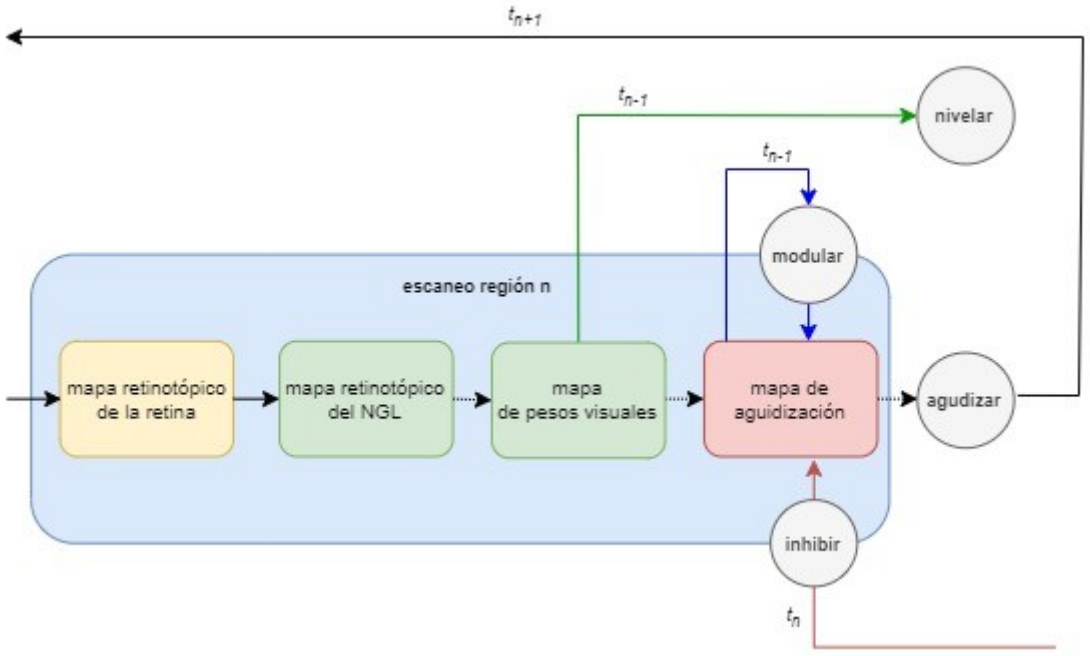


Figura 54 Mapas del escaneo con la agudización y la nivelación.

4.2.5 Escanear el espacio visual

A. Descripción de la tarea, del procesamiento visual y de la respuesta esperada en el espectador

La finalidad de esta tarea es la creación del algoritmo que procesa el escaneo del espacio visual, con una imagen como entrada y el mapa de prominencia del tejido interno como salida. La agudización y la nivelación se relacionan con un proceso iterativo de escaneo del espacio visual, donde cada movimiento tiene como fin obtener una región agudizada que es la siguiente a escanear, de tal manera que el proceso es guiado por sí mismo.

B. Los retos principales en relación con las limitaciones físicas, algorítmicas, temporales o psicológicas e las que dependen las composiciones de las imágenes

El principal reto es conseguir un equilibrio entre localizar regiones prominentes y obtener una representación global del tejido interno (con prominencias bajas, medias y altas). Además, facilitar el escaneo de cualquier región y no repetir regiones escaneadas con anterioridad.

La principal limitación tiene que ver con la arquitectura de los esquemas de referencia, la cantidad de campos receptivos y la cantidad de iteraciones necesarias.

C. Solución biológica del problema en relación con el comportamiento psicológico

El movimiento de ojos tiene una relación directa con el escaneo del espacio y su sincronización con los movimientos de la cabeza y el cuerpo (aspectos como la rotación y translación de las órbitas o la sincronización con la cabeza especialmente). Émile Javal introdujo el término «saccades» referido a los movimientos rápidos que realizan los ojos durante el escaneo del espacio visual, saltando de una región a otra (Javal, 1878). Este tipo de movimiento es complementado con las denominadas «fijaciones», que son las paradas que realizan los ojos en «centros de interés», planteadas por Busswell (Busswell, 1935). En cuanto a las imágenes y su relación con el movimiento de ojos, Stratton indica (Stratton, 1902) que los trazados que realizan los ojos en el escaneo de las imágenes que, a diferencia de la lectura de texto donde existen movimientos organizados en una misma dirección, pasan de una región otra de una manera irregular, siendo interrumpidos por algunos instantes, como descansando en algunas regiones concretas Existe una relación entre el movimiento de ojos con la percepción y los intereses cognitivos de alto nivel con procesos inconscientes de bajo nivel tal y como lo describió Stratton posteriormente (Stratton, 1906).

Buswell se centraría en los estudios de observación de sujetos experimentales mientras contemplaban imágenes (Busswell, 1935), para registrar la duración y posición de las fijaciones, tanto en imágenes simples como en complejas. Descubrió que existían «centros de interés» donde se concentraban las fijaciones. Posteriormente, contabilizó las medias en los tiempos de fijación entre personas formadas en arte y sin formar, entre

	Inicio	BIO	A/B	B/A	IN	Tipo	EX
CRISP	aleatorio	sí	no	sí	sí	simulador	no
SWIFT	aleatorio		no	sí	sí	aplicación	sí
EMMA	objetivo	sí	no	sí	no	aplicación integrada	sí
BU*TD	mapa	no	sí	sí	no	aplicación	sí

Tabla 3 Comparativa de simuladores de movimientos de ojos.
Nota: BIO: bioinspirado; A/B: procesamiento de Arriba hacia abajo; B/A: procesamiento de abajo hacia arriba; IN: inhibición; y EX: excentricidad.

niños o adultos, concluyendo que no existían diferencias concluyentes. Un punto importante de su trabajo fue la relación que existía entre la propia imagen y esos «centros de interés», es decir, qué efecto tenía una composición concreta en cuanto al movimientos de ojos, determinando que era menor que de la que se asumía.

Posteriormente, Yarbus demostró en un experimento en el que un sujeto observaba la misma imagen con diferentes tareas indicadas (Yarbus, 1967), que los trazados realizados por los ojos variaban, lo que le llevó a concluir que procesos de arriba-abajo y de abajo-arriba intervenían en el control de la percepción.

D. Comparación con otras soluciones de visión artificial

El interés por obtener los mapas de prominencia a través de la simulación de los movimientos de los ojos en visión artificial se inició con la creación de modelos de lectura automática de textos. En este área, destaca el trabajo de Underwood y colaboradores (Underwood y otros, 2006), que es donde tuvo una mayor aplicación en la primera década del siglo XXI.

El sistema más destacado es E-Z Reader realizado por Reichle y colaboradores (Reichle y otros, 2003). El objetivo de este trabajo fue la simulación del movimiento de ojos en la lectura para el reconocimiento de textos, pero tras su implementación, se estableció como un modelo eficaz que ha influenciado también a aplicaciones relacionadas con la inspección visual como: SWIFT (Engbert y otros, 2005) , CRISP (Nuthmann y otros, 2010), EMMA (Salvucci, 2001) o BU*TD (Peters & Itti, 2007). Su arquitectura la podemos resumir en tres elementos principales:

- Sistema de atención, que controla el tiempo, la información obtenida en cada momento y el proceso en general.
- La identificación de palabras que asocia los patrones obtenidos.
- El control ocular, donde está el generador de movimientos sacádicos del ojo.

A esta estructura hay que unir el funcionamiento del ojo, el ángulo de visión y el espacio del texto.

El desarrollo de modelos y simuladores a partir de E-Z Reader fue amplia en la década pasada, la Tabla 3 muestra cuatro modelos que ya hemos citado con sus características más relevantes:

- La iniciación del proceso.
- Si es bioinspirado.
- Si usan flujos abajo hacia arriba y arriba hacia abajo.
- Si utilizan la inhibición.
- Si es un simulador.
- Sí utilizan la excentricidad.

Hay dos cuestiones relevantes: el inicio aleatorio y la excentricidad. Exceptuando BU*TD (que utiliza los dos tipos de flujos), todos usan flujos de abajo hacia arriba, ya que la simulación del movimiento de los ojos en la lectura debe ser guiada por el propio texto, salvo que se estuviera intentando localizar una palabra o un trozo de texto concreto. Algunos trabajos como el de Brandt y Stark (Brandt & Stark, 1997) ahondan en el movimiento espontáneo de los ojos. Todos están enfocados al escaneo, siendo esta quizás la primera opción de un modelo que simula el movimiento de ojos. Por último, en relación con la excentricidad, como forma de representación del espacio visual, SWIFT sí lo plantea y EMMA desarrolla un sistema parecido a través de una distribución gaussiana en torno al centro del espacio visual.

E. Diseño de la solución

En relación con la experiencia del espectador, la visión artificial ha construido soluciones bioinspiradas a través de simuladores, como se ha visto en el caso de la lectura de textos planteado por Rayner y colaboradores (Rayner y otros, 1998). Además, este tipo de solución ha tenido éxito en otras tareas relacionadas con el procesamiento de imágenes. Por consiguiente, la implementación más adecuada es la de un simulador de movimiento de ojos para el escaneo del espacio visual.

Para construir este simulador, hay que implementar un proceso de escaneo iterativo, que se inicie en un punto del espacio visual y finalice cuando el valor de agudización sea inferior a un umbral mínimo de agudización. Si este umbral es igual a 0, el final de la iteración estará cuando todas las regiones del espacio visual tengan su valor de prominencia y el mapa de pesos visuales esté inhibido en su totalidad. La decisión del punto de inicio es una cuestión que ha tenido varios enfoques, como se puede comprobar en la Tabla 3, donde CRISP y SWIFT es aleatorio y EMMA, selecciona un objetivo y BU*TD analiza toda la imagen para seleccionar la región más prominente. Arnheim establece en su marco estructural el centro geométrico como el origen y final de cualquier relación en la composición. A partir de este criterio de Arnheim, el planteamiento más adecuado es que el proceso de iteración se inicie en el centro geométrico del espacio visual y desde ahí se seleccione la primera región a escanear.

El **Algoritmo 1** describe el proceso completo desde la entrada (el espacio visual donde está la imagen) hasta la salida (el mapa de prominencia). Hay dos posibles rutas, una para la iniciación y otra para las iteraciones. En la primera ruta, primer escaneo, no hay ningún CRA y el mapa de la retina está centrado en el espacio visual. En esta ruta, el objetivo es obtener el primer CRA, cuya posición será el centro del mapa de la retina para el siguiente escaneo. En este caso, el peso visual no tiene ni modulación, ni inhibición. En la otra ruta, si hay modulación e inhibición, y el mapa de actividad se utiliza para la nivelación. Posteriormente, en ambas rutas, se determina la vía preponderante, ON u OFF, y se localiza el CRA. Finalmente, se comprueba si el valor de agudización del CRA es superior a un umbral mínimo y, si es así, el proceso comienza de nuevo y, si no lo es, finaliza el escaneo y se devuelve el esquema de referencia como el mapa de prominencia del tejido interno.

Los subprocesos del algoritmo que corresponden a las subtareas que hemos analizado en el modelo. Tiene la siguiente estructura de entrada, función y salida:

- **Los mapas retinotópicos de la retina y el NGL**. Las arquitecturas de los mapas deben facilitar la existencia de dos focos de atención (el central y el externo), la preponderancia de la región inferior izquierda relacionada con anisotropía entre las regiones horizontal-vertical, izquierda-derecha e inferior superior. Estos mapas tienen como centro la región que se esté escaneando y van variando en cada escaneo.
    - Entrada: el espacio visual en píxeles.
    - Salida: mapa retinotópico NGL (el de la retina se establece como un paso intermedio)
- **Esquema de referencia del mapa de prominencia.** La arquitectura debe facilitar la existencia de dos focos de atención (el central y el externo), la preponderancia de la región inferior izquierda relacionada con anisotropía entre las regiones horizontal-vertical, izquierda-derecha e inferior superior. Es el mapa que representa las prominencia del tejido interno y, además, controla las regiones escaneadas.
    - Entrada: el espacio visual en píxeles.
    - Salida: esquema de referencia donde se van cargando los valores de prominencia para cada región escaneada. Al finalizar el proceso de escaneo, es el mapa de prominencia del tejido interno.
- **Valor de actividad**. El cálculo del peso visual en cada vía ON y OFF de cada campo receptivo del mapa NGL.
    - Entrada: el mapa NGL.
    - Salida: mapa con los pesos visuales.
- **Función relé en el NGL** La aplicación de la modulación (en relación con el valor anterior que tenía el campo receptivo en el t-1) e inhibición de los campos

receptivos que se encuentran en regiones ya escaneadas con anterioridad en el esquema de referencia.

  - **Entrada:** mapa de pesos visuales actual y el mapa de agudización en $t-1$ valor de inhibición (1 si el campo receptivo está en una región sin escanear y 0, al contrario).
  - **Salida:** mapa de agudización.

- **Proceso de agudización y nivelación**. Escaneo del espacio visual relacionado con los procesos de atención de agudización y nivelación.
  - **Selección ON/OFF**. La determinación de la vía preponderante.
    - **Entrada**: mapa de agudización.
    - **Salida**: mapa de agudización en la vía preponderante
  - **Agudización**. Selección del CRA_t.
    - Entrada: mapa de agudización en la vía preponderante.
    - Salida: CRA_t.
  - **Nivelación**. Cálculo del valor de prominencia del CRA_{t-1} a partir del mapa de pesos visuales.
    - Entrada: mapa de pesos visuales en t del CRA_{t-1}.
    - Salida: valor de nivelación que representa la prominencia del campo receptivo en el esquema de referencia donde está la región del espacio visual que representa el CRA_{t-1}.

Algoritmo 1. Proceso de escaneo

Input: $\boldsymbol{V}_{r,\theta,c}$
t=0
estado="activo"
via= "ON"
while estado es "activo"
if t=0
• Cargar mapa retina desde el centro geométrico del espacio visual:
$R^{area}_{i,j,v,o,0}$ desde $V_{0,0,c}$
else
• Cargar mapa retina desde la posición desde el CRA_{t-1}:
$R^{h}_{i,j,v,o,t}$ desde $V_{r,\theta,c}$ siendo el centro la posición en coordenadas polares en el espacio visual de CRA_{t-1}
• Cargar mapa NGL
$NGL^{h}_{i',j,v,o,t}$
• Cargar mapa de pesos visuales
if vía predominante ON

$$PV^{h}_{i',j,ON,t} = \frac{NGL^{h}_{i',j,ON,L,t}+NGL^{h}_{i',j,ON,M,t}+NGL^{h}_{i',j,ON,S,t}+NGL^{h}_{i',j,ON,LM,t}}{4}$$

else

$$PV^{area}_{i',j,OFF,t} = \frac{NGL^{h}_{i',j,OFF,-L,t}+NGL^{h}_{i',j,OFF,-M,t}+NGL^{h}_{i',j,OFF,-S,t}+NGL^{h}_{i',j,OFF,-LM,t}}{4}$$

if t=0
• Cargar mapa de pesos visuales AG:
$AG^{h}_{i',j,a,t} = PV^{h}_{i',j,a,t}$
else
• Nivelar mapa NGL y cargar valor de prominencia en el campo receptivo del esquema de referencia:
$ER^{h}_{p,q} = \sum_{i'=0}^{N}\sum_{j=0}^{M} PV^{dcha}_{i',j,a,t} * Gauss(x',y') + \sum_{i'=0}^{N}\sum_{j=0}^{M} PV^{izq}_{i',j,a,t} * Gauss(x',y')$
• Cargar mapa de agudización:
$AG^{h}_{i',j,a,t} = \left(PV^{h}_{i',j,a,t} - mod\left(PV^{h}_{i',j,a,t-1} - PV^{h}_{i',j,a,t}\right)\right) * I$
• Cambiar vía predominante ON u OFF, si la diferencia D de la convolución del mapa de agudización sin inhibición: $C_{a,t}$ y con inhibición $E_{a,t}$ es inferior a un umbral f.
$D < f$ siendo $D = |C_{a,t} - E_{a,t}|$
• Aplicar WTA en vía predominante de mapa PV y localizar al campo receptivo de PV que agudiza:
$CRA_t = \text{argmax}\left(AG^{h}_{i',j,a,t}\right)$
• Localizar campo receptivo en el esquema de referencia:
$ER^{h}_{p,q}$ donde la posición $V_{r,\theta}$ de $CRA_t \in \mathrm{CR^h_p}$ de $ER^{h}_{p,q}$
if valor de agudización < umbral mínimo de agudización
estado="finalizado"
return $ER^{h}_{p,q}$
t+1
loop

La Figura 55 muestra un esquema funcional del CRA_t del modelo Inner Fabric. Existen tres áreas funcionales principales en relación con el CRA_t: el espacio visual donde está la imagen, el centro de control donde está el esquema de referencia, y el área de escaneo donde están todos los mapas retinotópicos que tienen como centro la región que representa el CRA_t. En el espacio visual, se localiza la región del CRA_t (obtenido en el escaneo anterior) para cargar un nuevo mapa con este CRA_t como centro. Desde el mapa de actividad se nivela y el valor obtenido se carga en el campo respectivo del es-

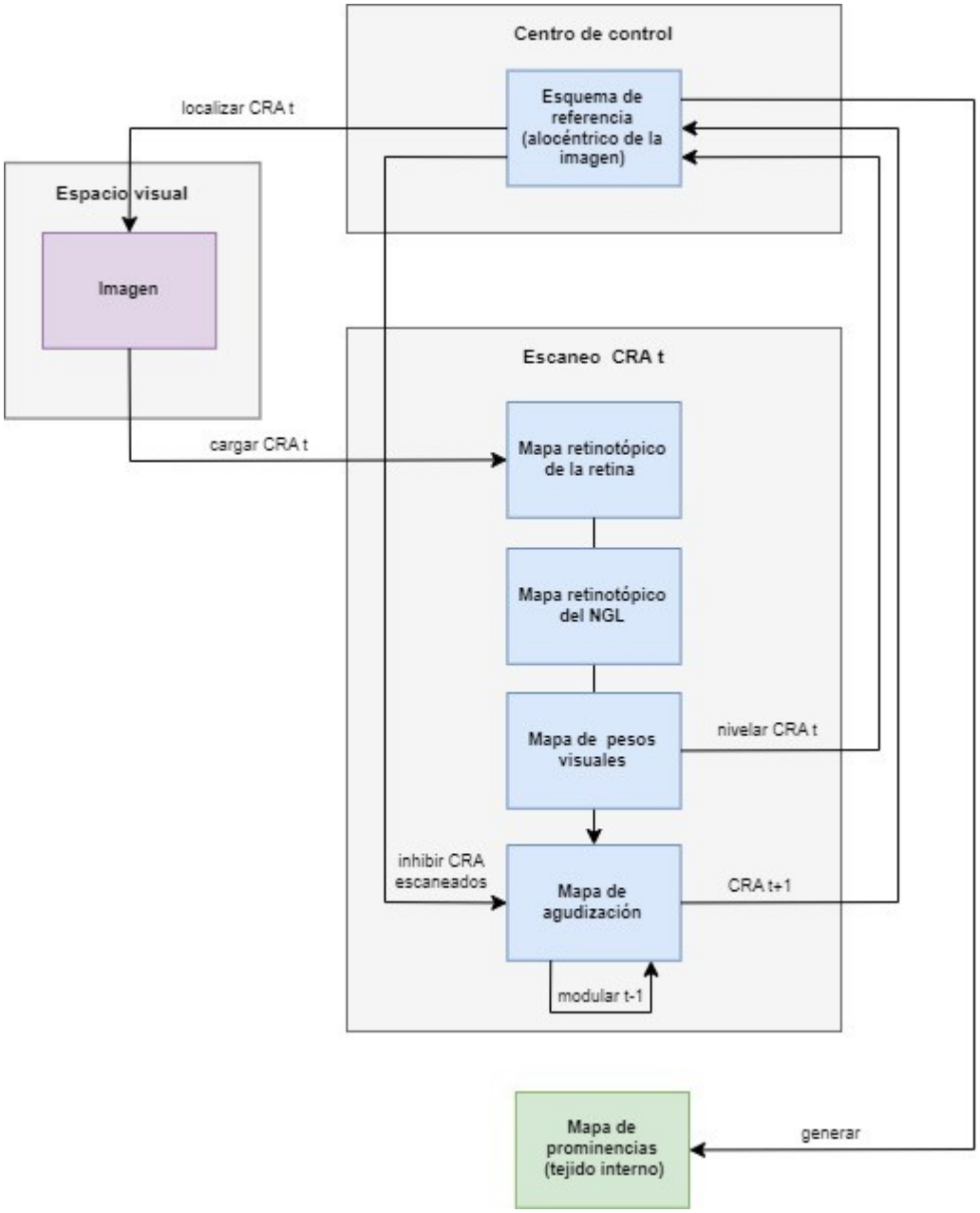


Figura 55 Esquema funcional del modelo Inner Fabric.

quema de referencia correspondiente a CRA_t. En el mapa de pesos visuales se inhiben los campos receptivos de las regiones del espacio visual ya escaneadas con los anteriores CRA. Se obtiene el CRA_{t+1} desde el mapa de agudización y el proceso comienza de nuevo.

4.3 Configuración del modelo Inner Fabric para imágenes artísticas y con criterios estéticos

Una de las cuestiones más importantes del modelo Inner Fabric es que es capaz en un mismo proceso de incluir prominencia tanto en la vía ON como en la OFF. Pero para que esta funcionalidad sea operativa es necesario ajustar los parámetros de las arquitecturas y, después, los coeficientes y pesos relacionados con el proceso de escaneado.

El espacio visual es la entrada del modelo Inner Fabric, pero, además, se mantiene como entrada en cada una de las iteraciones del proceso de escaneado. El esquema de referencia es el mapa de prominencia y, a su vez, es la herramienta de control del escaneado. ya que define el tamaño de las regiones y gestiona el proceso a través de la inhibición de lo que se ha escaneado.

El parámetro principal para ajustar las arquitecturas de los mapas retinotópicos es la cantidad de campos receptivos, N_{retina}, que establece el nivel de detalle de la representación para el escaneo y el tamaño de los campos receptivos. Por otro lado, el ancho del espacio visual, W, determina la resolución, y debe ser superior a N_{retina}. Para el NGL, el peso de cada región, w_r, implementa la preponderancia de la región inferior izquierda estableciendo el porcentaje de reducción de campos receptivos en cada cuadrante.

Para la arquitectura del esquema de referencia, el parámetro $N_{esquema}$ tiene la misma funcionalidad que en el mapa de la retina. El detalle de la información es inferior al del mapa de la retina, y los campos receptivos deben ser mayores para implementar la percepción horizontal. Por último, el peso μ_r de cada región, al igual que en los mapas retinotópicos, implementa la preponderancia por la región inferior izquierda, controlando el porcentaje de campos receptivos por cuadrante.

El coeficiente τ ajusta la intensidad de aplicación de la modulación en los campos receptivos del NGL. La memoria residual de los pesos visuales de los campos receptivos depende de este coeficiente, si es muy alto, se ajusta el estado actual demasiado a los anteriores, y a la inversa si es bajo. Encontrar un equilibrio es importante para que el escaneo sea lo más integrado posible y no resulte una colección de instantáneas independientes de las regiones del espacio visual.

En la selección de la vía preponderante, el umbral f establece el mínimo para que la región central sea figura y su entorno fondo, evaluando si el entorno es capaz de inhibir al centro, siendo el coeficiente ρ quien determina la relación del tamaño del centro y el entorno. El contraste entre figura y fondo permite la selección de regiones prominentes en ambas vías (ON y OFF), de tal manera que si la estimación del contraste figura-

fondo es muy exigente, tenderá a que una vía predomine sobre la otra y si no lo es, ambas vías predominarán por igual con independencia de las prominencia.

Por último, el umbral de agudización mínimo indica el nivel de exigencia del escaneo. Si este umbral es muy alto, es posible que sólo se escanee unas regiones concretas ya que el proceso es iterativo, localizando los campos receptivos que más agudizan en una representación del espacio visual pero centrado en una región concreta, sin que esto implique que se localicen en un orden secuencial de más a menos. Es decir, es posible que el campo receptivo que más agudiza en una iteración concreta su peso visual no sea inferior al anterior y tampoco superior al siguiente y, por lo tanto, si el umbral es alto, provocará que el proceso finalice sin localizar todas las regiones con prominencia alta.

Los parámetros para configurar son los siguientes:

- Espacio visual:
    - W, ancho del espacio visual.
- Mapas retina:
    - N_{retina}, número de campos receptivos en el eje horizontal y vertical (el total es $N_{retina} x N_{retina}$).
- Mapas NGL:
    - w_r, peso que determina la cantidad porcentual de reducción de campos receptivos de los mapas de la retina según el cuadrante (inferior- izquierda inferior-derecha, superior-izquierda y superior-derecha) en el eje vertical del mapa logarítmico.
- Esquemas de referencia:
    - $N_{esquema}$, número de campos receptivos para el eje horizontal.
    - μ_r, peso que define la cantidad porcentual de campos receptivos según el cuadrante del mapa de la retina (inferior- izquierda inferior-derecha, superior-izquierda y superior-derecha) en el eje vertical del mapa logarítmico.
- Agudización:
    - τ, coeficiente de la intensidad de la aplicación de la modulación que mantiene la memoria de los estados anteriores en el peso visual.
- Selección de vía predominante (ON/OFF):
    - f, umbral mínimo de la diferencia entre los mapas de AG sin inhibición del entorno y sin ella.
    - ρ, coeficiente que regula la relación del tamaño del centro con el entorno inhibitorio en la convolución del campo receptivo extraclásico.

4.3.1 Metodología para ajustar los parámetros, pesos, coeficientes y umbrales

4.3.1.1 Arquitectura de los mapas retinotópicos. Parámetros N_{retina} y W y peso w

Los parámetros configurables para la arquitectura de los mapas retinotópicos son dos: el valor N_{retina} que indica la cantidad de campos receptivos y el valor W que establece la cantidad de unidades máxima de r en espacio visual. Ambos parámetros tienen una incidencia relevante, tanto al nivel de detalle como el rendimiento de la computación en el escaneo, y el objetivo principal es obtener unas arquitecturas equilibradas para ambos. Para conseguir una combinación adecuada se realizarán pruebas con varios valores de N_{retina} y W hasta obtener la más satisfactoria.

Para el ajuste del peso w, se optará por una relación heurística que facilite la mayor cantidad de campos receptivos en el cuadrante inferior-izquierda y la menor en el cuadrante superior-derecho, mientras que los otros cuadrantes se ajustan en proporción en el orden de prioridad de la región izquierda sobre la derecha y de la inferior sobre la superior.

4.3.1.2 Arquitectura del esquema de referencia, parámetro $N_{esquema}$ y peso μ

Para el ajuste del esquema de referencia es importante la cantidad de campos receptivos $N_{esquema}$. Para ajustar este parámetro se probarán varios valores de $N_{esquema}$ analizando el tamaño de los campos receptivos de una posición concreta en cada cuadrante.

Para el ajuste del peso μ de cada región, se optará por una relación heurística que facilite la mayor cantidad de campos receptivos en el cuadrante inferior-izquierdo y la menor en el cuadrante superior-derecho.

4.3.1.3 Selección de vía predominante, prominencia en ON y OFF. Coeficiente ρ, relación con τ y el umbral f de la diferencia centro-entorno

Para ajustar los coeficientes ρ, τ y f es necesario analizar el comportamiento del modelo Inner Fabric con las distintas combinaciones de los tres en varias tareas relacionadas con la agudización de regiones en ON y OFF y la selección de la vía predominante. Los parámetros configurables, que ya hemos analizado y configurado, tienen una repercusión relativa en relación con el comportamiento del modelo Inner Fabric en el escaneo —afectan al nivel de detalle y al rendimiento de la computación. El coeficiente τ, que regula la aplicación de la modulación en el mapa de agudización, tiene una incidencia en la relación entre las regiones del espacio visual: una modulación alta reduce el contraste alto entre regiones y una modulación baja lo permite. Si no hay mucho contraste, el ajuste es mínimo. Su relación en la localización de campos receptivos agudizados en

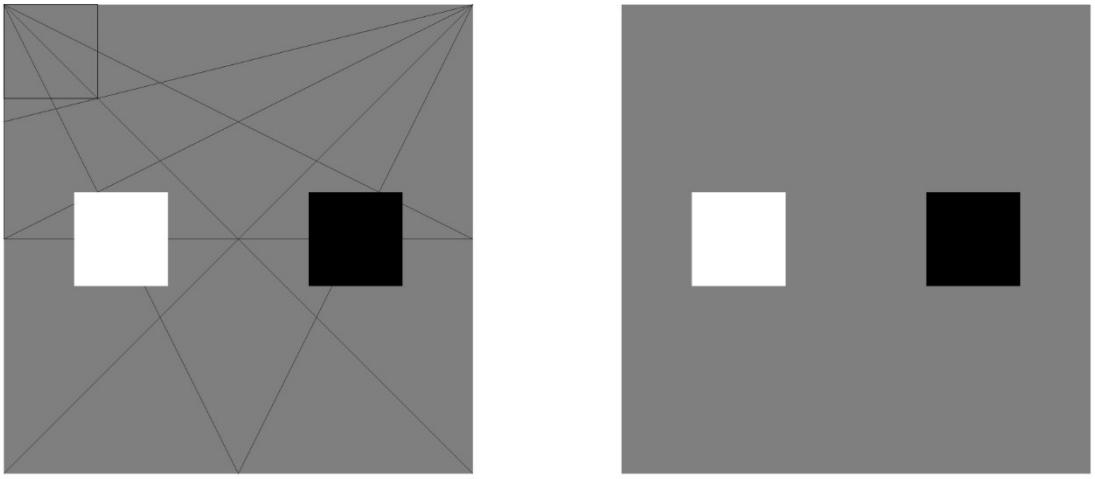


Figura 56 Imagen para el ajuste.
Nota: (a) creación de los elementos que serán figura en ON (cuadro blanco) y en OFF (cuadro negro) y su situación en un fondo neutro; y b) la imagen final.

ambas vías es importante, ya que los pesos visuales en la vía ON de una región, si son altos, tienen mucho contraste con los de la vía OFF de otra región, ya que en esta misma región su peso visual en ON, sería bajo al ser ambas vías opuestas —un valor 1 en la vía ON es 0 en la OFF y a la inversa.

El objetivo es obtener el valor de ρ (que regula el tamaño del centro y del área inhibitoria en los campos extraclásicos de los mapas NGL) en relación con τ (intensidad de la modulación de los pesos visuales de los campos receptivos de los mapas NGL), según un rango de valores f (umbral mínimo de diferencia en la función figura-fondo). Para esta finalidad, es necesario crear una imagen en donde exista una figura en cada vía en un espacio neutro y se pueda cuantificar si Inner Fabric las detecta en su vía, cuántas veces y los cambios de vía. En la Figura 56.a, la imagen está compuesta por un fondo gris neutro donde, a partir de varios ejes que relacionan las esquinas superiores con los lados laterales e inferior, se determina el tamaño de un cuadro que es situado en el eje central horizontal simétricamente: a la izquierda en blanco y a la derecha en negro. La situación del cuadro blanco en la mitad izquierda le da una mayor preponderancia al del negro por anisotropía de la representación, con lo que el objetivo es que en las combinaciones de los coeficientes obtengan la mayor similitud entre ambos en su detección, siendo siempre inferior la de OFF en relación con la de ON. Si fuera al revés, el resultado sería el mismo, pero la preponderancia sería del cuadro en la vía OFF en relación con la de la vía ON, ya que para Inner Fabric ambas vías son opuestas y se complementan. La Figura 56.b muestra la imagen final que usará en el proceso de ajuste. El uso del cuadro es para simplificar el ajuste manteniendo una relación entre las figuras (cuadro blanco y negro) y su espacio neutro que es también un cuadrado. Usar círculos, triángulos, rectángulos u otras formas irregulares, incluirían otros aspectos a la imagen que dificultaría la evaluación, por esa razón se utiliza relaciones regulares, homogéneas y simétricas. La situación de los cuadros en el eje horizontal se debe a la anisotropía de

los mapas que representan a la imagen, ya que hay una preponderancia de este eje sobre el vertical y facilitará el ajuste.

Para ajustar los tres coeficientes utilizamos los siguientes criterios secuenciales donde se van descartando las combinaciones:

1. **Similitud entre ON y OFF en la cantidad de CRA**. Si existe una superioridad en una vía sobre la otra –es una evaluación heurística donde la diferencia debe ser el doble–, la combinación de los coeficientes será desestimada, y será descartada para las siguientes pruebas.
2. **Similitud entre ON y OFF en la cantidad de CRA en la vía predominante de la figura**. Uno de los principales objetivos es que existan CRA en regiones que son figura, tanto en la vía ON como en la OFF, en este caso, cuadro blanco en ON y cuadro negro en OFF. La cantidad debe ser similar, algo superior en la vía ON ya que el cuadro se encuentra en la región izquierda y el negro en la derecha. La diferencia no debe ser alta para que la combinación de coeficientes pase al siguiente criterio.
3. **Similitud entre ON y OFF en la cantidad de CRA que no cambian de vía**. Los dos cuadros tienen dimensiones suficientes como para albergar varios campos receptivos agudizados en iteraciones seguidas en la misma vía (por ejemplo, el cuadro blanco que es figura en ON debe de obtener varios campos receptivos seguidos en la vía ON). En el espacio neutro, se debe mantener la vía preponderante en muchos CRA.

Se seleccionará la combinación de coeficientes que haya pasado los tres criterios según su orden y tengan la cantidad más similar entre ambas vías. En caso de igualdad entre varias combinaciones, habrá que realizar pruebas heurísticas en imágenes para seleccionar la más adecuada.

Para proceder, estimamos las siguientes combinaciones como referencia:

- f, umbral mínimo de diferencia en la función figura-fondo. En las pruebas realizadas previas, se ha comprobado que el intervalo de umbral óptimo aproximado es de $f = [0.1,0.3]$, por encima, no estima contraste figura-fondo y, por debajo, siempre determina que es figura. Para el análisis, se utilizarán como referencia los valores de $f \in \{0.1,0.2,0.3\}$.
- Las combinaciones entre el coeficiente τ y ρ en el intervalo $[0,1]$, y descartando los extremos 0 e 1 para evitar que en τ no se aplique la modulación si es 0 o se anule el valor actual del peso visual si es 1, y en ρ, el centro no exista si es 0 o el entorno si es 1. Los valores son:
  - $\tau \in \{0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9\}$
  - $\rho \in \{0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9\}$.

4.3.1.4 Umbral mínimo de agudización

El proceso de escaneo no es infinito, o bien localiza los campos receptivos de todas las áreas prominentes en el esquema de referencia—y por consiguiente se inhibe todo el espacio de los mapas NGL— o el valor de agudización obtenido en el escaneado de una región es inferior a un umbral mínimo de agudización. Al igual que otros parámetros, la elección de este valor depende mucho del objetivo: si somos muy exigentes con el umbral y la imagen no tiene valores altos en la región exterior, lo más seguro es que el proceso de escaneo se pare antes de analizar esta región; sin embargo, si el valor es menor, el escaneo detectará también campos receptivos prominentes con niveles bajos.

4.3.2 Ajuste de los parámetros, coeficientes, pesos y umbrales

4.3.2.1 La arquitectura de los mapas retinotópicos. Parámetros N_{retina}, W y peso w

Para configurar N_{retina} y W, se evalúan los siguientes valores de N_{retina}: 50, 100 y 200, y para W se utilizan 500 unidades con el fin de que la relación entre ambos sea: 1/10, 1/5 y un 2/5 respectivamente. Para el tejido interno, con una percepción horizontal sin detalles y enfocada en un análisis global de las regiones, 500 unidades para W es suficiente en las pruebas realizadas ya que cantidades mayores repercuten en el coste de computación sin encontrar mejoras importantes en el rendimiento. Para evaluar la resolución de la imagen y su incidencia, se utilizarán tres valores: 500px (relación de un píxel por unidad), 1000px (el doble de píxeles en relación con unidades) y 2000px (cuatro píxeles por unidad) y el resto de los coeficientes se han configurado para estas pruebas con valores medios, aunque su incidencia en los tiempos de procesamiento o cantidad de campos receptivos agudizados no es determinante.

La tabla 4 muestra los resultados para la combinación de valores de N_{retina} y la resolución en píxeles del espacio visual. El tiempo de ejecución crece exponencialmente según aumentan los campos receptivos y la resolución en píxeles del espacio visual. El aumento de campos receptivos agudizados es menor, pero pasa de los 63 en la combinación de $N_{retina} = 50$ y 500px, hasta los 92 en N_{retina} =200 y 2000px. El coste de

	Resolución		**Resolución**		**Resolución**	
	500 px		1000 px		2000 px	
N_{retina}	**CRA**	**Tiempo**	**CRA**	**Tiempo**	**CRA**	**Tiempo**
50	60	18”	63	1’04”	80	4’58”
100	63	22”	70	1’33”	84	5’47”
200	65	34”	76	1’54”	92	8’19”

Tabla 4 Estadísticas de la cantidad de CRA obtenidos según N_{retina} y la resolución.

computación con tiempos superiores a un minuto no justifica el aumento de los campos receptivos agudizados, siendo una combinación de $N_{retina} = 100$ y resolución de 500px con un tiempo de 26" la óptima, ya que la opción de $N_{retina} = 50$, obtiene una cantidad de campos receptivos agudizados inferior, en un tiempo muy cercano, 18".

Para el ajuste del peso w, en relación con la preponderancia de la región inferior izquierda, se utiliza el siguiente criterio:

$$w = \begin{cases} 1, & \pi \leq \theta_{i'} < \dfrac{3\pi}{2} \\ 0.7, & \dfrac{3\pi}{2} \leq \theta_{i'} < 2\pi \\ 0.4, & \dfrac{\pi}{4} \leq \theta_{i'} < \pi \\ 0.1, & 0 \leq \theta_{i'} < \dfrac{\pi}{4} \end{cases} \tag{42}$$

El criterio que se ha establecido es que el peso de la región inferior izquierda y el de la superior derecha estén en los extremos: 1 y 0.1 (que sería el valor mínimo posible con un solo decimal). Las otras dos regiones se ajustan a partir de este criterio de prioridad: primero, izquierda-derecha y, segundo, inferior-superior.

4.3.2.2 Arquitectura del esquema de referencia. Parámetro $N_{esquema}$ y peso μ

El esquema de referencia tiene cuatro regiones en relación con el eje vertical y horizontal, donde varía la cantidad de campos receptivos en las filas del mapa logarítmico. El primer criterio que planteamos es que $N_{esquema}$ sea múltiplo de 4 para facilitar la comparación entre arquitecturas y las diferencias en cada incremento de $N_{esquema}$, evaluando los siguientes valores $N_{esquema} \in \{8,12,16,20,24,28\}$

En segundo lugar, para configurar el peso μ, que indica la anisotropía en cada región, se aplica una reducción de campos receptivos según la región de 0.3 como máximo, para que las diferencias no sean excesivas (1/3 de diferencia entre la región inferior izquierda y la superior derecha), siendo el valor μ según el ángulo θ_p que indica en que cuadrante se encuentra:

$$\mu = \begin{cases} 1, & \pi \leq \theta_p < \dfrac{3\pi}{2} \\ 0.9, & \dfrac{3\pi}{2} \leq \theta_p < 2\pi \\ 0.8, & \dfrac{\pi}{4} \leq \theta_p < \pi \\ 0.7, & 0 \leq \theta_p < \dfrac{\pi}{4} \end{cases} \tag{43}$$

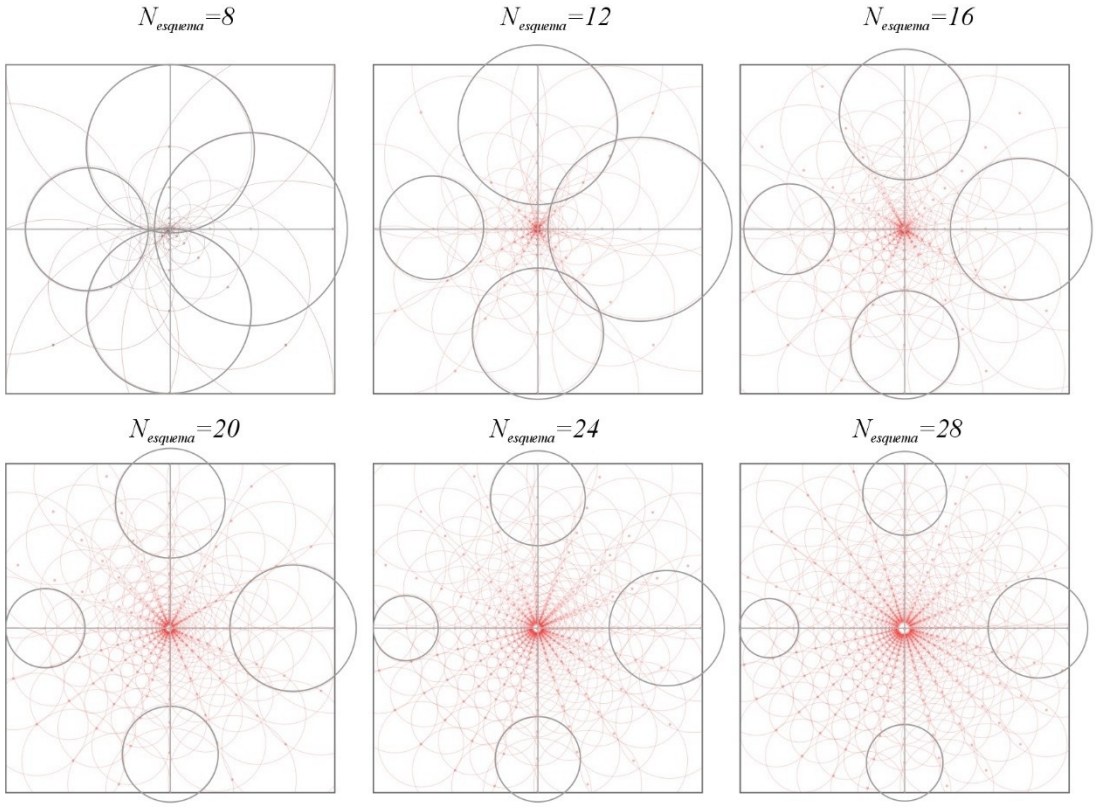


Figura 57 Arquitecturas de esquema de referencias según el valor $N_{esquema}$.
Nota: en cada eje se resalta el campo receptivo del extremo para poder evaluar las diferencias entre las variantes.

Para poder visualizar con mayor claridad y comparar las distintas arquitecturas, se selecciona los campos receptivos situados en la penúltima posición de las columnas, desde el centro en cada eje con el fin de ver la diferencia de tamaño según el cuadrante y valor de $N_{esquema}$. La Figura 57 muestra los resultados donde, si nos fijamos en los campos receptivos penúltimos en cada eje se solapan en $N_{esquema} \leq 12$ y se separan demasiado en $N_{esquema} \geq 20$. En $N_{esquema} = 16$, la arquitectura tiene un equilibrio en estos cuatro campos receptivos, ya que no se solapan, pero representan una región lo suficientemente grande para que la relación centro-exterior de Arnheim en su marco estructural esté equilibrada.

4.3.2.3 Umbral mínimo de agudización

Para obtener un mapa de prominencia en relación con una percepción global y tal como ya se ha analizado, lo más eficiente es tener un nivel bajo de umbral mínimo. Se establece en 0.1 para evitar valores cercanos a 0.

4.3.2.4 Selección de vía predominante: coeficiente ρ, relación con τ y el umbral f de la diferencia centro-entorno

Con Inner Fabric parametrizado y el coeficiente mínimo de agudización establecido, ejecutamos el escaneo en la imagen de ajuste con la combinación de valores del umbral

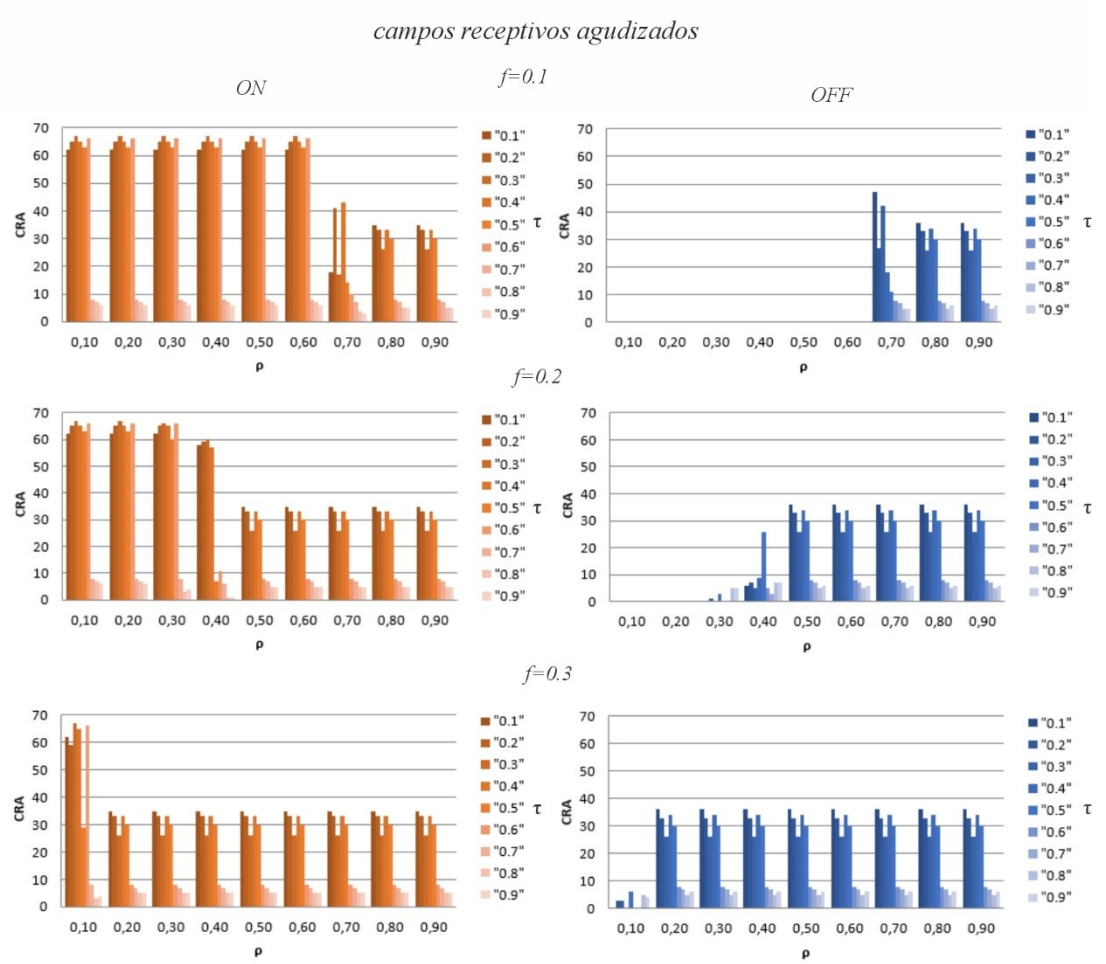


Figura 58 CRA en la vía ON y OFF para cada umbral de f según ρ y τ.

f para los conjuntos de coeficientes de τ y ρ. Realizamos las tres pruebas en el orden establecido con la finalidad de obtener una combinación final óptima.

1. Similitud entre ON y OFF en la cantidad de CRA

La Figura 58 muestra los gráficos por cada valor de f con la cantidad de campos receptivos agudizados en la vía ON y en la vía OFF (eje Y) y según las combinaciones de ρ (eje X) y de τ (valores por cada unidad ρ en el eje X)**.** Analizamos las combinaciones de ρ y τ que consiguen una cantidad de campos receptivos agudizados similares en ambas vías para cada valor de f y los intervalos de combinaciones que cumplen los criterios son los siguientes:

- Para $f = 0.1$ en $\rho \geq 0.7$ y $\tau \leq 0.5$
- Para $f = 0.2$ en $\rho \geq 0.5$ y $\tau \leq 0.5$.
- Para $f = 0.3$ en $\rho \geq 0.2$ y $\tau \leq 0.5$.

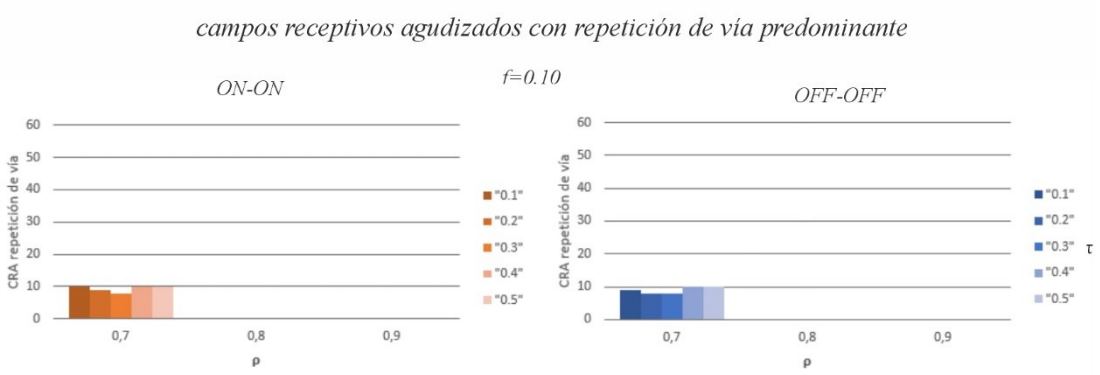


Figura 59 CRA que no cambian la vía predominante ON u OFF.
Nota: sólo con el umbral $f = 0.1$ existen valores.

2. Similitud entre ON y OFF en la cantidad de CRA en la vía predominante de la figura

En los intervalos donde existe una similitud entre los CRA en la vía ON y OFF, los campos receptivos agudizados en el cuadro blanco en la vía ON son 5 y en el negro en la OFF son 3 para todas las combinaciones seleccionadas en el apartado anterior. En este sentido no es posible eliminar ninguna combinación en el segundo paso de la evaluación.

3. Similitud entre ON y OFF en la cantidad de CRA que no cambian de vía

La Figura 59 muestra las gráficas con los resultados del porcentaje de CRA que mantienen la vía predominante para ON y OFF en f=0.1, donde únicamente con $\rho = 0{,}7$ existen campos receptivos que mantengan la vía. En f=0.2 y f=0.3, ningún CRA mantuvo la misma vía.

En relación con los valores de τ, las diferencias son mínimas entre 0.1 y 0.4, pero si observamos los datos de la Tabla 5 en el gráfico f=0.1 para $\rho = 0{,}7$ y los analizamos, existen diferencias sustanciales. La Tabla 5 muestra las cantidades de CRA según el coeficiente τ para ON y OFF. El valor de $\tau = 0.2$ tiene una diferencia en la cantidad de campos receptivos agudizados más pequeña en ambas vías con 14 campos, con 41 en la vía ON y 27 en la vía OFF.

τ	ON	OFF	diferencia	total
0.1	18	47	29	65
0.2	41	27	14	68
0.3	17	42	25	59
0.4	43	18	25	61

Tabla 5 Cantidad total de CRA según coeficiente τ para f=0.1 y $\rho = 0.7$

Combinación final.

Con estos datos podemos concluir que la mejor combinación para el umbral f y coeficientes ρ y τ en la imagen de ajuste es:

- $f = 0.1$ y $\rho = 0.7$ el valor $\tau = 0.2$.

La Figura 60 muestra el proceso de escaneo en la imagen de ajuste. Cada punto blanco corresponde a la posición de CRA en la vía ON y el negro en la vía OFF. Para facilitar la visualización, se han representado los CRA del esquema de referencia (Figura 60.a) con una escala de temperatura donde el azul son los valores altos de prominencia en la vía OFF, cian los medios y rojo los altos en la vía ON y amarillo, los medios, siendo verde para ambos, los valores bajos de prominencia. Además, se han incluido dos tipos de mapas de prominencia, uno con discriminación de las vías ON y OFF (Figura 60.b) y otro sin ella (Figura 60.c).

En el primer mapa, la prominencia en la vía ON alta se representa en el rojo y en la vía OFF en el azul, siendo el verde la baja para ambos. En el segundo, la prominencia alta es en rojo y la baja en azul. En el primer mapa, debe haber CRA en el cuadro blanco

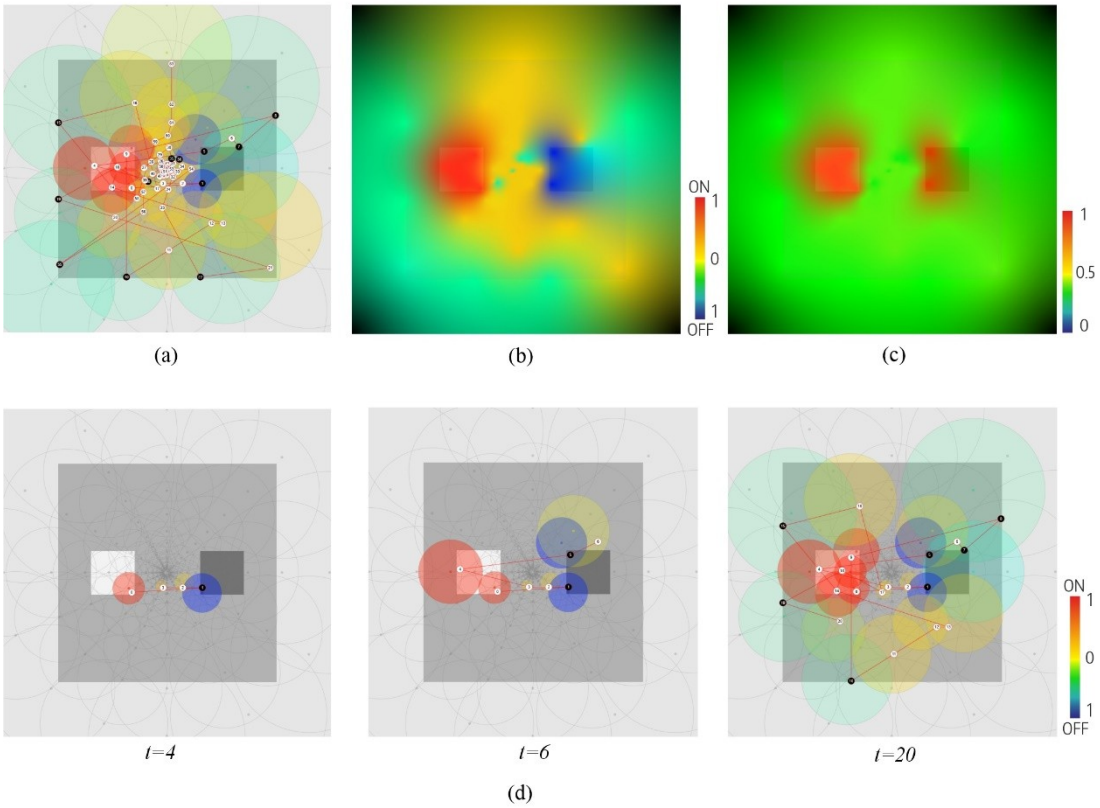


Figura 60 Análisis de la configuración final en la imagen de ajuste.
Nota: (a) esquema de referencia con el trazado; (b) mapa de prominencia del tejido interno con discriminación ON y OFF; (c) mapa de prominencia del tejido interno sin discriminación ON y OFF; y (d) secuencia de CRA en varios intervalos de la secuencia de escaneado: t=4, t=6 y t=20.

con prominencia alta en ON (rojo), en el cuadro negro en OFF (azul) y en el fondo neutro en ambos, pero con valores medios, amarillo en ON y cian en OFF. En el esquema de referencia, donde se muestra el trazado de escaneo (ver Figura 60.a), existen regiones rojas y azules en los dos cuadros, y por el trazado se puede visualizar que en la región neutra hay una mayor cantidad de CRA en la vía ON, con prominencia media en la región central y en la vía OFF, en los extremos. Esta distribución coincide con el comportamiento previsible al detectar el cuadro blanco en ON, el negro en OFF y la región neutral con ambos, aunque existe una clara predominancia de la vía ON en la región neutra. El mapa de prominencia con discriminación de ON y OFF (ver Figura 60.b) muestra claramente lo comentado, mientras que en el mapa de prominencia sin discriminación de ON y OFF (ver Figura 60.c) visualizamos con claridad que existe una prominencia alta en ambos cuadros y media en toda la región neutra.

En relación con la arquitectura de los esquemas de referencia y los mapas retinotópicos, se puede comprobar fácilmente la excentricidad y la anisotropía. En el mapa de prominencia que discrimina la vía ON de la OFF (ver Figura 60.b), existen más CRA en el cuadro blanco (cinco) que se sitúan en la región de la izquierda que en el negro (tres), que se sitúan en el de la derecha. También hay más CRA en la región inferior que en la superior, y ninguno en la esquina superior derecha. En este sentido, en igualdad de condiciones, aunque un cuadro esté en blanco y otro en negro, hay una preponderancia en el escaneo de la izquierda sobre la derecha y de la región inferior con la superior.

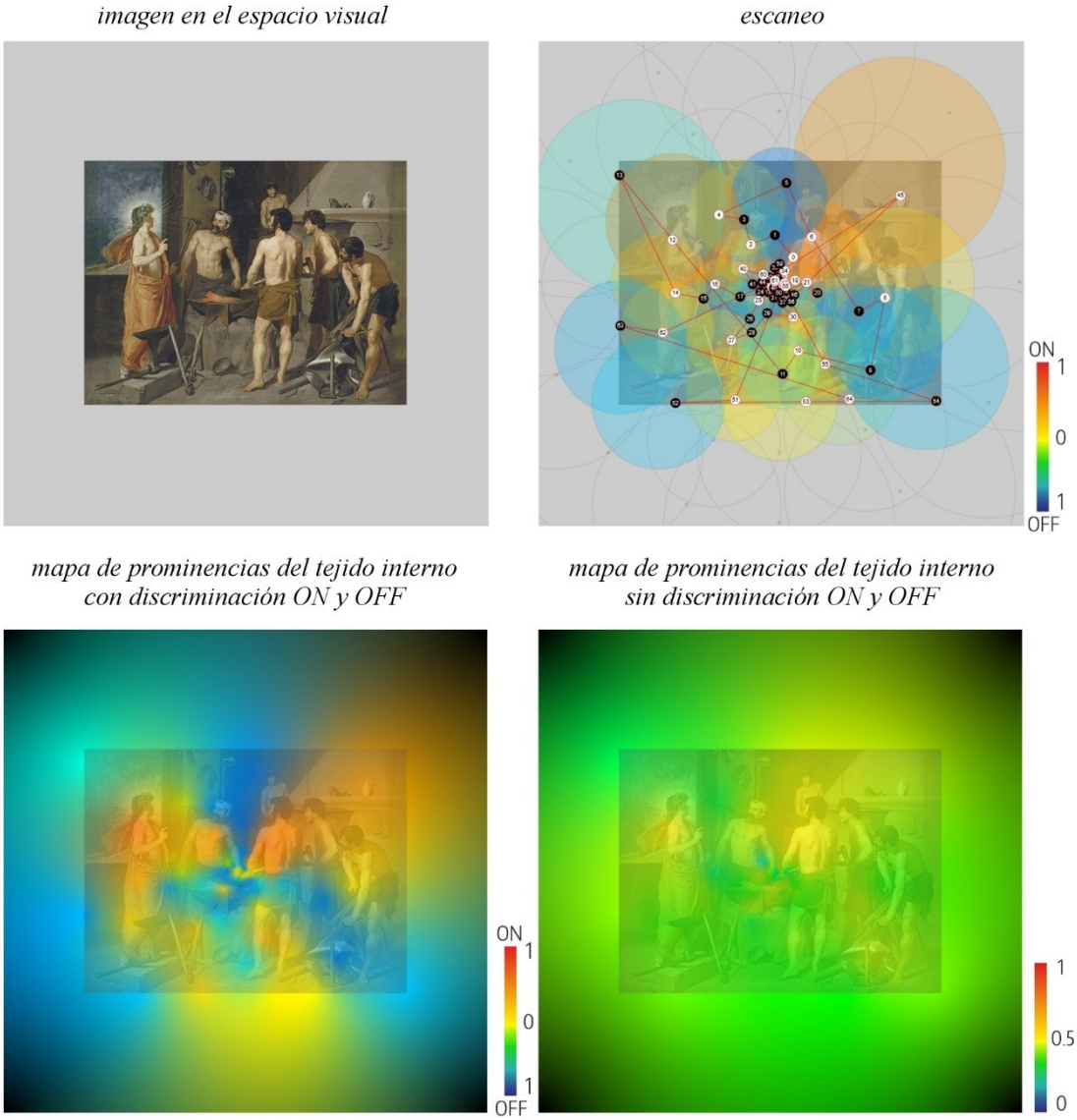
En cuanto a los mapas retinotópicos, si analizamos el trazado de las primeras veinte iteraciones de escaneado, podemos comprobar que hay una tendencia a la región central, luego a la región de la izquierda inferior cercana al centro, después la de la derecha, luego a la superior izquierda, a la superior derecha y después a las regiones externas. Para ampliar la visualización del comportamiento de Inner Fabric ajustado, la Figura 60.d muestra tres momentos del proceso de escaneo: t=4, t=6 y t=20. En t=4, Inner Fabric ha localizado dos regiones prominentes en ambos cuadros, primero en el blanco, después en el negro, y luego en la región neutra. En t=6, Inner Fabric ha localizado otros dos campos en ambos cuadros, y en t=20, todas las regiones de ambos cuadros, gran parte de la región neutra y los límites de la imagen.

4.3.3 Representación del mapa de prominencia

Las dos principales aplicaciones de los mapas de prominencia del tejido interno son para su visualización por parte de un experto y para su uso en aplicaciones de visión artificial. El esquema de referencia lo hemos representado de dos maneras: en coordenadas polares y con mapas logarítmicos. Para un experto que tenga que visualizar la imagen, la primera opción es la más adecuada, ya que es más fácil relacionar el mapa con la imagen. Pero también lo es para las aplicaciones en visión artificial, sobre todo en redes neuronales artificiales.

Para representar las prominencias, se usa un mapa kriging donde se ha utilizado la posición de cada CRA en el espacio visual. Este tipo de representación es la más adecuada para facilitar la visualización de las prominencias, ya que permite la interpolación. Otra opción era usar la posición de los campos receptivos de los esquemas de referencia, pero en los campos receptivos grandes, como sucede en la zona exterior, la posición del centro y la del CRA pueden ser muy distantes y distorsionar el efecto de la prominencia real.

La Figura 61 tiene los tres tipos de mapas usados: el del esquema de referencia con los campos receptivos con un color que representa su prominencia y el trazado de CRA, el mapa kriging donde se discrimina ON y OFF y el mapa kriging sin discriminar ON y OFF. Para los mapas que discriminan ON y OFF, la escala que se utiliza de temperatura destina los cálidos a la vía ON y los fríos a la OFF. En este tipo de mapas se visualiza



«La Fragua de Vulcano» por Diego Rodríguez de Silva y Velázquez, (1630)

Figura 61 Representaciones del mapa de prominencia y esquema de referencia.

con facilidad la relación de prominencia con el contraste entre ON y OFF, mientras que los mapas donde no hay discriminación, hay menos contraste, y la visualización es más compleja.

4.3.4 Resultados experimentales

Para comprobar el ajuste de Inner Fabric en imágenes artísticas y con criterios estéticos, se analizan cuatro pinturas diferentes, tanto por la composición como por el estilo y la temática. La primera obra es «El cardenal» de Rafael (1510-1511), una obra renacentista con una estructura basada en un triángulo, donde el fondo oscuro contrasta con la figura (ver Figura 62.a); en segundo lugar, «El triunfo de San Agustín» de Claudio Coello (1664), donde se establece un eje vertical ligeramente inclinado y un contraste alto (Figura 62..b); en tercer lugar, «La lucha de los mamelucos» de Francisco de Goya (1814), con un uso del espacio y el contraste expresionista, donde la composición se convierte en un gran mosaico de elementos prominentes entremezclados en una rejilla (ver Figura 62.c); y por último, «Chicos en la playa» de Joaquín Sorolla (1909), donde una diagonal construye la composición con elementos simples (ver Figura 62..d).

Con la finalidad de visualizar los resultados de Inner Fabric, la Figura 63 muestra el trazado del escaneo sobreimpreso en el esquema de referencia y los mapas de prominencia del tejido interno resultantes, uno con discriminación de la vía ON y OFF y otro sin ella. En el caso de la obra de Rafael, donde existe una estructura basada en un triángulo, en el mapa de prominencia con discriminación ON y OFF, la región superior domina los CRA con prominencia alta en OFF, con excepción de parte del rostro con una pequeña prominencia alta en ON. Por otro lado, en la región inferior, la prominen-

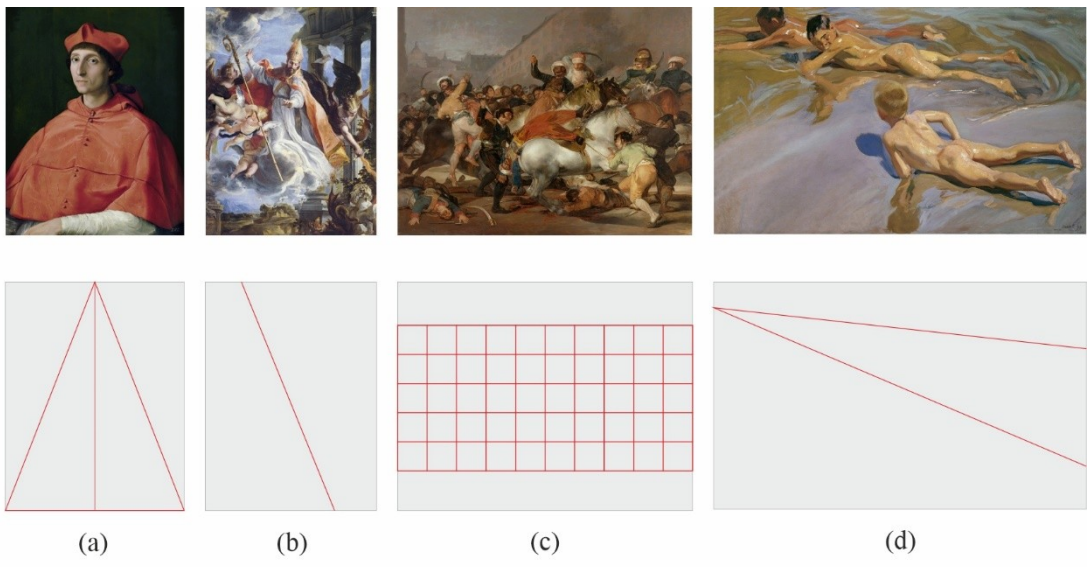


Figura 62 Ejemplos de esquemas de composición en pinturas.
Nota: (a) «El Cardenal» de Rafael (1510-1511); (b) «El triunfo de San Agustín» de Claudio Coello (1664); (c) «La lucha con los mamelucos» de Francisco de Goya y Lucientes (1814); y (d) «Chicos en la playa» de Joaquín Sorolla (1909). La parte inferior muestra la estructura de sus composiciones.

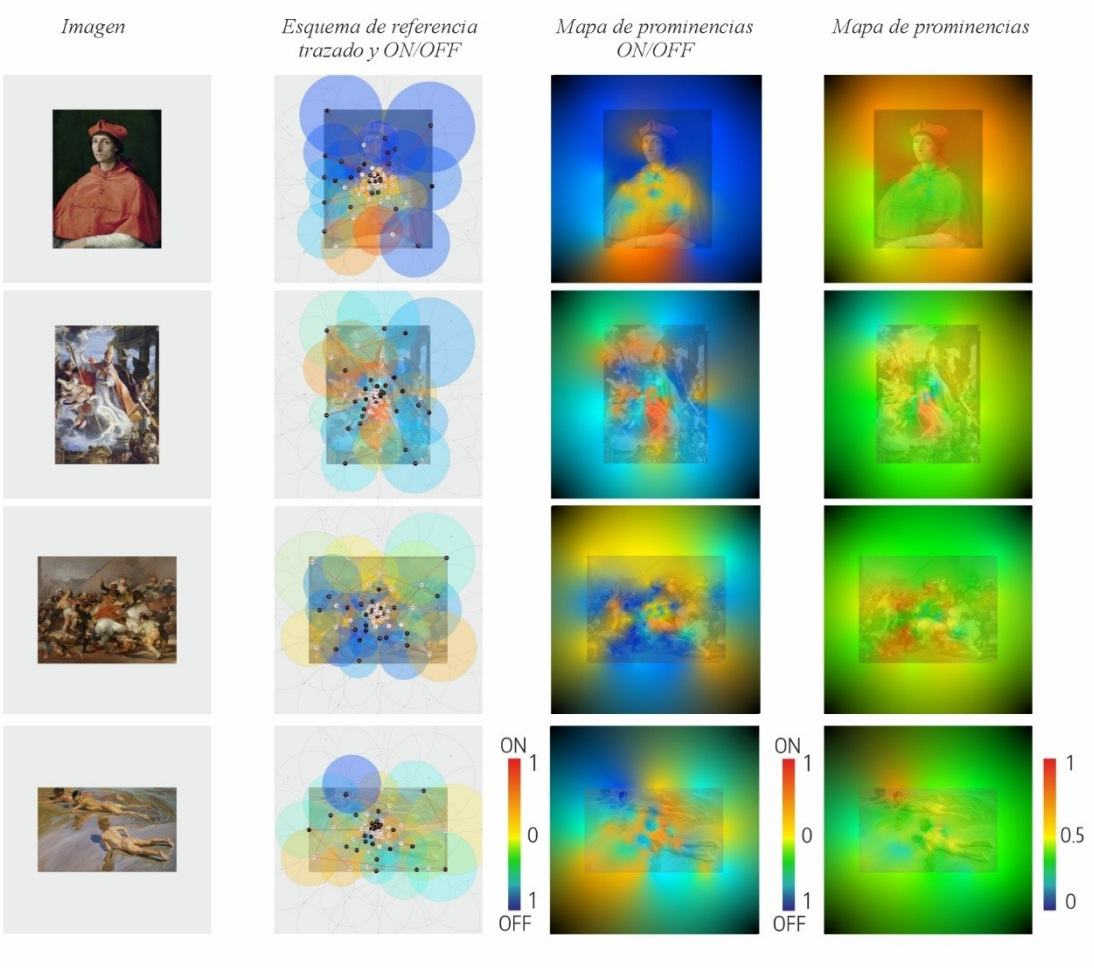


Figura 63 Ejemplos de esquemas de referencia y de los mapas del tejido interno.

cia alta está en la vía ON, aunque menor que en la superior. Toda la parte media, el cuerpo de la figura, combina prominencias medias de ambas vías. Si analizamos el mapa de prominencia que no discrimina ON de OFF, hay una prominencia alta en la parte superior, menor en la inferior, y media en toda la región central. En este caso, el equilibrio entre la parte superior e inferior se mantiene por esa diferencia de prominencia, mayor en la región más inestable (superior) y menor en la más estable (inferior) –según la preponderancia de la región inferior izquierda de la imagen.

En el caso de Coello, la estructura de la composición se organiza en torno al eje vertical con una diagonal que va desde la región inferior derecha a la región superior izquierda. La mayor prominencia se encuentra en la parte inferior media en la vía ON, mientras que, en la parte superior a ambos lados del eje vertical, las regiones son prominentes en valores medios, en ON en la parte izquierda y en OFF en la derecha. La combinación de la prominencia en OFF (región superior-derecha) y en ON (región superior-izquierda) provoca el dinamismo en torno al eje vertical.

En la obra de Goya, el tejido interno se establece con la alternancia de regiones pequeñas de prominencias en ON y OFF que representan a las figuras de la batalla. En el

mapa de prominencia sin discriminación de ON y OFF se ve con claridad cómo se genera un círculo alrededor de la región central.

En la obra de Sorolla, toda la composición es guiada por una diagonal que va de la región superior izquierda a la inferior derecha. El mapa de prominencia con discriminación ON y OFF muestra que la región superior de la diagonal es prominente en OFF y en la del medio en ON, aunque, a ambos lados de esta diagonal existen varias regiones prominentes también en ON. Sin embargo, el mapa de prominencia sin discriminación muestra con mayor claridad este eje diagonal, con la mayor prominencia en la región superior y con el contrapeso de equilibrio en la región media con una prominencia medio-alta (amarillo).

Los mapas de prominencia representan el tejido interno que, como hemos visualizado en los ejemplos, se relaciona con la composición que se analiza con los elementos visuales como: contornos, formas, color, texturas, etc. Con la configuración realizada, Inner Fabric obtiene mapas de prominencia para representar el tejido interno con regiones prominentes en la vía ON y OFF dentro de un mismo proceso. Por consiguiente, Inner Fabric escanea el espacio visual ajustándose a los criterios de una percepción horizontal, a la existencia de dos focos de atracción (central y externo), a la prominencia de la región inferior izquierda o la relación entre las regiones en un proceso de agudización y nivelación. En estos ejemplos, es posible visualizar la relación que existe entre el tejido interno y la composición de la imagen, sin que las prominencias detectadas se hayan obtenido a través de las características visuales.

Las distintas operaciones que ejecuta Inner Fabric para obtener el mapa, así como la arquitectura del esquema de referencia y los mapas retinotópicos, facilitan la implementación del marco estructural de Arnheim. Para ver con mayor claridad esta

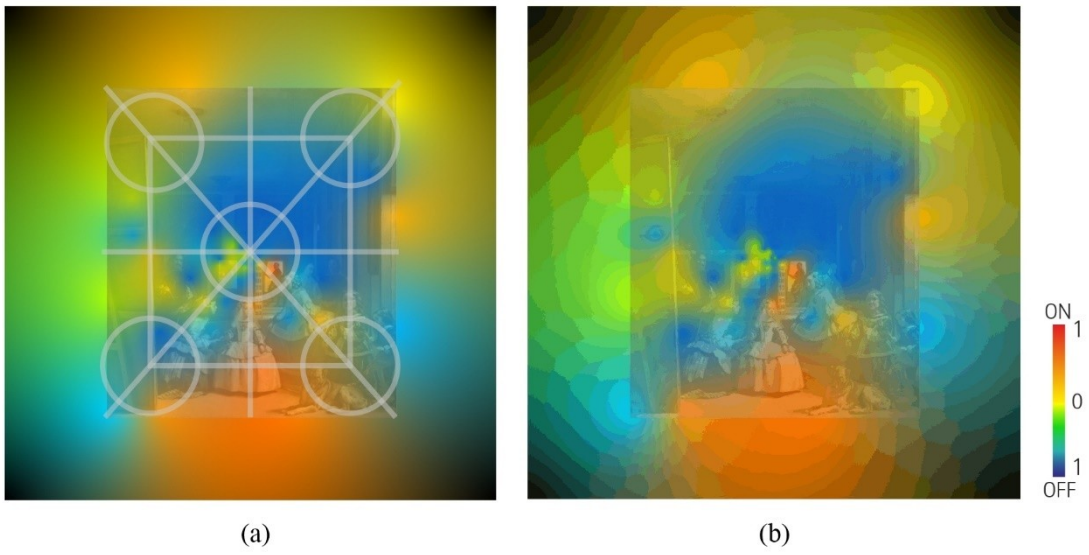


Figura 64 Marco estructural y el mapa de prominencia del tejido interno.
Nota: (a) mapa estructural sobreimpreso en el mapa kriging de prominencia (b) versión del mapa de prominencia a 8 bits. «Las Meninas» de Diego Rodríguez de Silva y Velázquez (Ver el análisis completo en el Anexo 2).

implementación, la Figura 64 presenta un ejemplo en una pintura conocida («Las Meninas» de Velázquez) donde en la Figura 64.a se ha sobreimpreso el marco estructural de Arnheim. La representación que ofrece el mapa, tamaño de las áreas de las prominencias, más pequeñas en la región central y mayores en las regiones externas, demuestran que Inner Fabric se ajusta al marco estructural. En la Figura 64.b, el mapa se ha reducido su profundidad de color a 8 bits con 255 colores con la finalidad de reducir las transiciones y ver más claro el tamaño de las prominencias. Se puede observar una diferencia tanto en la cantidad de prominencias como en el tamaño entre la esquina inferior izquierda y la superior derecha. La cantidad es menor en la parte superior derecha, mientras que el tamaño es mayor, y a la inversa en la inferior izquierda. Existe una mayor representación de la parte inferior izquierda, lo que se relaciona con la excentricidad y la anisotropía de las arquitecturas retinotópicas y del esquema de referencia, que tiene su repercusión en los mapas. En el Anexo 2, se pueden consultar más ejemplos de pinturas de la colección del Museo del Prado.

5 Casos de uso

La representación del tejido interno como mapa de prominencia permite tanto el análisis por parte de un experto como su uso en tareas de visión artificial. Es decir, el mapa en sí es una herramienta que muestra las regiones más relevantes y permite tener una visión global de la composición sin características visuales, lo cual, para un experto, supone tener la posibilidad de analizar las composiciones de las imágenes con independencia de la temática, el estilo o si es arte abstracto o figurativo. Para la visión artificial, el mapa es una herramienta para procesos de bajo nivel semántico y para tareas donde se utilice la composición, tanto en la generación de imágenes como en su procesamiento y análisis.

Para poder evaluar su capacidad y funcionamiento, se han llevado a cabo una serie de experimentos basados, por un lado, en las innovaciones que utiliza Inner Fabric y, por el otro, en la aplicación del mapa de prominencia en tareas de visión artificial. Una de las principales innovaciones es el sistema de color OCC, que permite convertir el sistema de color RGB a una escala jerárquica de una sola dimensión (que es comparable a la conversión de color a blanco y negro). Con la finalidad de controlar la conversión, se añade una función que aumenta los colores cálidos y reduce los fríos. El sistema de conversión completo se patentó con el nombre de «Método y sistema para convertir una imagen digital de color a escala de grises» (España Patente nº 201831253, 2017). Para poder evaluar esta conversión, se comprueba que se mantiene la información de color aplicando la prueba de Ishihara (Clark, 1924) –utilizada en humanos para detectar problemas con la discriminación del color– y, además, se compara los resultados de esta prueba con otras técnicas de conversión de color a blanco y negro.

Otro aspecto relevante es la relación que se establece en Inner Fabric entre el tejido interno y la estructura del NGL. La conexión con áreas de la corteza visual, como la V4, indica la posibilidad de que exista en la extracción de características visuales, como la forma y el color, una aportación de la información precortical del NGL. Para analizar esta relación, un segundo experimento implementa una conexión de las capas iniciales (equiparable al NGL) a las profundas (equiparable al V4) en redes neuronales convolucionales. Después, evalúa si las capas profundas mejoran su selectividad al color y, por consiguiente, la señal poco procesada, comparable al tejido interno según Inner Fabric, juega un papel relevante en facilitar la extracción de características visuales.

Por último, para evaluar el uso del mapa de prominencia del tejido interno en tareas de visión artificial, se han creado dos aplicaciones con imágenes artísticas. La primera es un clasificador de tipos de composición basado en las categorías de Arnheim de central, binaria, jerárquica y atonal (Arnheim, 1956). Este clasificador se implementa en una red neuronal artificial mediante aprendizaje supervisado gracias a un *dataset* de casos sintéticos, creado con reglas simples sin necesidad de etiquetado humano, utilizando a los mapas de prominencia del tejido interno. El segundo es un buscador CBIR que localiza imágenes de artistas de distintos estilos con tejidos internos similares. Esto posibilita

localizar imágenes con composiciones parecidas sin la necesidad de extraer las características visuales y realizar tareas complejas de análisis.

5.1 Conversión de color a escala de grises OCC++

Con el surgimiento de las cámaras y los televisores digitales, y después de los monitores para equipos informáticos durante el siglo XX, el procesamiento de las imágenes se estableció como un campo multidisciplinar en donde encontraba cobijo un amplio abanico de especialistas, científicos y técnicos, pero también de creativos, diseñadores y artistas. Uno de los problemas con el que se encontraron fue el de la conversión a escala de grises (blanco y negro para el gran público) de imágenes en color, ya que, en la conversión, se pierde parte de la información que esta característica visual ofrece en la percepción (Bala & Eschbach, 2004).

Una de las principales herramientas utilizadas al componer imágenes es el contraste de color, que establece relaciones jerárquicas y de oposición entre las regiones de la imagen (Pawlik, 1976). El contraste significa diferencia, lo que implica orden y distancia. Se ha demostrado empíricamente que existe una relación natural entre los principios objetivos de ordenación del color y la percepción humana (Puhalla, 2008). Sin embargo, la forma en que los seres humanos categorizan el espectro de colores visibles es uno de los problemas fundamentales de la ciencia cognitiva (Loreto y otros, 2012) e, independientemente del procesamiento interno que se haga en el espacio de color, la distancia euclidiana resultante no se corresponde con la percepción humana de las diferencias entre colores (Gravesen, 2015).

Al mismo tiempo, los colores representan categorías cognitivas, estableciendo relaciones semánticas a distintos niveles de abstracción. Las categorías de bajo nivel incluyen «cálido», «frío», «claro», «oscuro» (Arnheim, 1956); (Berlin & Kay, 1991), «claro-cálido» y "oscuro-frio" (Rosch, 1975) mientras que en los niveles superiores cabe esperar conceptos complejos como que el verde signifique «esperanza» o el rojo «pasión» (Pawlik, 1976). Por lo tanto, el análisis del contraste de las abstracciones cromáticas de bajo nivel es independiente del contenido (frío frente a cálido, claro frente a oscuro, etc.).

El sistema de color creado en esta tesis, OCC, y que se ha utilizado para relacionar el color con la actividad neuronal con el fin de obtener el peso visual, establece una jerarquía que posibilita la conversión de una imagen de color a escala a grises. Para poder controlar la conversión, se incluye una variación de OCC que permite regular la calidez y frialdad (denominado OCC++). El método y sistema está patentado como «Método y sistema para convertir una imagen digital de color a escala de grises» (España Patente nº 201831253, 2017).

5.1.1 Conversores de color a escala de grises

La conversión de color a escala de grises es una tarea habitual en las aplicaciones de edición de imágenes y visión por ordenador. Cuando convertimos una imagen en color

a escala de grises, las propiedades de matiz, luminancia y saturación deben mantenerse. Algunos conversores aplican pesos a cada canal RGB (Bala & Eschbach, 2008), (Majewicz & Smith, 2013), (Ng, 2013) para obtener las diferencias entre los rangos de los canales. Este tipo de conversores asumen cuestiones de la percepción humana para su desarrollo y aplicar cálculos científicos para estimar los pesos. Los conversores que utilizan pesos más destacados son CIECAM97, L*a*b* lum, XYZ lum o YCrCb Lum, y algunas de las patentes: «Method and device for use in converting a colour image into a grayscale image». (US8355566B2), «Method and apparatus for converting a color image to grayscale» (US8594419B2), «Method of converting color image into grayscale image and recording medium storing program for performing the same» (US8526719B2) o «Method and device for use in converting a colour image into a grayscale image» (US7382915B2). El sistema más implementado en aplicaciones informáticas y usado en librerías de visión artificial, Matlab o Photoshop como ejemplo, es la Recomendación BT.601 de la Unión Internacional de Telecomunicaciones (UIT).

Existe una segunda línea de conversores que ponderan el valor del color RGB original por el contraste en cada región de la imagen para establecer una diferenciación según la relación de cada píxel con su entorno (Kuk y otros, 2011). El contraste local de cada píxel estima su relevancia dentro de la imagen con el fin de determinar su posición en la escala de grises final. Para llevar a cabo esta tarea, es necesario aplicar algoritmos y procesamientos concretos en cada píxel (o en cada grupo de píxeles) que dependen de circunstancias locales, como en Color2gray (Gooch y otros, 2005), globales con el resto de la imagen (Ma y otros, 2015), o, incluso, con ambas (Du y otros, 2014); (Kuk y otros, 2010). Algunas de las patentes más destacadas de estos métodos son: «Color to monochrome conversion» (US4977398A), «Mapping of color images to black-and-white textured images» (US5153576A), «Image processing for converting color images into monochrome pattern images» (US5726781A), «System for black and white printing of colored pages» (US5898819A), «Printing black and white reproducible color documents» (US5701401A) o «Color transforming method» (US6101272A).

Otra opción destacable es «Method for converting a video signal into a black/white signal» (US4257070A), que utiliza canales con propiedades visuales: intensidad, matiz, textura y efecto pictórico, y aplica umbrales, o (Seo & Kim, 2013) que utiliza el análisis de componentes principales para reducir la dimensionalidad.

5.1.2 Descripción del conversor de color a escala de grises OCC++

La organización de los colores en una jerarquía es un tema de estudio amplio y ha sido objeto de análisis inicialmente por artistas y, sobre todo a partir de mediados del siglo XX, por la psicología del arte (Pawlik, 1976). Existen diferentes modelos para representar el color, y el consenso es difícil en el campo del arte, aunque se puede establecer que todos coinciden en que existen unos colores primarios. A partir de estos colores primarios, a través de la mezcla entre sí, se dan los colores secundarios, y estos, a su vez, a los terciarios, y así hasta llegar a distintas combinaciones. Desde la esfera de color de Runge

(Runge, 2010), a la de Munsell (Munsell, 1915), pasando por sistemas de ruedas o pirámides, pero en ningún caso se establece una jerarquía hasta llegar al modelo de hexágono de Kueppers (Kueppers, 1982). El modelo de Kueppers sí establece una jerarquía con su «romboedro» que va del blanco, pasando por el amarillo, rojo, verde y azul, hasta el negro. El determinar esta jerarquía es un paso necesario para obtener una escala donde las propiedades de cada color queden reflejadas en la transformación a niveles de gris, pero resulta difícil determinar una conversión efectiva desde el sistema RGB.

Además, en las artes visuales, la distinción entre colores cálidos y fríos es muy común (Arnheim, 1956), lo cual permite establecer el contraste de color también por categorías. Berlin and Kay (Berlin & Kay, 1991), después de estudiar distintos lenguajes alrededor del mundo, demostraron que los colores son universales e independientes a sensibilidades culturales. E. Rosch (Rosch, 1975) descubrió una tribu en Nueva Guinea (Los Danis) con solamente dos categorías: claro-cálido y frío-oscuro. Ellos aprendieron los colores rojo, verde, azul y amarillo (los colores primarios para los procesos de colores opuestos) mucho más rápido que los otros. Una justificación de este hecho es que existe una estructura neuronal que analiza las relaciones entre colores opuestos y, por tanto, la percepción visual del contraste entre las categorías cálido y claro opuestas a frío y oscuro.

Existen muchos estudios sobre el contraste de colores que analizan la relación con colores vecinos y opuestos. Como resultado, se han propuesto diferentes sistemas de color, basados en estructuras como ruedas, esferas o pirámides (Runge, 2010); (Itten, 1992); (Klee, 1961); (Munsell, 1915) Entre ellos, sobresale el espacio de color propuesto por Kueppers, donde un entramado romboédrico establece una relación jerárquica entre los colores (Kueppers, 1982). En los últimos 150 años, hemos tenido dos teorías principales sobre el procesamiento del color: la teoría tricromática (Young, 1801), (Helmholtz, 1852) y la teoría de colores opuestos (Goethe, 1840); (Schopenhauer, 1816); (Hering, 1885). Sin embargo, a finales del siglo pasado, la neurociencia demostró que ambas teorías coexisten en la corteza visual; el ojo capta la luz mediante un proceso tricromático, pero después convierte los datos adquiridos en una estructura de procesos oponentes mediante operaciones neuronales (Hubel, 1988). La sensación de un color es el resultado de procesos neuronales, como propuso Land en su teoría Retinex (Land, 1977).

Schopenhauer describe en su libro sobe la visión y el color (Schopenhauer, 1816) la relación entre el color y la división de la actividad de la retina que permite establecer una jerarquía, de tal manera que el blanco corresponde a la actividad plena y el negro a la ausencia total de actividad, mientras que el amarillo es ¾ de actividad, el rojo y el verde ½, y el azul ¼. Esta descripción completa está en el apartado «4.2.3. Calcular los pesos visuales».

El principal problema en la conversión del color a partir de sus tres dimensiones: matiz, luminosidad y saturación es que la combinación de los tres canales RGB obtiene una escala de grises que mide la luminosidad. Con el uso de pesos en cada canal, se introdu-

ce además el matiz, pero la combinación final no es fácil partiendo del sistema RGB y sin incluir la saturación. En los sistemas que analizan el contraste local, funciona con eficacia cuando éste es alto, pero no cuando es medio o bajo. Esto provoca que los colores, en regiones donde el contraste no es muy alto, se encuentren en valores bajos en la escala de grises final, aunque sus colores originales tengan un valor mayor en sus propiedades de luminosidad, matiz o saturación. Además, la valoración local o global, donde se compara cada píxel con su entorno, establece una escala final ligada a la propia estructura de la composición de colores de la imagen, pero es complejo comparar los resultados entre varias imágenes, al depender la escala final del contraste interno de cada imagen.

A partir de la teoría de Schopenhauer, que relaciona el color con la actividad de la retina, se convierten los tres canales RGB en cuatro canales (L, M, S, y LM) en la vía ON, y (-L, -M, -S y -LM) en la vía OFF. Partiendo de los conceptos de actividad dividida de la retina de Schopenhauer, las vías opuestas ON y OFF transportan los valores de actividad e inactividad, siendo su suma, «la actividad completa» (AC), constante. La media de los cuatro canales es el valor de actividad que determina una posición jerárquica, que describimos en las ecuaciones (26) en la vía ON y (27) en la vía OFF, donde se mantienen las relaciones de luminosidad, matiz y saturación. Para el conversor de color a escala de grises, sólo usaremos la vía ON con la ecuación (26) que calcula el valor de actividad:

$$A = (LM + L + M + S)/4 \tag{44}$$

donde $L = R, M = G, S = B$ y $LM = min(R + G, 1)$

El valor A sería, por lo tanto, el valor en la escala de grises en la conversión. Para poder controlar la calidez y frialdad, se aplica la siguiente ecuación que permite ponderar la calidez y frialdad:

$$A' = A + \left(a * \left(\frac{1}{1 + e^{-3(L-M)}} - 0.5\right)\right) \tag{45}$$

donde A' es el nuevo valor de escala de grises y a es un coeficiente que pondera la calidez o frialdad de un color con dos valores constantes configurables (W para colores cálidos y C para colores fríos):

$$si\ L > M\ \ a = W\ sino\ a = C;\quad 0 \leq W, C \leq 1 \tag{46}$$

En la patente, el control de la calidez y frialdad se consigue a través del canal L. En estudios posteriores, se ha comprobado que la conversión tenía problemas con los colores fríos, con lo que, tras un análisis y reconfiguración del sistema, se cambió el control desde el canal L a L y M.

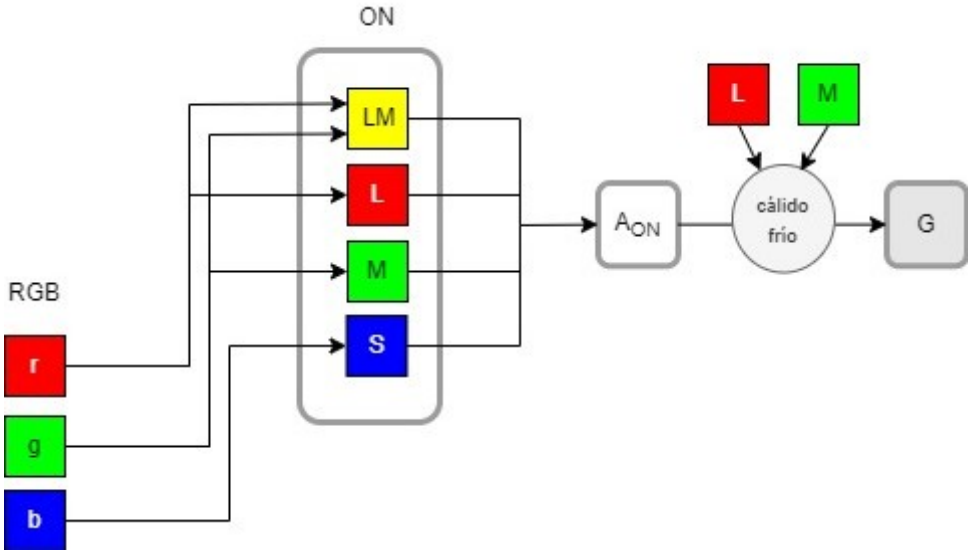


Figura 65 Esquema del funcionamiento OCC++.

La Figura 65 muestra el esquema completo del sistema donde se ve la relación entre los canales RGB y el sistemaOCC, y el control de calidez y frialdad. El valor G sería el resultado final en la escala de grises.

5.1.3 Análisis objetivo de la discriminación del color

En la conversión de color a escala de grises, el principal objetivo es mantener la máxima expresividad de la combinación de los tres canales RGB en uno sólo. Para poder evaluar si esta expresividad en la conversión se mantiene, hay que analizar la discriminación del color de la conversión. Para este fin, existe la prueba de Ishihara (Clark, 1924), que es utilizada para detectar problemas con la discriminación al color en humanos. Una de las principales aplicaciones de la prueba es la detección del Daltonismo, que es una discapacidad de origen genético que provoca que los receptores de la retina no sean sensibles a un rango de longitud de onda concreto o a varios. Es decir, que la distinción entre las gamas rojas y verdes no es posible o es mínima, e incluso que, si afecta a los tres tipos de conos, tampoco se distinga el azul.

La prueba original está compuesta de 38 cartas que representan en un círculo una textura en varios colores. En cada textura se esconde un número o unas líneas que no son visibles si se tiene afectación de un tipo determinado de receptor. Cada bloque de cartas tiene como finalidad detectar uno o varios problemas concretos:

- En la carta 1 y 38, los números deben ser visibles para todos los humanos.
- En las cartas de la 2 al 9, debe ser visible con claridad un número para la visión normal y otro diferente si el problema es en la detección del rojo o del verde.
- En las cartas de la 10 al 17, debe se visible con claridad un número para la visión normal y ninguno cuando hay problemas en la detección del rojo o del verde.

- En las cartas de la 18 al 21, no debe ser visible ningún número para la visión normal y visible un número para la visión con problemas de detección del rojo o del verde.
- En las cartas de la 22 a 25, debe ser visible un número de dos cifras para la visión normal, y el primero cuando hay problemas en la detección del rojo y el segundo con la del verde.
- En las cartas 26 y 27, las líneas deben ser visibles para la visión normal, y sólo visible la parte superior cuando hay problemas de detección del rojo y la inferior con la del verde.
- En las cartas 28 y 29, las líneas son sólo visibles para la visión con problemas en la detección del rojo o del verde.
- En las cartas del 31 al 33, los números son sólo visibles para la visión normal.
- En las cartas 34 y 35, la visión normal detecta la línea verde y cuando hay problemas en la detección del rojo y del verde sólo detecta la línea violeta.
- En las cartas 36 y 37, la visión normal detecta la línea naranja y cuando hay problemas en la detección del rojo y del verde sólo detecta la línea violeta.

La prueba determina si existe protanomalía, cuando hay una detección de la gama del rojo débil, o protanopía, cuando no la detecta. También, deuteranomalía, cuando hay una detección de la gama del verde débil, o deuteranopía, cuando no la detecta.

Para poder comparar entre los tipos de conversores con OCC++, se convierten las cartas de color a escala de grises usando tres tipos de conversores: BT.601 (con pesos), usando el análisis de componentes principales (PCA), color2gray (usando contraste de vecindario) con un contraste bajo y alto. Se aplica 0.4 para los coeficientes W y C de OCC++. Los resultados se comprueban por inspección visual a partir de la descripción de cada carta.

Resultados experimentales

La Figura 66, Figura 67 y Figura 68 muestran los resultados para las 38 cartas y los tres conversores. Para facilitar la comparativa, la Tabla 6 recopila las cartas que sí discriminan el color, si es con contraste bajo o alto, y si hay alguna diferencia entre la gama del rojo o del verde. Se concluye que:

- El conversor por pesos (BT.601) no es capaz de discriminar los números de las cartas 1 y 38, que cualquier humano (con o sin daltonismo) es capaz de hacer. Las cartas del 2 al 5, sí discrimina los números sin problemas, pero en la 8 y 9, de la 14 a la 17 y de la 20 y a la 21, lo consigue, pero con un contraste muy bajo. Las cartas entre la 23 y 27, que permite diferenciar si existe problemas con el rojo y el verde, discrimina sólo los números que detectaría un humano con deuteranomalía, o sea con problemas con la gama del verde.

- El PCA, tampoco discrimina las cartas 1 y 38. Al igual que el conversor de pesos, también consigue discriminar las cartas del 2 al 5. Con contraste bajo la 12. De la 23 a la 27, donde sólo discrimina el número o línea que detectaría un humano con deuteranomalía o sea con problemas con la gama del verde. Finalmente, discrimina la 31, la 36 y 37.

- Color2gray, con la conversión de contraste bajo, sí discrimina débilmente la 1, de la 2 a la 5; la 15 y la 17. De la 22 a la 27, sólo los números relacionados con la con deuteranomalía o sea con problemas con la gama del verde. Color2gray, con la versión de contraste alto, discrimina perfectamente la carta 1, de la 2 a la 9 y de la 15 a la 26. De la 27 a la 31 con mayor contraste los números o líneas que detectaría un humano con deuteranomalía o sea con problemas con la gama del verde, y con menor contraste los números y líneas que detectaría un humano con protanomalía, o sea con problemas con la gama del rojo. Por último, discrimina sin problemas de la 32 a la 38.

- OCC++ discrimina todas las cartas, además no tiene problemas con la gama del verde y tampoco con la gama del rojo.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
BT.601		C	C	C	C			c	c					c	c	c	c		
PCM		C	C	C	C							c							
color2gray B	C	C	C	C	C										C	C	C		
color2gray A	C	C	C	C	C	C	C	C	C						C	C	C	C	C
OCC++	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C

	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
BT.601				V	V	V	V	V											C
PCM				V	V	V	V	V				C					C	C	
color2gray B			V	V	V	V	V	V											
color2gray A	C	C	C	C	C	C	C	Vr	Vr	Vr	Vr	Vr	C	C	C	C	C	C	C
OCC++	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C

Tabla 6 Resultados de la prueba de discriminación del color para los conversores.
Nota: B, contraste bajo; A, contraste alto; C, discriminación completa; R, discriminación con problemas en la gama del rojo; V, discriminación con problemas en la gama del verde; mayúscula: discriminación con contraste alto; y minúscula: discriminación con contraste bajo.

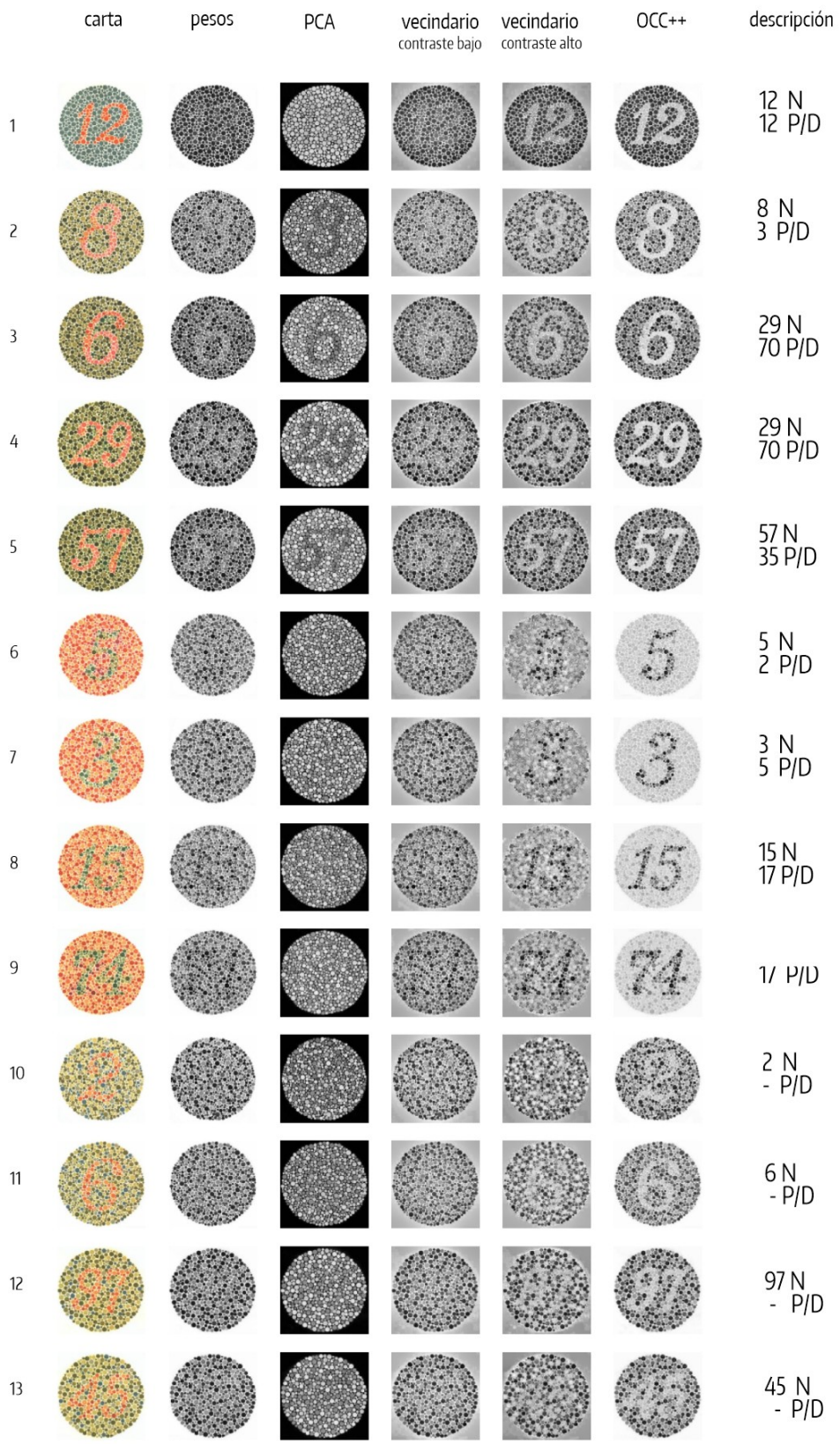
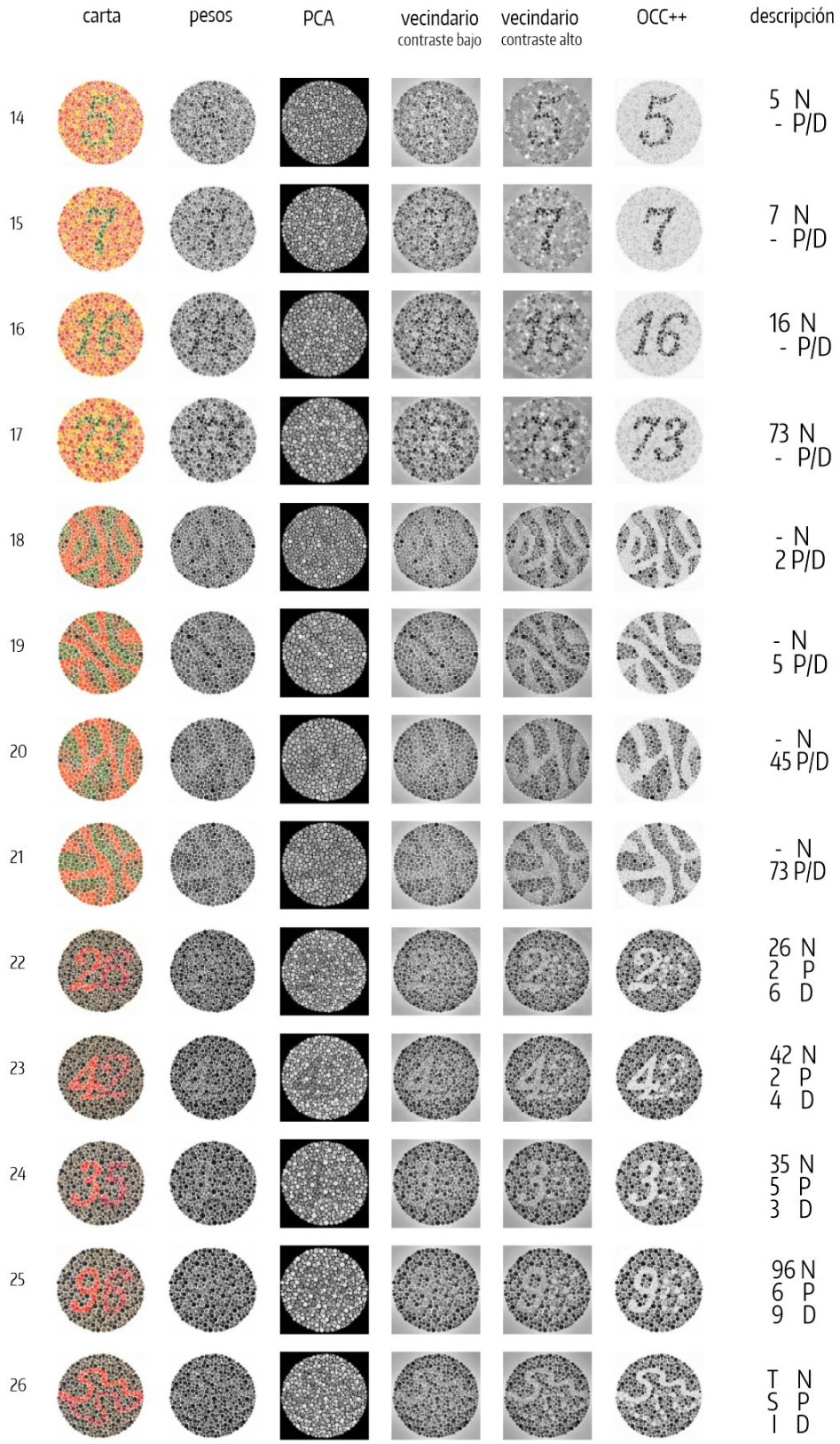
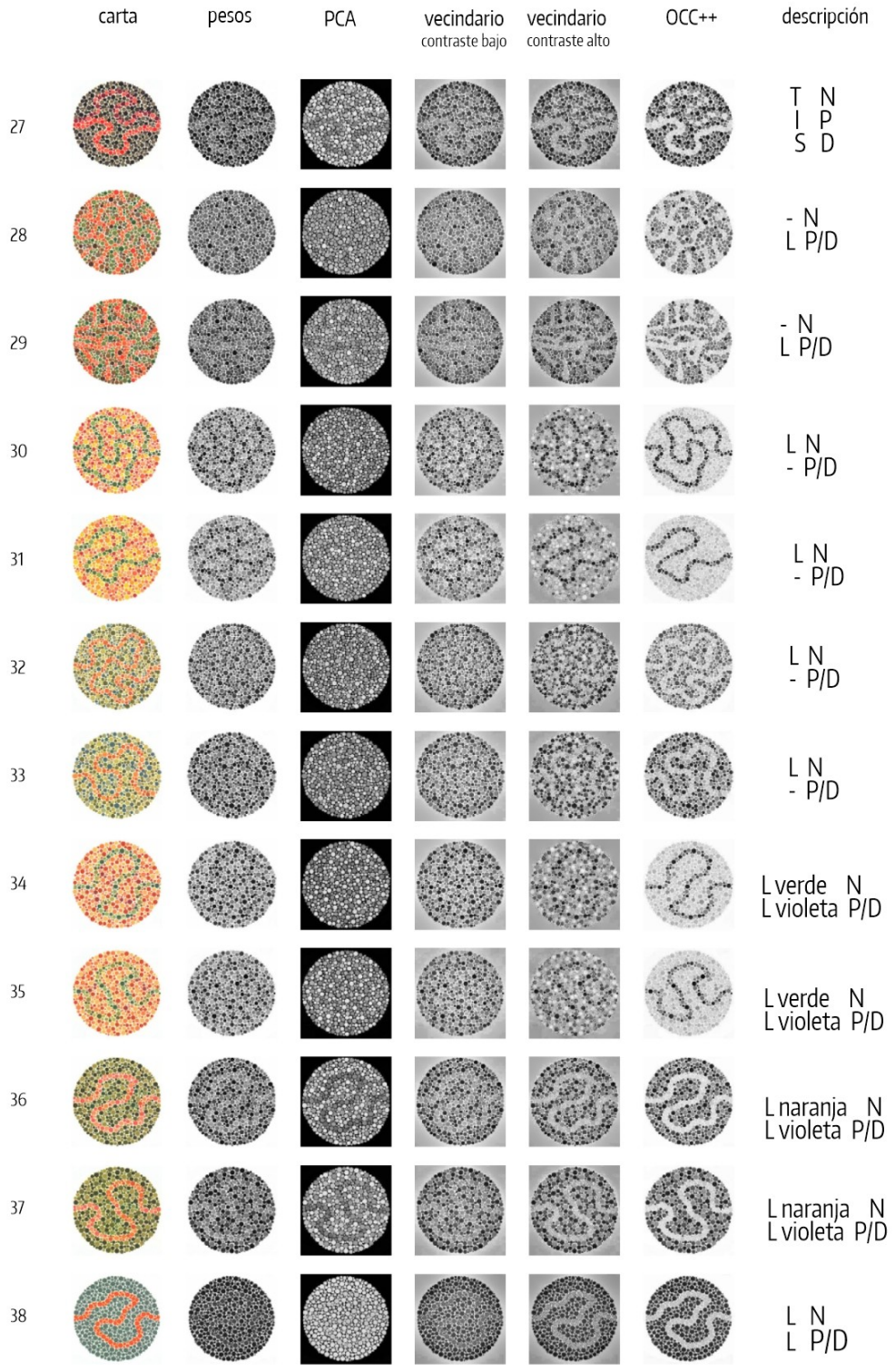


Figura 66 Resultados de la prueba de Ishihara, cartas de la 1 a la 13.



N: Normal P: Protanomalía (débil) o Protanopía (nulo) al rojo D: Deuteranomalía (débil) o Deuteranopía (nulo) al verde
L: Lineas I: Linea inferior S: Linea superior

Figura 67 Resultados de la prueba de Ishihara, cartas de la 14 a la 26.



N: Normal P: Protanomalía (débil) o Protanopía (nulo) al rojo D: Deuteranomalía (débil) o Deuteranopía (nulo) al verde
L: Lineas I: Linea inferior S: Linea superior

Figura 68 Resultados de la prueba de Ishihara, cartas de la 27 a la 38.

5.1.4 Análisis cualitativo de la conversión a escala de grises

Para completar la evaluación, se compara los conversores en una imagen. La Figura 69 muestra una conversión de color a escala de grises de la pintura «Impresión, amanecer» (1872) de Claude Monet usando los cuatro tipos: con pesos (BT.601), PCA, con contraste de vecindario (color2gray) y OCC++. Esta pintura tiene un contraste bajo en la gamas frías y alto entre la gama cálida y fría, con el azul como gama predominante, lo cual ayuda a evaluar la calidad de cada conversión. El principal contraste de color está en el sol y su reflejo (categoría de cálido) en las aguas (categoría de frio), donde se visualiza con claridad el problema detectado en la prueba (ver mapa cálido/frio de la Figura 69.b): el conversor que usa los pesos (BT.601), muestra poco contraste entre las regiones cálidas y las frías y el PCA, al contrario, muestra un contraste muy alto. Color2Gray mantiene este contraste, pero algo bajo, y OCC++ muestra un contraste fuerte, tanto en el sol como en los reflejos, con la misma intensidad que se percibe en la versión de color. Por otro lado, las regiones de azul donde no hay un contraste muy alto (por ejemplo, la región central de la derecha del puerto con la grúa y el agua), Color2Gray consigue un contraste bajo en las aguas, mientras que OCC++ establece un contraste algo más alto, muy cercano a la versión en color. Sólo OCC++ muestra un contraste de las regiones cálidas de la parte superior con la intensidad que la versión en color consigue mezclando rojos con azules. Los conversores BT.601 y color2Gray reducen este contraste y PCA lo aumenta en exceso.

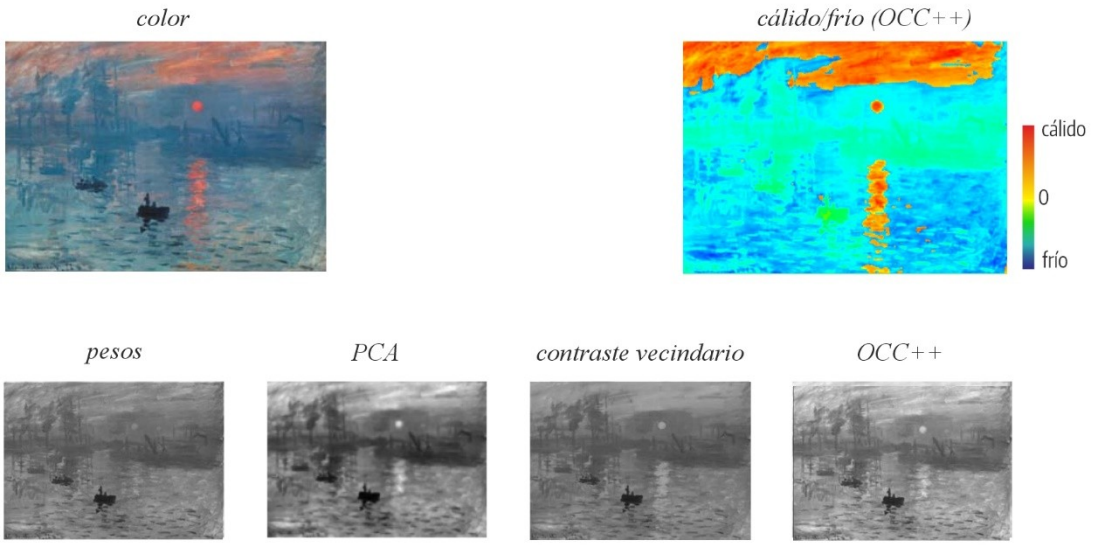


Figura 69 Ejemplo de conversión de color a escala de grises en una pintura
Nota: «Amanecer» de Claude Monet (1872); imagen en color; mapa de temperatura de las categorías cálido/frio y conversión a escala de grises con cuatro tipos de conversores.

5.1.5 Discusión

Uno de los principales problemas de los conversores de color a escala de grises convencionales es la baja discriminación del contaste de color, que se demuestra con facilidad usando la prueba de Ishihara, y que ninguno es capaz de pasar al completo. BT.601 que utiliza pesos, PCA a través del contraste entre los canales y color2Gray con contraste del vecindario presentan deuteranomalía, es decir, baja discriminación de la gama del color verde. En el conversor BT.601, el canal con un mayor peso es el rojo, y esto puede ser la causa, aunque probablemente no sea la única. En el caso del PCA, es diferente, ya que a priori su conversión se basa en una comparativa de los tres canales en cada píxel y no hay diferencia de peso entre ellos. Pero es posible que la falta de contraste entre los canales —la prueba dificulta hasta a un humano con visión normal el reconocimiento de los números y líneas—, sea la causa más lógica de los problemas de este conversor en la prueba.

En el caso de color2gray, sucede algo parecido en relación con el contraste: con una parametrización con contraste bajo presenta problemas parecidos a los de los otros dos conversores, mientras que, con otra parametrización con contraste alto, pasa la prueba casi al completo. En OCC++, con una configuración medio-baja de la diferenciación de los colores cálidos y fríos (se ha usado 0.4 para ambos) es capaz de discriminar todas las cartas de la prueba. Esto indica que la diferenciación entre cálido y frío introducida en OCC++, gracias a la estructura de cuatro canales en un sistema de procesos de colores opuestos, es determinante para mantener el contraste del color, incluso si es bajo, como sucede en algunas de las cartas donde los números tienen poco contraste con el fondo.

5.1.6 Conclusiones

El sistema de color OCC transforma el espacio de color RGB en una estructura de procesos de colores opuestos. Esto permite que se establezcan relaciones opuestas y complementarias, facilitando el análisis del contraste de color. La diferenciación entre la categoría cálida y fría en el sistema OCC++ permite una mejora considerable gracias a un mayor control del contraste del color. Los resultados de la prueba de discriminación del color demuestran que un conversor por pesos tiene problemas con el rojo y verde en muchas de las cartas, y especialmente con la gama del verde; que el PCA tienen problemas cuando el contraste entre los canales es bajo; y que el conversor por contraste de vecindario sólo mejora si el contraste de su conversión es aumentado.

De una manera global y en comparación con las otras soluciones, OCC++ tiene, además, las siguientes mejoras técnicas:

- En relación con los conversores que evalúan de manera independiente cada píxel (BT.601 y PCA), mantiene mejor el contraste del color en la conversión.
- En relación con los conversores que evalúan el contraste por vecindario, OCC++ obtiene resultados similares o incluso mejores en la conversión, pero

mejora el rendimiento computacional al realizar la evaluación independiente de cada píxel.

5.2 Mejora de la selectividad al color en arquitecturas CNN con LSC

En la solución planteada para Inner Fabric, se ha relacionado al NGL y su estructura, en dos vías paralelas y con canales separados de la información visual, con el tejido interno de la composición. Esta relación se ha basado en el concepto de la percepción horizontal de Ehrenzweig (Ehrenzweig, 1967), por la cual, la información se procesa sin las características visuales. La justificación es que la corteza visual extrae las características en áreas muy especializadas en una percepción vertical, y el tejido interno podría estar distribuido por todas ellas, mientras que en el LGN no lo está.

En el apartado «2.2.7. Procesamiento de la percepción visual», se introduce una idea sobre la conexión directa del NGL con áreas de la corteza visual como la V4, donde se procesa la forma y el color. La convención general es que, en el procesamiento, el flujo de información pasa del NGL únicamente al área V1 de la corteza visual. Inicialmente no se había planteado si podía existir otras conexiones que no siguieran un esquema secuencial. Ferrera y colaboradores (Ferrera y otros, 1994) comprobaron que en el área V4 se mantenía información de la estructura del NGL. Siguiendo el esquema secuencial, parecía que esa estructura se mantenía a pesar de las distintas operaciones que realizaban las áreas visuales. Sin embargo, esto no era muy viable, aunque revelador, ya que indicaba que la estructura del NGL tenía algún efecto en la formación de la sensación de color en el área V4. Posteriormente, se comprobó que existía una conexión directa del NGL y el área V4 en un estudio realizado con humanos (Arrigo y otros, 2016).

La conexión directa del NGL con el área V4 sugiere que podría existir una influencia de la información sin procesar en un nivel semántico bajo, como la del tejido interno, con otra procesada en un nivel semántico alto, como la de la composición con elementos visuales, sin que esta conexión sea a través sólo de pasos intermedios. Con esta idea, se realizó una investigación usando el modelo CNN, el cual está bioinspirado en la corteza visual y en las operaciones de extracción de características y asociación de patrones. En este tipo de red neuronal artificial, el flujo de información es secuencial, desde la entrada hasta la salida, pasando de la detección de bordes hasta la extracción de características cada vez más complejas en las últimas capas. En este escenario, la conexión del LGN con el área V4 se podría simular como una conexión de las primeras capas de la red con las últimas. A este tipo de conexión la denominamos LSC (*long skip connection*) y el objetivo de la investigación fue analizar y evaluar varios modelos CNN con *datasets* donde el color era una característica relevante. La investigación está publicada en (Sanchez-Cesteros y otros, 2023), y en esta sección se hace un resumen de los resultados y cuestiones relevantes.

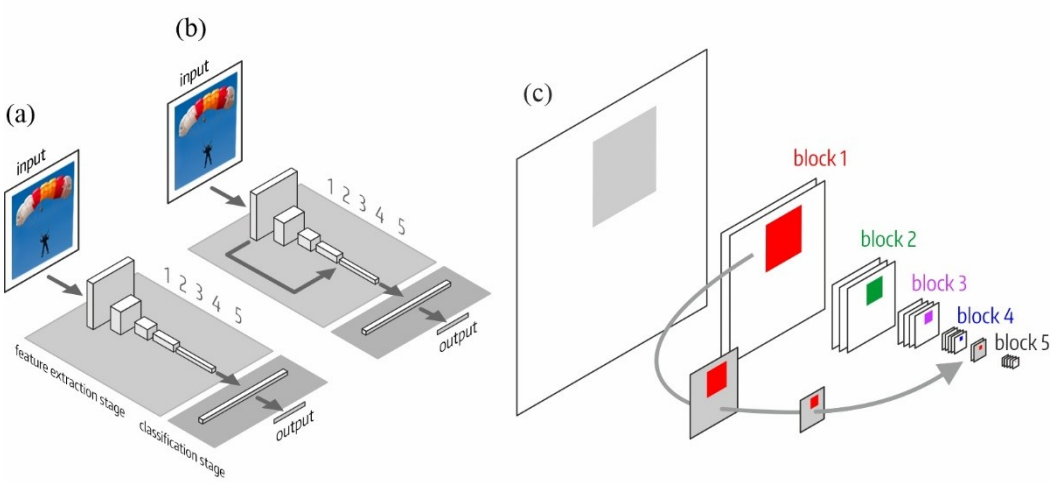


Figura 70 *Long skip connection* en la arquitectura CNN.
Nota: (a) arquitectura CNN convencional; (b) LSC en la arquitectura CNN; y (c) detalle del tamaño de un campo receptivo correspondiente a una neurona del último bloque en el resto de los bloques y en la imagen.

5.2.1 Descripción de la investigación y hallazgos relevantes

El objetivo es modificar tres tipos de arquitecturas CNN en la tarea de clasificación y comparar el rendimiento entre los modelos originales y las variantes, y entre todos ellos en la precisión de la tarea. La Figura 70.a muestra la arquitectura convencional de una CNN con el área de extracción de características y la de asociación. Las capas se agrupan por bloques, que se delimitan por la reducción del tamaño de los filtros que aplican en la salida (*subsampling*). La Figura 70.b muestra la arquitectura anterior con la LSC propuesta, conectando la salida del bloque primero con la salida del bloque cuarto, con la finalidad de que ambas sean la entrada del último bloque, el quinto.

Para obtener una visión completa de la aplicación de LSC, en la investigación se implementó en tres tipos de arquitecturas: VGG16, como CNN convencional; Densenet121, como arquitectura que aplica *skip connection* en las capas dentro de cada bloque para recuperar características; y Resnet50, con *skip connection* residuales entre las capas de cada bloque. Los seis modelos fueron entrenados en cuatro *datasets*: dos pertenecientes Imagenet con clases de temática variada (Imagenette con 10 clases y Tiny Imagente con 200 clases) y dos específicas, una de pájaros (de 315 clases) y otra de flores (de 102 clases).

Una vez entrenados los modelos en cada uno de los *dataset*, se analizaron la precisión obtenida en los *dataset* de validación, comprobando que las variantes con LCS mejoraban a sus correspondientes modelos hasta entre un 2% y un 10%. Además de este dato, el objeto del análisis era la selectividad al color, que se relaciona con la influencia de la información con poco procesamiento del primer bloque en el bloque último. Es decir, del NGL en el área V4. Para poder evaluar la selectividad del color en las capas, se implementó una metodología existente para obtener el índice de selectividad al color (CSI) y un parche que muestra el campo receptivo de píxeles que activan a cada neurona (FN,

Feature Neuron). Además, se creó un nuevo método para ordenar los filtros de las capas según su CSI, matiz del FN y valor de activación, pero por filtros en vez de neuronas. Este nuevo método permitía inspeccionar visualmente los filtros. Las pruebas que se realizaron fueron las siguientes:

- Ablación de filtros de las capas según su CSI, para evaluar las diferencias de precisión obtenidas con la finalidad de comprobar sí la mejora en la precisión era debida a un valor alto de CSI.
- Experimento con dos *dataset* de un pájaro: uno con su silueta monocroma y otro con la textura. El objetivo era ver la respuesta a seis tipos de matiz, unos muy presentes en el *dataset* y otros poco presentes, para comprobar si los modelos con LSC eran capaces de mejorar su selectividad al color, extrayendo más características tanto en matices muy presentes como poco presentes.

La Figura 71 muestra los resultados de la diferencia entre el modelo original y el de LSC de la disminución de la precisión por ablación en cuatro rangos de CSI: bajo, medio-bajo, medio-alto y alto. En las tres arquitecturas, la diferencia fue mayor en los rangos altos de CSI, lo cual indicaba que con LSC había una relación entre la mejora de precisión y el aumento de selectividad al color.

En el segundo experimento, se comprobó que había una mejora en las características extraídas en los modelos LSC, sobre todo en matices poco presentes. La Figura 72 muestra un ejemplo de VGG16, que es donde mayores diferencias se encontraron (ver los resultados de los otros modelos en (Sanchez-Cesteros y otros, 2023)). Además, se comprueba que el modelo LSC tiene filtros con selectividad al color en todos los matices y detecta la figura del pájaro completa por el rango del CSI alto.

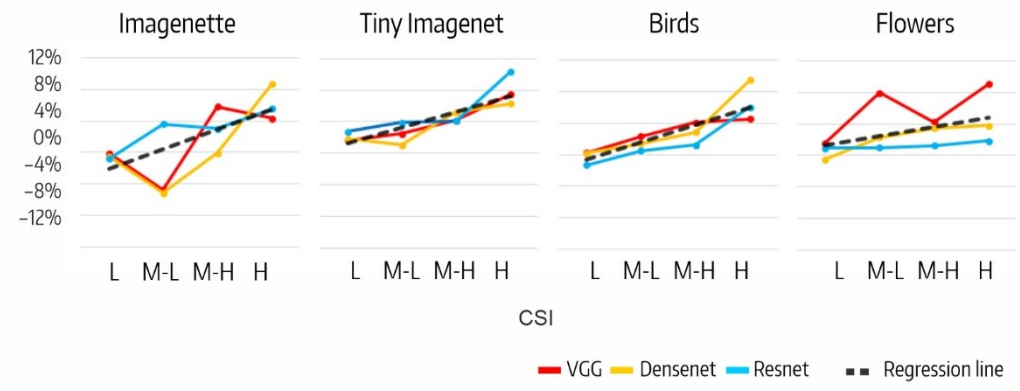


Figura 71 Resultados de la diferencia por la disminución de precisión por ablación.
Nota: entre el modelo original y con LSC (por tipo de modelo y *dataset*).

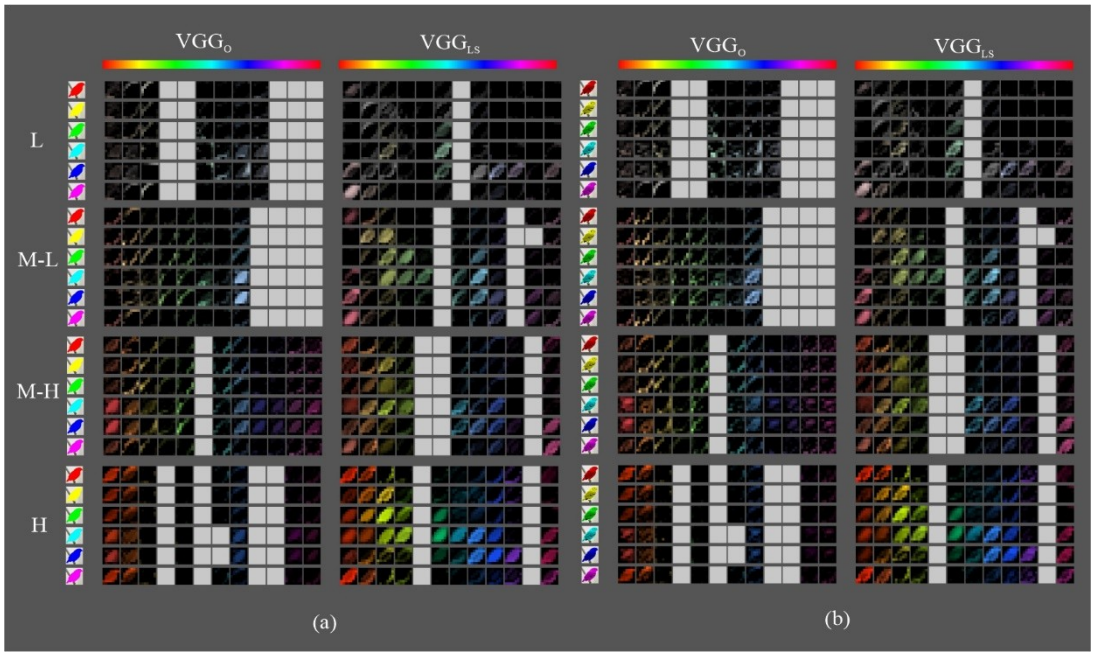


Figura 72 Resultados del experimento del matiz en la selectividad del color
Nota: en la arquitectura VGG16; (a) siluetas monocromas y (b) siluetas con textura.

5.2.2 Conclusiones

La conexión del NGL con el área V4 muestra que existe una interacción directa entre una estructura de información en un nivel de procesamiento bajo sin características visuales y la de extracción de características visuales en un procesamiento de nivel alto. En los experimentos, se demostró que la red aumentaba su selectividad al color gracias a la LSC y que, además, las características visuales mejoraban en el último bloque. La relación entre el tejido interno y la composición se establece de una manera similar, es decir, entre un procesamiento de tensiones (en un nivel semántico bajo) y otro de elementos visuales (en un nivel semántico alto), lo cual invita a concluir que la relación entre ambas estructuras puede ser similar a la analizada en este estudio.

5.3 Clasificación por tipo de composición de imágenes artísticas usando el tejido interno

Para D. Dondis, la composición de una imagen se relaciona con la búsqueda del equilibrio a partir de un proceso por donde, por un lado, se agudiza sus regiones y, por el otro, se nivelan (Dondis, 1974). Este proceso de equilibrado establece, a su vez, una serie de tipologías de composición que Arnheim estableció en cuatro tipos, con la finalidad de poder analizar las tensiones sobre el marco estructural (Arnheim, 1956). En este marco estructural, Arnheim indicaba que existían dos focos de atracción: el centro geométrico y el exterior. En el sistema planteado por Dondis, este marco y los focos de atención establecen los ejes de equilibrio, que pueden desplazarse como si se trataran del centro de una balanza y las regiones agudizadas los pesos.

La relación del marco estructural con la agudización y la nivelación, y con los ejes de equilibrio, determina cuatro clases: la central, la binaria, la jerárquica y la atonal. En la central, la agudización se encuentra en la región central del marco estructural y, por consiguiente, su nivelación es por la propia posición de los pesos visuales al no necesitar de otra región para equilibrar. En la binaria, se establece un equilibrio entre dos regiones agudizadas opuestas a partir de los ejes de equilibrio. En la jerárquica, tres o más regiones agudizadas con intensidades distintas crean una jerarquía entre ellas. En la atonal, gran parte de las regiones estarían agudizadas en un mismo nivel, usualmente medio-alto.

Los límites entre las clases no son tan estrictos y es posible que una composición pertenezca a varias clases simultáneamente en grados distintos. Por ejemplo, una composición jerárquica de tres regiones, con dos regiones con un nivel parecido de agudización y una tercera más bajo, podría pertenecer en un grado alto a la clase binaria y en uno más bajo a la jerárquica.

Detectar las regiones que se agudizan a partir de los elementos visuales (líneas, formas, color, textura, etc.) es una tarea que usualmente realizan expertos en el análisis de composiciones con criterios de juicio de valor, con lo que es difícil obtener una clasificación consensuada y, por lo tanto, etiquetar un *dataset* se convierte en una tarea compleja y con la posibilidad de incluir muchos sesgos. Sin embargo, el tejido interno, que es la estructura subyacente a la composición por debajo de los elementos visuales, puede facilitar la automatización de la tarea de clasificación, al representar las tensiones básicas de la composición en una sola dimensión.

5.3.1 Trabajos relacionados

En visión artificial, el interés por la clasificación de imágenes con composición creativa se ha centrado especialmente en las fotografías. La mayoría de los clasificadores que se han creado han utilizado las tecnologías vigentes en cada momento en visión artificial, con lo que podemos encontrar sistemas que realizan la clasificación por características, por segmentación y más recientemente con redes neuronales artificiales. Todos tienen en común el uso de metodologías, por ejemplo, la regla de los tercios, la localización de la línea de horizonte o las diagonales.

En el sistema propuesto para fotografías de exterior de Lee y colaboradores (Lee J.-T. y otros, 2018), se clasifican en primer lugar en nueve categorías según la regla de los tres tercios, para después usar un clasificador FT_CNN (variante de una red neuronal de convoluciones que aplica un algoritmo basado en la tolerancia de fallos). Cada clase es un detector de un elemento visual de la composición, como la línea de horizonte, las diagonales, las formas triangulares, etc. En otros sistemas utilizan técnicas de recuperación de características visuales en las imágenes para clasificarlas por la composición (Maeda y otros, 1999). En estos sistemas, el objetivo es asociar un tipo de clase a un tipo de composición, a través de la agrupación de todas las características visuales que la definen. Más recientemente, hay sistemas que establecen patrones de la composición

	Tamaño	Activación	Parámetros a entrenar
Input	14x14		0
Flatten	196		0
Dense 1	256	ReLU	50432
Dense 2	512	ReLU	131584
Output	4	Softmax	2052

Tabla 7 Arquitectura de la red neuronal artificial para el clasificador.

basados en la dominancia de algún elemento visual: línea, textura, forma, etc. (Zhang y otros, 2021).

Otros trabajos han evaluado la estética de las imágenes a través de la composición. Destaca OSCAR (Yao y otros, 2012), que permite relacionar las composiciones de fotografías a través de algunas características visuales como la forma, la textura o el color, permitiendo la recuperación de imágenes, tanto en color como monocromas. El sistema utiliza una clasificación por categorías de composición: como horizontal, vertical y central, usando el algoritmo KNN que utiliza la proximidad para hacer clasificaciones o predicciones. Además, se relacionan las características visuales con la etiquetación por feedback de humanos sobre una escala estética. En este tipo de proyectos, se suelen extraer patrones simples como la línea de horizonte, diagonales, o formas como rectángulos o triángulos evidentes en la imagen.

5.3.2 Descripción del clasificador

El objetivo es construir un clasificador con una red neuronal artificial para los tipos de composición, que utilice los mapas de prominencia del tejido interno como entrada en vez de las imágenes. El tejido interno, representado como un mapa de prominencia, tiene una ventaja importante al tener una sola dimensión, ya que se relaciona la agudeza y la nivelación de cada región en una sola prominencia, y no es necesario extraer características visuales. Además, el tejido interno, al representar sólo prominencias, puede facilitar la creación de *datasets* sintéticos en un área, como el de las imágenes artísticas, donde la evaluación es por juicio de valor y la etiquetación por expertos es muy subjetiva.

5.3.2.1 Arquitectura

Un aspecto importante del mapa de prominencia del tejido interno es que es una representación espacial de la imagen y, por consiguiente, la posición de cada prominencia es importante. Esto quiere decir que no es posible usar una CNN, y para el diseño de la arquitectura se usará una red con capas *dense* (*fully-connected*) que mantenga la relación espacial de las prominencias como una característica importante.

Para facilitar el procesamiento de los mapas en la red neuronal, el tamaño que utiliza el modelo como entrada es de 14x14, que es un tamaño suficiente para procesar la información del tejido interno (global y sin detalles). Para facilitar la transición entre las prominencias, se aplica la función kriging al mapa. La arquitectura del modelo (ver Tabla 7) tiene una capa de entrada de 14x14x1 que es redimensionada a un vector de 196 neuronas, y utiliza dos capas: la primera de 256 neuronas y la segunda de 516. La función de activación es ReLU, no se aplican *bias* y los pesos se inician con la distribución Gorot uniforme. La capa de salida es *softmax*, para tener un la probabilidad de cada clase.

5.3.2.2 Conjunto de datos

La ventaja de usar los mapas de prominencia del tejido interno es que permite diseñar casos sintéticos aplicando reglas con facilidad. Estas reglas sirven para posicionar las prominencias de acuerdo con cada clase y pudiendo controlar la cantidad y las distancias.

Clase central

Para crear los casos de la clase central, se sitúan un conjunto de prominencias en la región central aleatoriamente con niveles altos y otro conjunto con valores bajos en el resto del espacio. Los umbrales de los valores de prominencia, así como el tamaño de la región central, son parámetros configurables para ajustar en el generador.

Para establecer la posición de las prominencias de la región central, creamos la matriz $C_{r,\theta}$ en coordenadas polares, donde el valor de cada r es:

$$r = r_{min} + (r_{max} - r_{min}) * \gamma \tag{47}$$

siendo r_{min} la distancia mínima que puede ocupar y r_{max} la distancia que limita la región central, e γ un número aleatorio en el intervalo $[0,1]$. El ángulo θ es:

$$\theta = \theta_{min} + (\theta_{max} - \theta_{min}) * \gamma \tag{48}$$

siendo:

$$\theta_{min} = \frac{2\pi}{N} * i \tag{49}$$

$$\theta_{max} = \frac{2\pi}{N} * (i + 1) \tag{50}$$

donde i es cada prominencia y N el total de prominencias que habrá en la zona central.

El valor de la prominencia será:

$$C_{r,\theta} = (1 - A_{alto}) * \gamma \quad (51)$$

donde A_{alto} es el valor mínimo para una prominencia alta. El valor de prominencia alta está en el intervalo $[A_{alto}, 1]$.

Para el resto de las prominencias, que tendrán un valor bajo y que completarán el mapa, creamos la matriz $E_{r,\theta}$ donde r puede ser cualquier distancia aleatoria γ, y θ se obtiene con la ecuación (48), pero dependiendo de la cantidad N de prominencias bajas que se estimen necesarias. El valor de prominencia se obtiene aplicando la siguiente ecuación:

$$E_{r,\theta} = A_{bajo} * \gamma \quad (52)$$

donde $A_{bajo} < A_{alto}$.

Clase binaria

La clase binaria se caracteriza por tener dos regiones prominentes opuestas en relación con un eje (horizontal, vertical o diagonal). Su creación sigue el mismo proceso que la clase central, pero se diferencia en que sólo generamos dos prominencias altas y opuestas en su posición. La primera prominencia se crea aplicando las ecuaciones (47) y (48) para obtener la posición en r y θ, y la segunda, opuesta a la primera en θ. Los valores de prominencia serán aleatorios en el intervalo $[A_{alto}, 1]$. El resto de las prominencias estarán posicionadas usando $E_{r,\theta}$.

Clase jerárquica

La clase jerárquica sigue el mismo criterio que la binaria, pero añadiendo una tercera prominencia y sin aplicar la oposición en θ. La distancia r es aleatoria como la central y la binaria para las tres prominencias, siendo el mínimo y el máximo parametrizables para que no estén muy cercanas. Los valores de prominencia dependen de un valor aleatorio dentro de tres intervalos parametrizables: alto en $[A_{alto}, 1]$, medio-alto en $[A_{medoAlto}, A_{alto}]$ y medio en $[A_{medio}, A_{medioAlto}]$. El resto de las prominencias estarán posicionadas usando $E_{r,\theta}$ con $A_{bajo} < A_{medio}$.

Clase atonal

La clase atonal se caracteriza por tener más de tres regiones con prominencia media, usando los mismos criterios que las otras clases, con un valor $N > 3$ y la distancia aleatoria en todo el mapa. Los valores de prominencia son aleatorios en el intervalo $[A_{medio}, A_{medioAlto}]$. El resto de las prominencias estarán posicionadas usando $E_{r,\theta}$ con $A_{bajo} < A_{medio}$.

5.3.3 Descripción experimentos

Al no existir clasificadores previos o *datasets* etiquetados para tipos de composición, para evaluar el clasificador se plantea el uso de dos *datasets* con pinturas de distintos periodos y estilos (desde el Renacimiento hasta el arte Pop del siglo XX) de artistas relevantes y otro de artistas no relevantes.

La sintaxis visual de Dondis define la composición como un equilibrio de pesos visuales, donde el dinamismo se consigue a partir de la agudización y la nivelación, es decir, a partir de crear desequilibrio para posteriormente buscar el equilibrio. En esta definición, las imágenes con composiciones atonales son las que más dinamismo deben obtener y, por consiguiente, son las que más se utilizan, a diferencia de las centrales, donde el dinamismo se reduce a la zona central, que como Arnheim indica es la que produce quietud, y Dondis define como «aburrimiento». En la evaluación de la clasificación, el porcentaje de composiciones atonales tiene que ser superior al de composiciones centrales. Para las clases binaria y jerárquica, la distribución será menor, pero ambas deben de tener proporciones similares, ya que comparten muchas características comunes.

La evaluación consistirá en un estudio estadístico de los porcentajes de composiciones en cada clase y de un análisis visual de tipo experto, para comprobar que se cumplen los planteamientos de Arnheim y Dondis.

5.3.3.1 Generación del *Dataset* para el entrenamiento del clasificador

El *dataset* se genera con 4000 casos, 1000 por clase, siendo una cantidad adecuada para la arquitectura del clasificador y las cuatro clases. Para el *dataset* de entrenamiento se utilizan 800 por cada clase (un 80% del total) y para el de validación 200 (un 20% del total) como es usual. Los parámetros utilizados para la generación de los casos son los siguientes:

- Clase central:
    - El tamaño de la imagen es 14x14: $ancho = 14$.
    - Para la región central: $N = 6$.
    - Para las regiones de prominencia baja: $N = 12$.
    - La distancia mínima de la región central: $r_{min} = ancho * 0.01$.
    - La distancia máxima de la región central: $r_{max} = ancho * 0.4$.
    - El umbral mínimo de la prominencia alta: $A_{alto} = 0.8$.
    - El umbral máximo de la prominencia baja: $A_{bajo} = 0.4$.
- Clase binaria:
    - El tamaño de la imagen es 14x14: $ancho = 14$.
    - Para las regiones de prominencia baja: $N = 12$.

- La distancia mínima de la región binaria: $r_{min} = ancho * 05$.
- La distancia máxima de la región binaria: $r_{max} = ancho * 0.7$.
- El umbral mínimo de la prominencia alta: $A_{alto} = 0.8$.
- El umbral máximo de la prominencia baja: $A_{bajo} = 0.4$

- Clase jerárquica:
  - El tamaño de la imagen es 14x14: $ancho = 14$.
  - Para las regiones de prominencia baja: $N = 12$.
  - La distancia mínima de la región jerárquica: $r_{min} = ancho * 03$.
  - La distancia mínima de la región jerárquica: $r_{max} = ancho * 0.7$.
  - El umbral mínimo de la prominencia alta: $A_{alto} = 0.9$.
  - El umbral mínimo de prominencia medio-alto: $A_{medioAlto} = 0.7$.
  - El umbral mínimo de prominencia medio: $A_{medio} = 0.5$.
  - El umbral máximo de la prominencia baja: $A_{bajo} = 0.4$.
- Clase atonal:
  - El tamaño de la imagen es 14x14: $ancho = 14$.
  - Para la región atonal: $N = 8$.
  - Para las regiones de prominencia baja: $N = 12$.
  - La distancia mínima de la región atonal: $r_{min} = ancho * 03$.
  - La distancia mínima de la región atonal: $r_{max} = ancho * 0.7$.
  - El umbral mínimo de prominencia medio-alto: $A_{medioAlto} = 0.7$.
  - El umbral mínimo de prominencia medio: $A_{medio} = 0.5$.
  - El umbral máximo de la prominencia baja: $A_{bajo} = 0.4$.

La imagen final se genera como un mapa kriging con la posición de las prominencias. La Figura 73 muestra 25 ejemplos para cada clase, donde se ve que los límites no son muy estrictos entre la binaria y la jerárquica y entre la jerárquica y la atonal, pero muy estrictos entre la central y la atonal. La diferencia entre binaria y jerárquica es una tercera región de prominencia media o la simetría de dos regiones con prominencia media-alta y alta.

El modelo es entrenado con los siguientes criterios: batch de 32, ratio de aprendizaje de 0.001 y con 12 epochs, obteniendo un 96% de precisión en el *dataset* de validación.

5.3.3.2 *Datasets* para la evaluación: «Best Artworks of All Time» (BAAT) y Midjourney

Uno de los principales *datasets* de pinturas es Wikiart (Saleh & Elgammal, 2015), pero para la finalidad de este experimento, excede en la cantidad de autores y, además, incluye artistas de diferente repercusión en la historia del arte. Sin embargo, hay un *dataset* que incluye a 50 autores de distintas épocas y estilos, «Best Artworks of All Time»

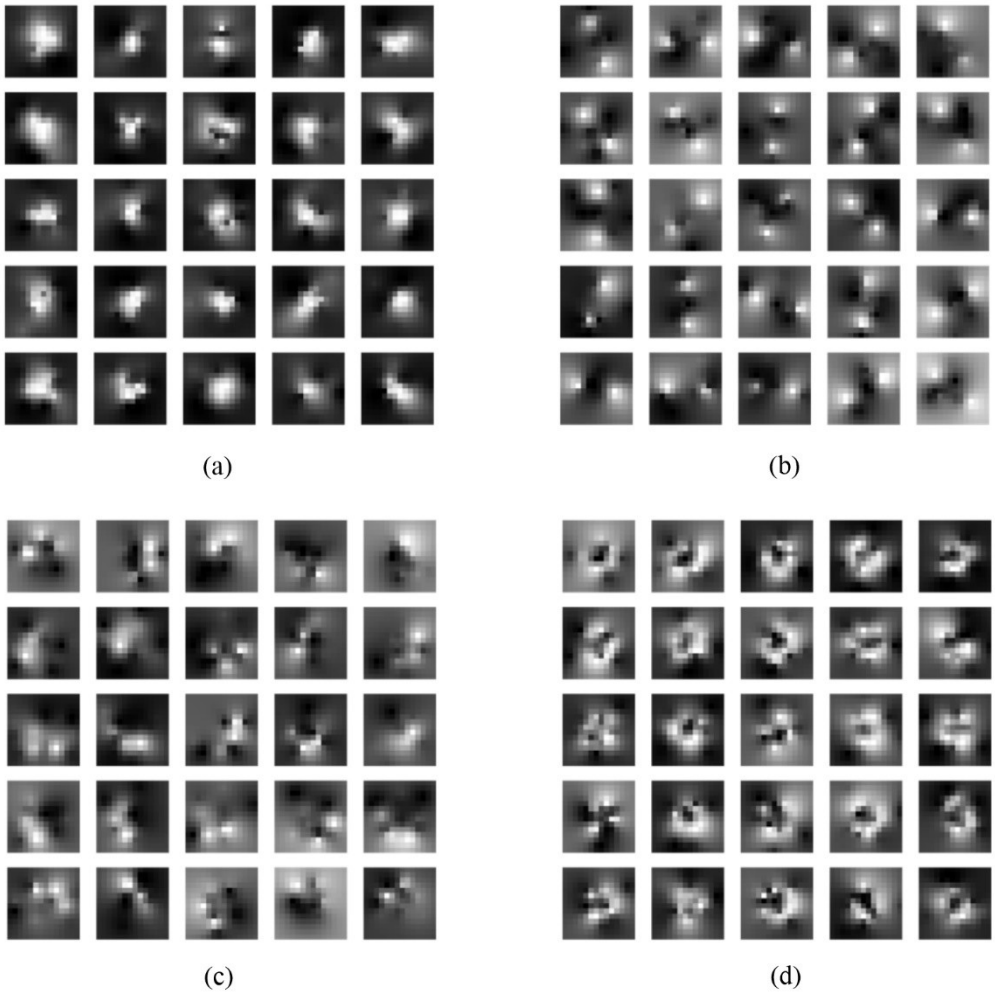


Figura 73 Ejemplos de mapas de prominencia generados para cada clase
Nota: (a) central; (b) binaria: (c) jerárquica; y (c) atonal.

(BAAT) (Icaro, 2023), todos con una gran repercusión en la historia del arte, tanto por su calidad como por su influencia. Por otro lado, incluimos un *dataset* que recopila imágenes generadas en Midjourney en 2022 (LDMTWO, 2023), un modelo de IA generativa entrenado con millones de imágenes capturadas en Internet de distintas fuentes, incluyendo artistas relevantes y noveles.

BAAT tiene 8436 pinturas de 50 artistas, 44 estilos distintos y 31 países (entre Europa y América). La cantidad de pinturas varía según el artista, siendo Vicent van Gogh el que tiene la mayor cantidad con 877 y Jackson Polloc el que menos con 24. Esto no tiene una especial repercusión para el objetivo del experimento. El segundo *dataset*, imágenes generadas por Midjoruney, se utiliza una versión reducida (LDMTWO, 2023), que incluye 3842 imágenes extraídas de la primera versión.

La Figura 74 muestra una selección aleatoria de 100 imágenes de BAAT y la Figura 75 de Midjoruney. En ambos *datasets*, hay imágenes apaisadas, verticales, retratos, paisajes, abstractas, escenas de interior y exterior, primeros planos, con diversidad de estilos y acabados, monocromos y polícromos, etc.

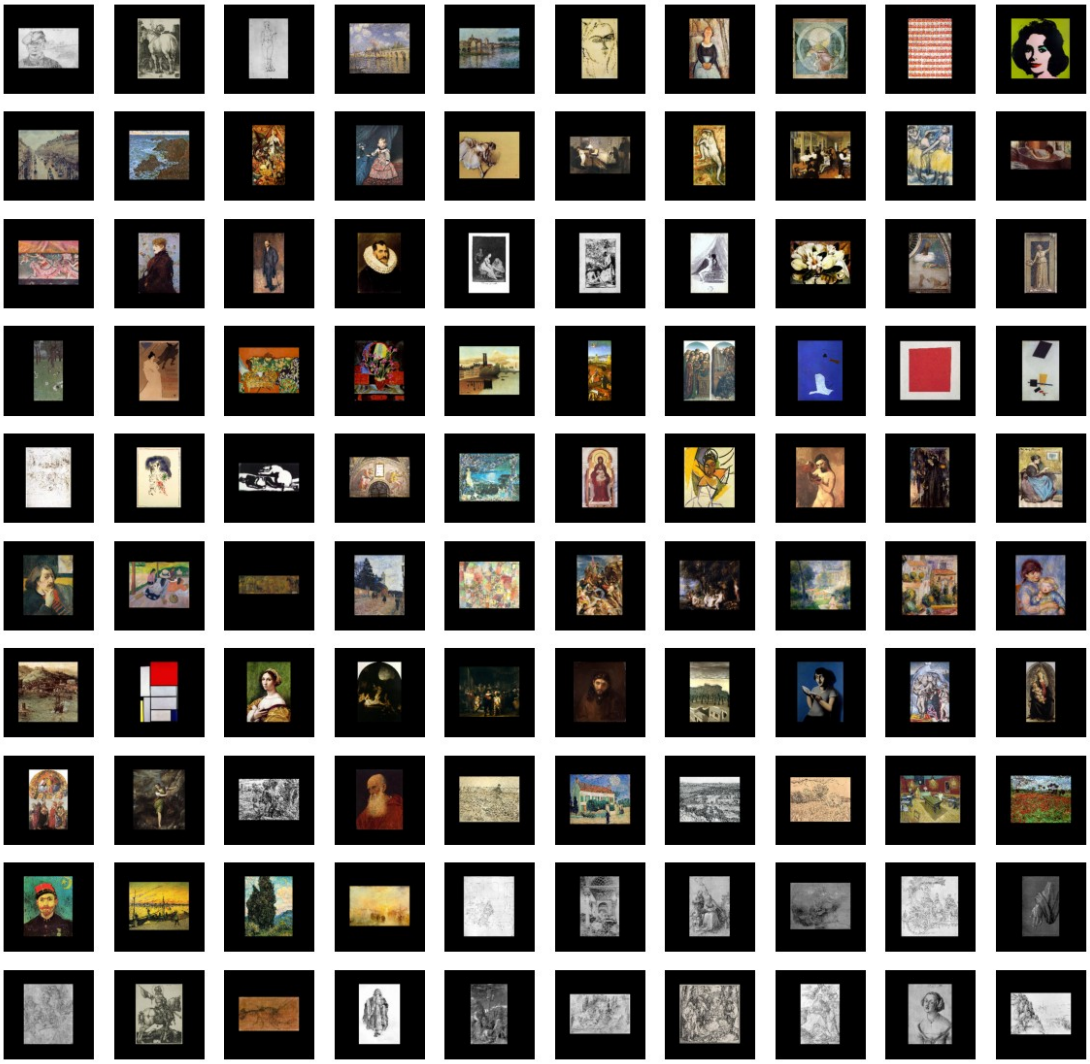


Figura 74 Ejemplos de imágenes del *daataset* BAAT.
Nota: existe una variedad temática y formal.

Figura 75 Ejemplos de imágenes del *dataset* Midjourney.
Nota: existe variedad temática y formal.

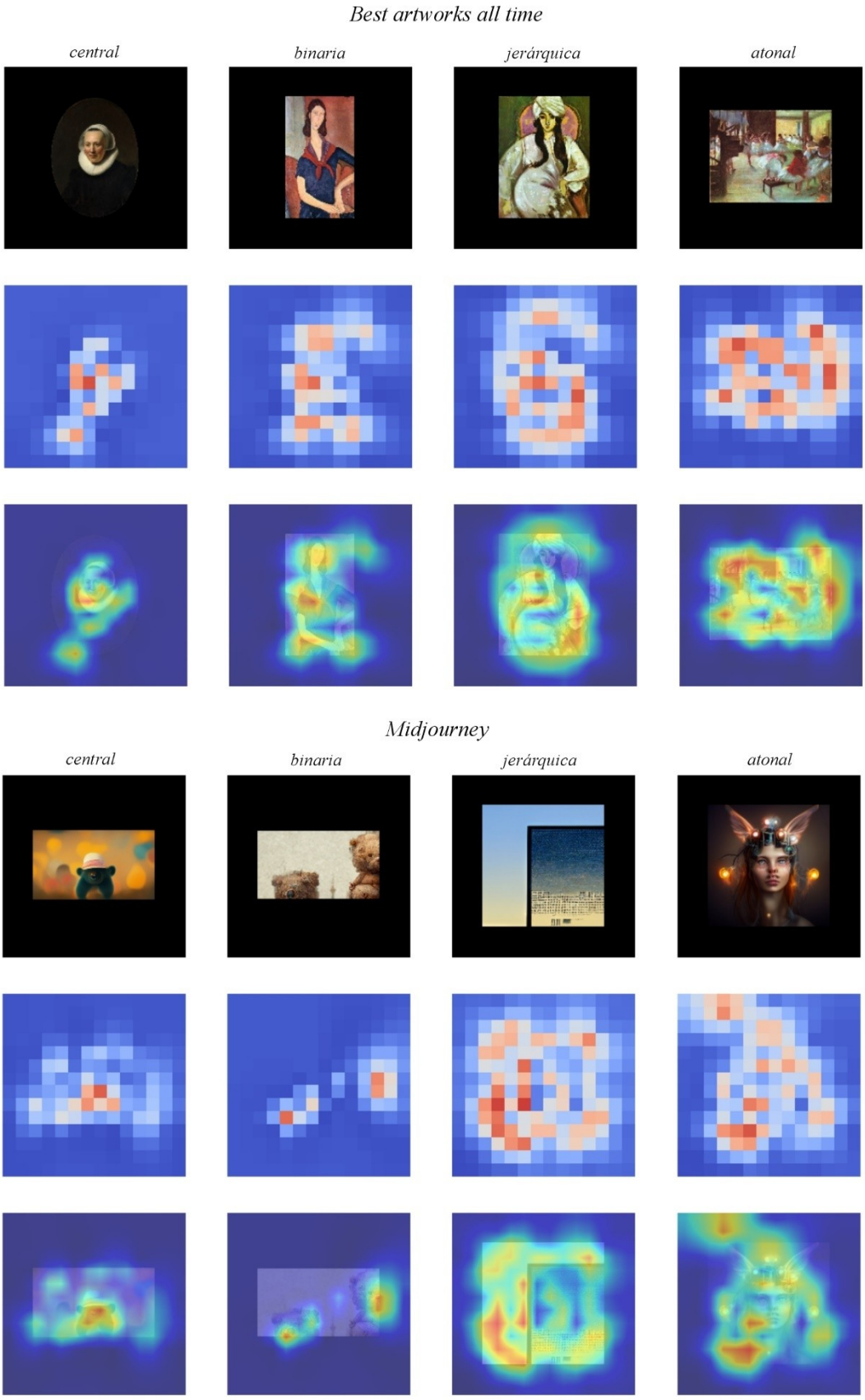


Figura 76 Ejemplos para cada clase de BAAT y Midjouney.
Nota: primera fila, imagen; segunda fila, mapa del tejido interno; y tercera fila: mapa de calor del tejido interno sobreimpreso a la imagen.

Dataset	**Central**	**Binaria**	**Jerárquica**	**Atonal**
BAAT	0.3 %	10.48%	31.28%	57.92%
Midjourney	0.55%	20.57%	22.34%	56.52%

Tabla 8 Porcentaje de imágenes por clase y *dataset*

5.3.4 Resultados experimentales

La Tabla 8 muestra el porcentaje de imágenes por cada clase en cada *dataset*. La mayor cantidad de imágenes fueron clasificadas en la clase atonal, seguido de la jerárquica y la binaria. La clase central es la que menor porcentaje obtiene. Los resultados son muy similares en ambos *datasets* en la clase central y atonal, y con diferencias de un 20% en las clases binaria y jerárquica en BAAT, y de un 2% en Midjourney.

La Figura 76 muestra un ejemplo de imagen por cada clase en ambos *datasets*. Para facilitar el análisis, se incluye el mapa de prominencia del tejido interno y una superposición de la imagen con este mapa. En ambos *datasets*, se puede comprobar cómo las imágenes de cada clase se ajustan a la definición de central, binaria, jerárquica y atonal. Se puede observar que la complejidad de la composición aumenta con claridad desde la central hasta la atonal. En la clase central, el contraste entre la figura y el fondo es muy alto, tanto en la pintura (un retrato en un formato circular) como en la infografía digital. En la clase binaria, es más evidente en los dos osos que en el retrato de la joven, donde está muy cerca de la clase jerárquica. El edificio encaja perfectamente en la clase jerárquica si analizamos su composición con tres elementos visuales: cristales de la fachada, cielo azul y el reflejo solar. En el caso del retrato de la mujer de BAAT, la composición en forma de «S» facilita que exista una jerarquía. Por último, la clase atonal es más evidente en los ejemplos de ambos *datasets*, ya que existen más de tres elementos visuales prominentes.

5.3.5 Discusión

Los resultados de la clasificación de ambos *datasets* muestran que el mayor porcentaje de imágenes es en la clase atonal (57.92% y 56.52% respectivamente), mientras que el menor es en la central (0.3% y 0.55%, respectivamente). La preferencia por una composición más dinámica, como la atonal, es común tanto para los artistas relevantes como para el generador entrenado con imágenes de diversas fuentes. Esta diferencia es bastante grande, e indica que el dinamismo en la composición es un elemento importante, tanto que la preferencia por tejidos internos con muchas regiones prominentes es muy superior a la de una prominencia central.

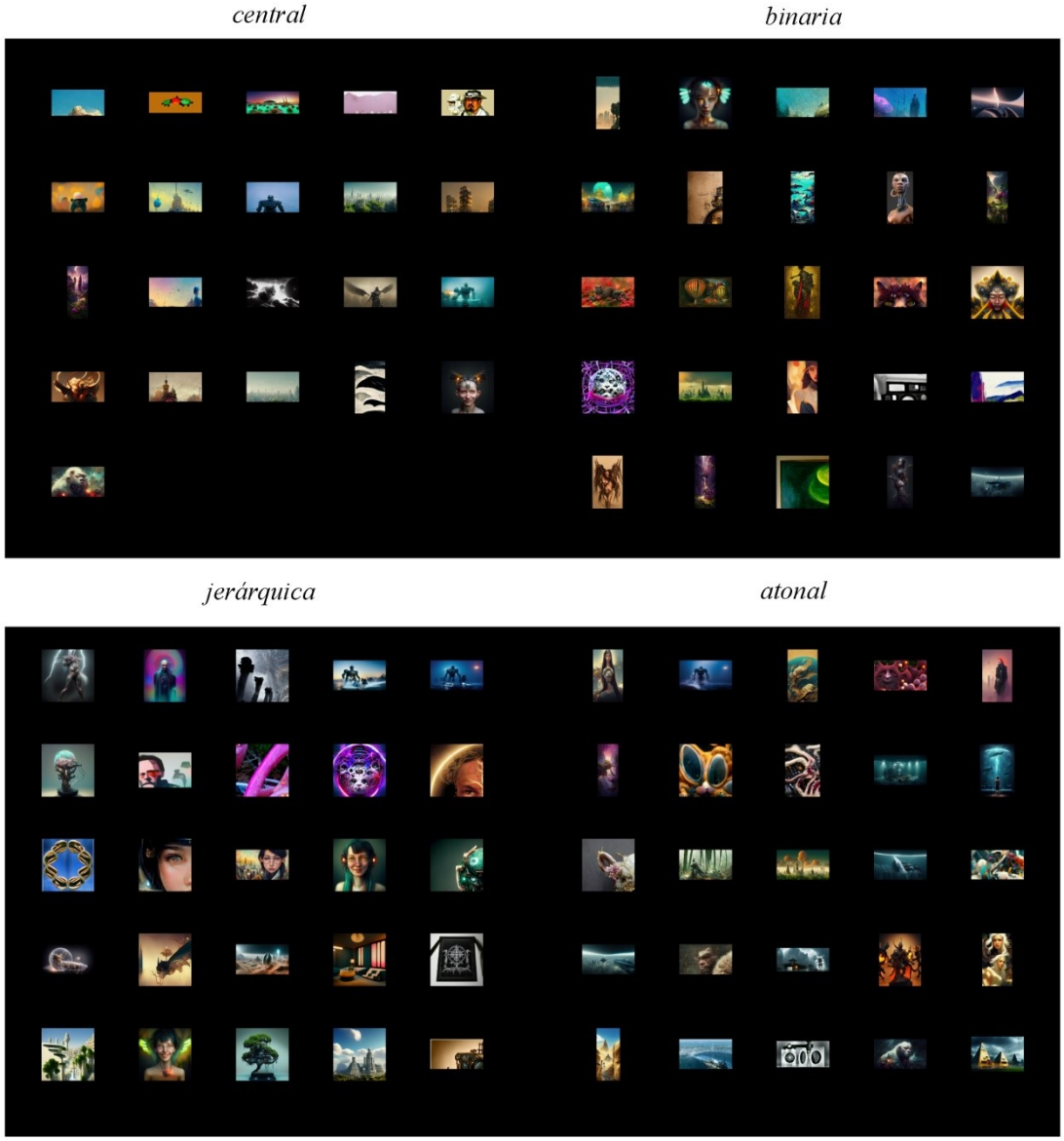


Figura 77 Resultados por tipo de clase en el *dataset* Midjoruney.

Por otro lado, la inspección visual en las imágenes clasificadas permite comprobar que hay una relación entre los mapas y las clases (Figura 76). La complejidad de las composiciones (elementos visuales presentes y estructuras de la composición) aumentan desde la clase central a la atonal, y, si bien las clases central y atonal no arrojan duda en su clasificación, en la binaria y en la jerárquica los límites son más abiertos. Las Figura 78 y Figura 77 muestran más ejemplos de imágenes de ambos *datasets*, donde la comparación entre la sencillez de la clase central se contrapone con la complejidad de la atonal. En la inspección visual, las imágenes clasificadas se ajustan a las clases, sobre todo a la clase central y atonal. Aunque, como ya se ha visto, las clases binaria y jerárquica son la que más ambigüedad ofrecen.

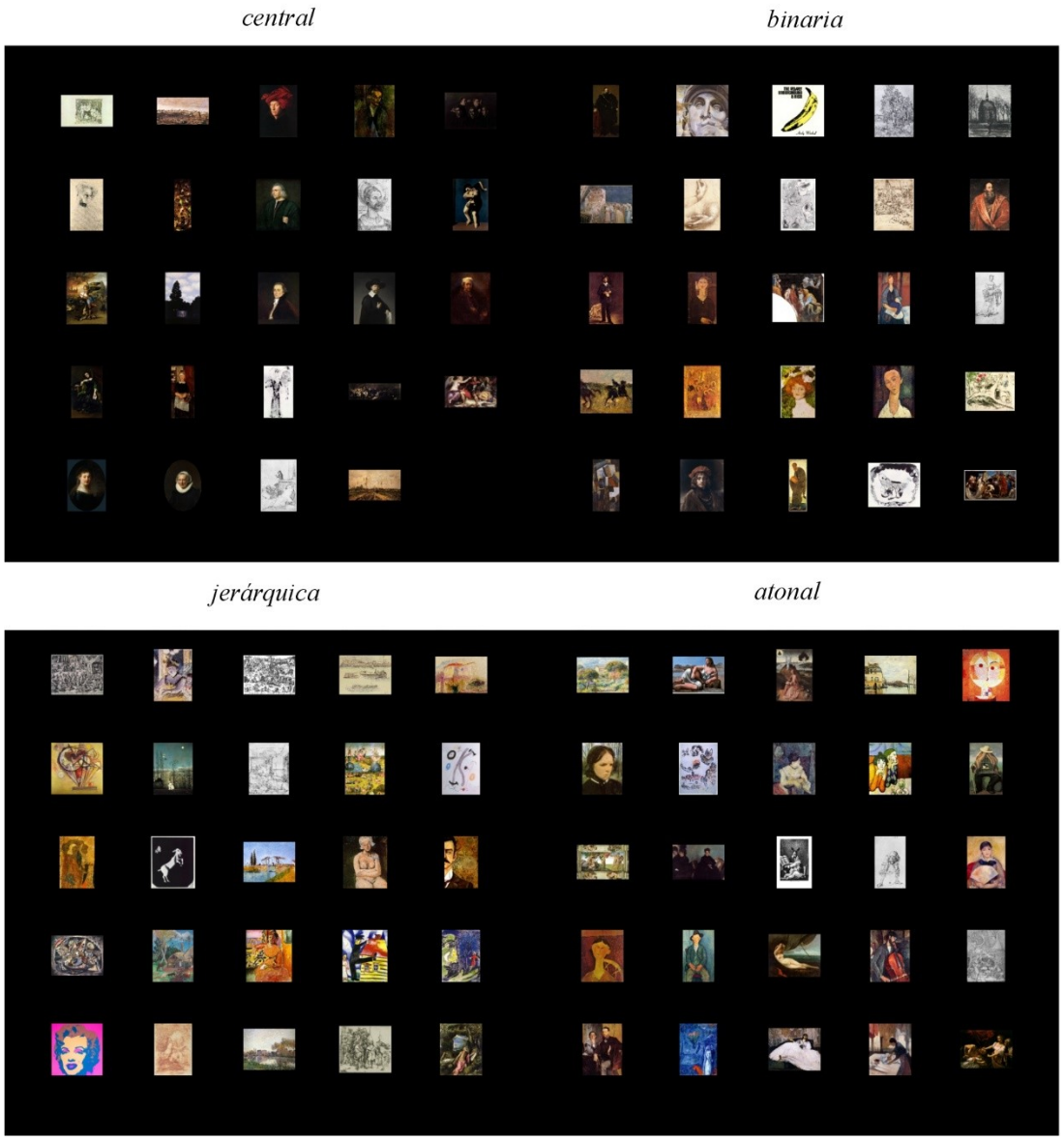


Figura 78 Resultados por tipo de clase en el *dataset* BAAT.

5.3.6 Conclusiones

La representación como mapa de prominencia del tejido interno de la composición facilita tareas como la clasificación en la tipología de Arnheim, ya que simplifica la representación usando prominencias en vez de las características visuales. En este caso, además, posibilita la generación de un *dataset* sintético que facilita la labor de etiquetación subjetiva que sería necesaria, ya que, a partir de un conjunto de reglas de generación se puede construir los casos etiquetados.

La evaluación se ha realizado con dos *datasets* distintos, uno de autores relevantes de todas las épocas y otra generada con modelos de IA entrenados con *datasets* de imágenes obtenidas en Internet desde fuentes muy diversas, algunas con autores relevantes y otras más noveles. El porcentaje de cada clase es coincidente en ambos *datasets* y se re-

laciona con el hecho de que un tipo de composición plantea un mayor dinamismo que otros, y es elegida preferentemente a la hora de componer. En la inspección visual de los resultados, se ha observado aspectos comunes en las composiciones de las imágenes en ambos *dataset* según cada clase, siendo más simples las centrales que las atonales.

Con este clasificador podemos obtener, a partir del mapa del tejido interno de una imagen, la clase de composición sin tener que recurrir a algoritmos complejos que tengan que analizar las características visuales como contornos, formas, colores, texturas, etc. Evidentemente, esto implica una ventaja sustancial en el procesamiento de este tipo de imagen, ya que, conocer su clase facilita operaciones más complejas sobre su composición o sobre la extracción de más información.

5.4 Búsqueda de imágenes por el tejido interno

La búsqueda de imágenes artísticas a partir de las características visuales es una tarea compleja, tanto para un experto como para una aplicación informática. Localizar una pintura de Picasso en una base de datos de arte, es una tarea relativamente fácil, si existe una etiquetación elaborada previamente. Pero localizar una pintura similar por su composición, con una imagen como criterio de búsqueda, plantea problemas técnicos al tener que relacionar la extracción de características visuales con la estructura de una composición. Si, además, no buscamos una imagen parecida, ni con la misma gama de color, ni los mismos elementos visuales, y tampoco con la misma temática, es incluso más complejo.

Uno de los principales problemas es la etiquetación de las características visuales, ya que necesitan expertos muy cualificados, tanto por la complejidad de extraer las características como por la de discriminar las que son relevantes. En este campo, existen soluciones que permiten la búsqueda visual a partir de imágenes utilizando técnicas de CBIR (*Content Based Image Retrieval*) y el avance del campo es prometedor, sobre todo tras los éxitos recientes en la visión artificial.

La composición de una imagen es una cuestión de interés tanto para los expertos que estudian la historia del arte como para los propios creadores, donde la búsqueda visual facilita, tanto el análisis de obras de arte a partir de obras anteriores como el estudio a partir de obras similares. Sin embargo, la complejidad de analizar las composiciones y su representación dificulta una búsqueda visual, ya que, a diferencia del contenido visual que se extrae desde los píxeles, la composición es una interpretación a partir de la imagen. Además, hay que tener en cuenta, que todos estos procesos se relacionan con aspectos formales de la imagen (color, textura, formas, movimiento, espacio, etc.) que el creador ha utilizado para componer.

El tejido interno, en su representación en forma de mapa de prominencia, puede facilitar las tareas de búsqueda en este contexto, tanto por el hecho de que sólo tiene una dimensionalidad que indica la capacidad de atracción de cada región, como porque mantiene las relaciones espaciales originales. En este apartado, se va a construir un motor de búsqueda visual que, a partir de una imagen como criterio de búsqueda, recupere

imágenes de un *dataset* con tejidos internos similares, los cuales son una representación subyacente de las composiciones.

5.4.1 Trabajos relacionados

Desde el surgimiento de la Web a partir de los años 90 y, sobre todo, con la irrupción de las cámaras digitales al inicio del siglo XXI, las bases de datos de imágenes han crecido exponencialmente hasta límites insospechables en décadas anteriores. Por esa razón, desde los inicios, surgieron soluciones para la gestión de estas bases de datos como, por ejemplo, Webserver de 1996 (Frankel y otros, 1996) que realizaba búsquedas a partir de imágenes con técnicas de recuperación basadas en el contenido (CBIR, *Content Based Image Retrieval*). Esta iniciativa utilizaba algoritmos de visión artificial aplicando técnicas de vecindad, filtros, o umbrales a los píxeles.

Los motores de búsqueda se convirtieron en un campo floreciente en una época temprana de la Web, incluso antes del surgimiento de Google. En el estudio del campo de CBIR de 1999, *Image Retrieval: Past, Present and Future* (Rui y otros, 1999), se dedicaba un apartado a los motores de búsqueda que usaban una imagen como criterio de búsqueda. Estos motores buscaban las imágenes a partir de una serie de características visuales, como la similitud de colores, texturas, formas, etc. En los siguientes años, se crearon buscadores a partir de los avances en el campo de la visión artificial como el uso de Pagerank (Jing & Baluja, 2008), de *Bag of Features* en (Jégou y otros, 2010), de CNNs (Babenko y otros, 2014), o, más reciente, Vision Transformer (Baldrati y otros, 2022).

Otra variante interesante, es la búsqueda de imágenes a partir de esquemas realizados a mano, que fue implementada en los años 90 en la aplicación Magic Brush de Microsoft y, recientemente, en (Sabry y otros, 2023), usando en ambos casos los *autoencoders*. El uso de esta técnica se aplica en proyectos donde hay pocos casos etiquetados como en (Petscharnig y otros, 2017), mezclando técnicas del aprendizaje automático con el procesamiento de las imágenes con la finalidad de extraer características relevantes. Esto indica que el campo sigue en evolución y adaptándose a los avances tecnológicos, tanto en la creación de las bases de datos como en la tecnología para extraer características de las imágenes.

Dentro del campo de las imágenes artísticas, existen varias iniciativas recientes que parten de los avances de los motores de búsqueda que hemos comentado. En 2016, Leonardo y colaboradores estudian la posibilidad de utilizar las CNN para localizar patrones visuales en bases de datos de pinturas destinados a historiadores del arte (di Lenardo y otros, 2016). En el proyecto de Seguin y colaboradores, se plantea utilizar la capacidad de las CNN para localizar patrones visuales con independencia de su posición en las imágenes gracias a la convolución (Seguin y otros, Tracking Transmission of Details in Paintings., 2017),. En «*The Replica Project*» (Seguin, 2018), se implementa la construcción de un motor de búsqueda visual para historiadores del arte. *Neural-based cross-model* (Gong y otros, 2023) utiliza técnicas recientes como el uso de texto+imagen aplicando VSE (*Visual Semantic Embed*) para obtener representaciones de las imágenes

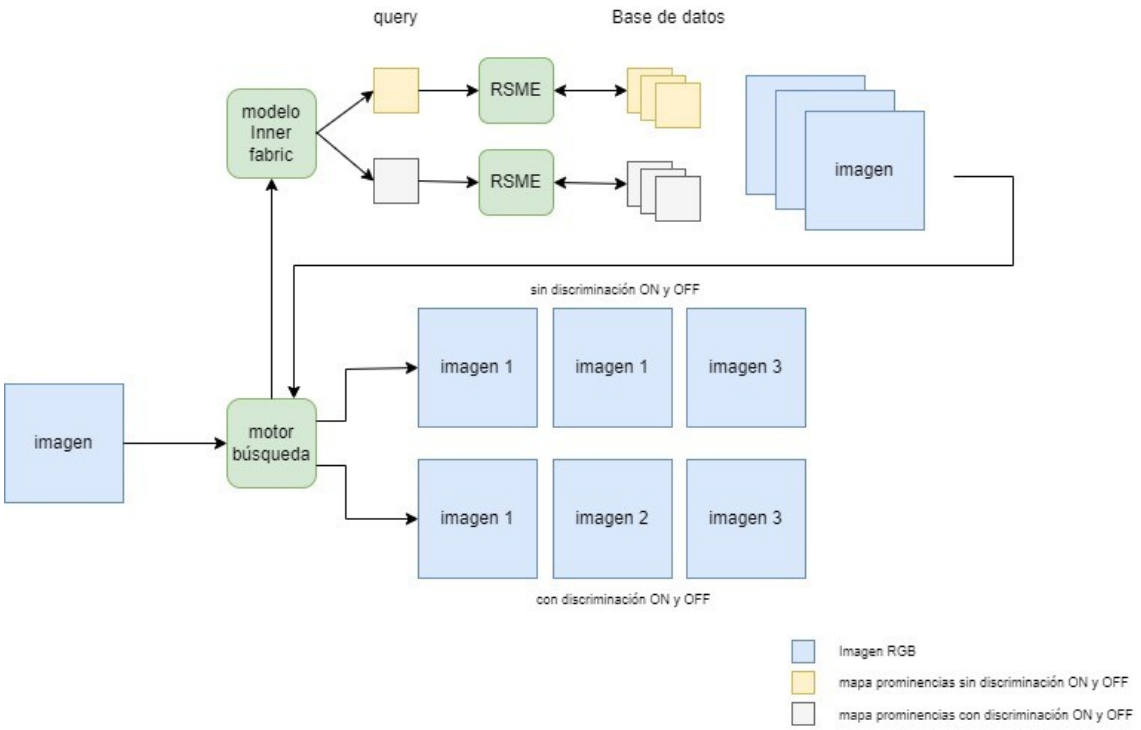


Figura 79 Esquema general del motor de búsqueda.

en forma de encoder, donde se asocia las descripciones visuales y semánticas y se facilita la búsqueda.

5.4.2 Descripción del modelo

EL uso de los mapas del tejido interno como criterio de búsqueda es algo novedoso y que puede facilitar la recuperación de imágenes con composiciones parecidas. La Figura 79 muestra el esquema del motor de búsqueda diseñado para esta finalidad. En la primera parte, el motor obtiene los mapas de prominencia del tejido interno desde la imagen con Inner Fabric, los cuales serán los criterios de búsqueda*:* uno con discriminación ON y OFF y otro sin discriminación. En la segunda parte, se crea una lista ascendente de los mapas según el RMSE (Raíz Cuadrada del Error Cuadrático Medio) entre cada mapa de tejido interno del *datasets* y el criterio de búsqueda. Para este experimento se utiliza un método de cálculo de error estándar, pero puede ser sustituido por otros métodos y algoritmos dependiendo de los resultados que se quieran obtener en la búsqueda. En la tercera parte, se devuelve la lista ordenada recuperando las imágenes de la base de datos correspondientes a los mapas como resultado de la búsqueda. El buscador permite usar mapas con discriminación ON y OFF y sin ella.

5.4.3 Descripción del experimento

Para poder evaluar el funcionamiento del buscador, realizaremos búsquedas con 50 imágenes elegidas al azar en el datase de obras de arte BAAT (Icaro, 2023).

En cada búsqueda, evaluaremos los resultados con los siguientes criterios heurísticos:

- Formato de la imagen, si es horizontal o vertical, y si se mantienen las proporciones o no del criterio de búsqueda.
- Características visuales presentes: líneas, colores, texturas, etc. El objetivo es comprobar que se obtengan resultados que no sean homogéneos a las características visuales del criterio de búsqueda.
- Estructura general de la composición. El objetivo es que las composiciones de los resultados mantengan la estructura general similar al criterio de búsqueda (si es horizontal, vertical, circular, diagonal, etc.).

Una vez comprobado cada resultado, se obtendrá una media del porcentaje de cumplimento de los requisitos en los resultados de un conjunto de imágenes del *dataset*. Para tener un resultado estadístico, se evaluarán cincuenta imágenes seleccionadas al azar.

5.4.4 Evaluación y análisis de los resultados

La Tabla 9 muestra el porcentaje medio del cumplimiento de cada criterio heurístico para las cincuenta búsquedas realizadas. Para los mapas sin discriminación ON y OFF, el requisito de formato similar con la imagen como criterio de búsqueda se cumple en un 94 % y el de que no exista una similitud de las características visuales en un 88 %. Ambos son porcentajes altos, sin embargo, el requisito de estructura general se cumple sólo en un 59%. Para las búsquedas que usan la discriminación de ON y OFF, los resultados son inferiores para el formato (73%) y, también, para las características visuales distintas (64%). Sin embargo, la similitud con la estructura básica de la composición aumenta hasta un 65%. En el primer caso, cuando no hay discriminación, las regiones prominentes pueden ser tanto en ON y OFF, lo cual implica que no exista una similitud en las características visuales. Esto justifica la diferencia de un 14% cuando la búsqueda sí discrimina ON y OFF. La diferencia más llamativa está en el formato, ya que a priori, la discriminación de ON y OFF debería tener una media mayor. Inspeccionando las imágenes (ver Figura 80, Figura 81 y Anexo 1) se puede observar que la discriminación de ON y OFF es más compleja de obtener (por o menos más exigente) lo cual debe incidir en la dificultad de encontrar mapas similares tanto en un mismo formato como con las características visuales parecidas.

La Figura 80 muestra un ejemplo de búsqueda con mapas con y sin discriminación de ON y OFF. La primera imagen (parte superior izquierda) es el criterio de búsqueda, una imagen de un cuadro de Degás, y las quince siguientes, los resultados por orden ascendente. En la búsqueda sin discriminación de ON y OFF, la similitud del mapa de la imagen como criterio de búsqueda con los resultados es mayor que con los mapas con

discriminación, sobre todo a partir de la tercera imagen. El formato de las imágenes, sin discriminación de ON y OFF, es bastante similar en todos los casos, mientras que cuando hay discriminación, sólo las imágenes que tienen mapas más similares mantienen un formato parecido. Esta diferencia entre los mapas de los resultados también tiene una repercusión, sobre todo, en la estructura general de la composición. En los mapas de la imagen como criterio de búsqueda, las regiones prominentes están en la parte superior, tanto en la izquierda como en la derecha, y en la parte inferior izquierda. La región central tiene una prominencia media, y una prominencia baja el entorno, tanto por la derecha como por la izquierda. Este patrón representa una composición donde existe una prominencia en la parte superior (grupo de jinetes) y parte inferior izquierda (área de hierba del primer plano). En el caso de esta imagen estas prominencias están en ON (es visible en el mapa que discrimina ON y OFF), y lo podemos comparar con la última imagen, la pintura de Modigliani, donde la prominencia está en OFF (área marrón oscuro del fondo detrás del desnudo). Ambas imágenes tienen la misma estructura básica en su composición (diagonal que va de la parte superior izquierda a la inferior derecha), pero sus características visuales (línea, forma, color y textura) son distintas. Esto se puede comprobar con el resto de las imágenes en distintos niveles.

En el caso de los mapas con discriminación ON y OFF, sólo los que mantienen una similitud en las características visuales tienen una misma estructura básica, compuesta por la diagonal que ya hemos comentado y, además, una temática parecida. Es decir, las imágenes que mantienen un mapa muy similar con discriminación ON y OFF son paisajes: 1, 2, 5, 9, 12, 13 y 14 (leyendo de arriba hacia abajo y de izquierda a derecha, después de la imagen como criterio de búsqueda).

Para completar esta evaluación, la Figura 81 muestra los resultados para una imagen de un cuadro de Frida Kahlo, la cual tiene una estructura de composición basada en un triángulo. En esta composición, la región interior del triángulo está en ON y la exterior en OFF. En el mapa sin discriminación, la mayor prominencia está en la región del fondo, mientras que en la que sí discrimina, podemos ver la región central en ON separada de la del fondo en OFF. En relación con esta circunstancia, en la búsqueda sin discriminación, hay resultados donde la región interior del triángulo está en OFF y el fondo en ON (claramente en la 9, 10, 13 y 14). Sin embargo, cuando hay discriminación, todas las imágenes responden al patrón comentado: de figura en ON y fondo en OFF.

ON/OFF	Formato de la imagen similar	Características visuales diferentes	Estructura básica de la composición similar
Sin discriminación	94 %	88%	59%
Con discriminación	73%	64%	65%

Tabla 9 Porcentaje medio de las imágenes que se adaptan a los criterios.
Nota: las varianzas son 2.3%, 1.8% y 3.6%, respetivamente, sin discriminación de ON/OFF y 5.7%, 8.8% y 4.9%, respectivamente, con discriminación de ON/OFF.

Figura 80 Resultados de la búsqueda para una imagen de un cuadro de Degás.
Nota: los resultados de la búsqueda con el nombre del autor y la diferencia media de error, el criterio de búsqueda es la primera imagen de la parte superior izquierda.

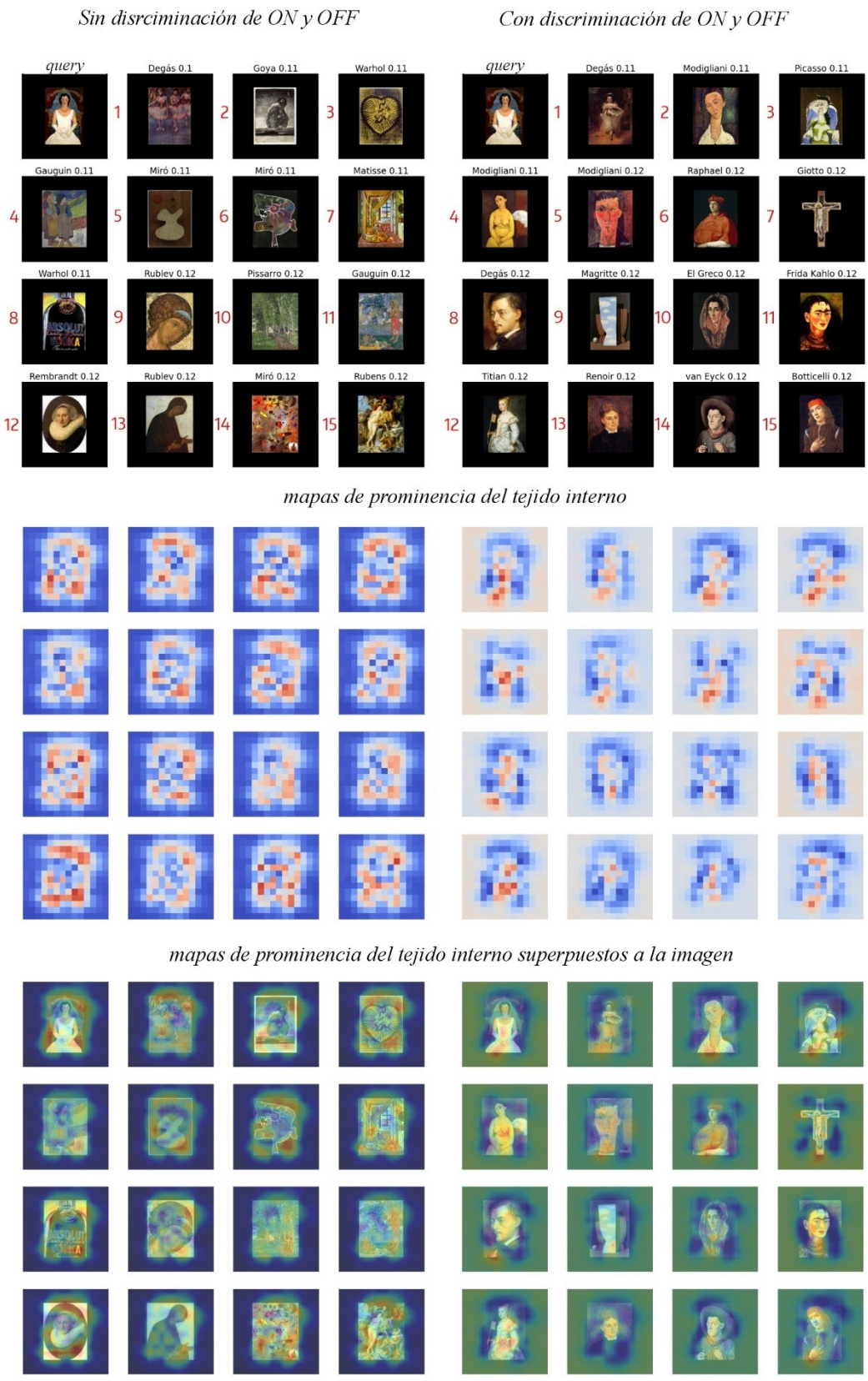


Figura 81 Resultados de la búsqueda para una imagen de un cuadro de Frida Kahlo.
Nota: los resultados de la búsqueda con el nombre del autor y la diferencia media de error, el criterio de búsqueda es la primera imagen de parte superior izquierda.

5.4.5 Conclusiones

La búsqueda de imágenes en bases de datos a partir de características visuales o del contenido es un área que tiene un importante desarrollo dentro de la visión artificial. El objetivo de este experimento ha sido el uso de los mapas de prominencia del tejido interno de las composiciones para evaluar su aplicación. Para la comparación entre el criterio de búsqueda y el *dataset*, se ha utilizado el RSME por ser el sistema más simple de comparación entre imágenes. Para evaluar los resultados de la búsqueda, se han utilizado una serie de criterios heurísticos para comprobar la similitud del formato de las imágenes obtenidas (vertical/horizontal y sí mantiene las proporciones), la diferencia de las características visuales presentes en la imagen y la similitud de la estructura básica de la composición. A partir de una selección de cincuenta imágenes al azar, se ha comprobado estos requerimientos obteniendo un porcentaje final medio para la búsqueda con mapas sin y con discriminación de ON y OFF.

A partir de los datos estadísticos, se ha podido concluir que la búsqueda con discriminación de ON y OFF localiza menos imágenes que mantengan el mismo formato, con menos diferencia en las características visuales y mayor similitud de la estructura básica de la composición. Una inspección de los mapas en dos casos concretos, uno más complejo en su composición con una diagonal y otro más sencillo con una estructura triangular de figura y fondo muy delimitados, ha demostrado que los mapas muy similares mantienen formatos y estructuras básicas de la composición semejantes, aunque sus características visuales sean distintas.

6 Conclusiones

Una de las principales claves conceptuales de esta tesis ha sido tratar a la imagen como un objeto visual. Cuando se compone, se utilizan los elementos visuales, como puntos, líneas, contornos, formas, colores o texturas, con una finalidad concreta. Después, la imagen es percibida y, a través de la extracción de características, la composición es reconstruida para comprender esa finalidad original. Además, como objeto visual, la imagen es condicionada por la realidad visual que se percibe. Para una rana, el movimiento es su realidad visual. Los árboles, el cielo azul, o las flores amarillas no existen, salvo que se muevan. Obviamente, una imagen en color de un amanecer no tendría sentido para la rana. Por esta razón, la composición de la imagen depende de la realidad visual y de cómo se percibe.

En este contexto, se introdujo la idea de una estructura subyacente a la composición que representara las tensiones y que se denomina tejido interno. La idea de tener una representación por debajo de los elementos visuales y sin necesidad de extraer sus características, fue una de las principales motivaciones de esta tesis. Conseguir esa representación se planteó como el principal objetivo.

Por otro lado, el mayor problema de la composición en las imágenes es el propio sistema de percepción. La teoría de la Gestalt demostró con sus experimentos que existían principios y leyes en la percepción que condicionan cómo se deberían de componer las imágenes, pero también las limitaciones. Si dibujamos un círculo sin cerrar, lo percibimos cerrado, aunque sin cerrarlo no sea un círculo. Otro aspecto que la psicología del arte ha estudiado, es la existencia de una estructura de fuerzas en el espacio visual de la imagen, donde destaca la atracción de la región central y, a su vez, de las regiones externas. Pero no sólo hay dos fuerzas de atracción en el espacio, sino que hay regiones donde existe una mayor tensión que en otras. Es decir, que además de situar los elementos visuales en el espacio vacío, es importante decidir dónde situarlos. Todas estas cuestiones obligaban a un análisis del sistema de percepción, ya que todos estos condicionantes dependen de su funcionamiento y no de la imagen por sí misma.

La conexión entre la estructura subyacente y la composición es uno de los principales problemas analizados. La cuestión principal fue localizar qué área del sistema de percepción visual podría representar esta estructura subyacente. La modularidad de la corteza visual y la dificultad para localizar un área concreta que tuviera esa representación, dirigió los esfuerzos al NGL. Una cuestión relevante fue analizar si la estructura del NGL podía representar al tejido interno e intervenir en la percepción de la composición a través de una conexión directa con el área V4. Con esta finalidad, se diseñó un experimento con redes de neuronas artificiales convolucionales (bioinspiradas en la corteza visual) para conectar las primeras capas, que se relacionan con la retina y el NGL, con las últimas capas, que se relacionan con el área V4. Los resultados demostraron que los modelos con esta conexión mejoraban tanto su precisión en tareas de

clasificación como su capacidad para extraer las características visuales vinculadas con la selectividad al color. Esta investigación se publicó en (Sanchez-Cesteros y otros, 2023).

Para poder obtener una representación del tejido interno, el planteamiento fue construir un modelo de visión artificial con una metodología bioinspirada. Para adaptar mejor esta metodología al objetivo, se incluyó el enfoque de la psicología del arte. En este sentido, el modelo, denominado Inner Fabric, es un ejemplo de transversalidad que combina neurociencia, psicología del arte y visión artificial. La investigación se centró en la modelización de los procesos necesarios con el fin de obtener una representación del tejido interno, a través de las relaciones entre las regiones de la imagen y sus pesos visuales. La complejidad principal estuvo en desligar la composición de las operaciones de extracción de características visuales, que suceden en un nivel consciente, y descender a un nivel más básico donde sólo existen tensiones. Esto, por otro lado, tenía un problema mayor, ya que en visión artificial no es común trabajar por debajo de la extracción de las características visuales.

Un primer problema fue transformar el espacio de píxeles a una representación en mapas retinotópicos con la excentricidad y la anisotropía entre sus regiones. Con este tipo de representación, se implementaba el marco estructural de la composición, donde hay un foco de atracción en la región central y otro en la externa, además, de una preponderancia de la región inferior izquierda. Esto conducía a una arquitectura funcional, es decir, a que la propia arquitectura fuera quien facilitara, por ejemplo, que existiera una mayor atracción de la región central por la excentricidad.

Un segundo problema fue cómo representar los píxeles en RGB del espacio visual en los mapas retinotópicos, donde los datos se estructuran en vías paralelas y segregadas. Esto obligó a analizar los sistemas de color, y a proponer uno nuevo, OCC, que convertía los píxeles de RGB a un sistema de colores opuestos y complementarios con un nivel de actividad e inactividad neuronal. Esta relación facilitó posicionar cada color en una escala jerárquica y obtener el peso visual. En un análisis posterior del sistema OCC, se comprobó que había una relación entre esta escala con una conversión de color a escala de grises, lo que llevó a patentar el sistema y realizar una investigación enfocada en esta cuestión. En esta investigación, se amplió el sistema con la inclusión de las categorías de cálido y frio, consiguiendo un conversor que pasaba con éxito, y mejor que otros conversores, la prueba de Ishihara, utilizada para detectar en el ser humano los problemas en la discriminación de colores.

Un tercer problema fue cómo construir el modelo para que escaneara la imagen a partir de un sistema atencional basado en el movimiento de ojos, el cual debía simular los procesos inconscientes relacionados con la percepción horizontal del tejido interno. Recorrer la imagen, determinar los pesos visuales y la relación entre ellos, conectaba el escaneo de la imagen con los procesos de la agudización y de la nivelación planteados desde la psicología del arte.

En este tesis, el alcance eran las imágenes artísticas y con criterios estéticos, aunque Inner Fabric se podría ajustar para cualquier tipo de imagen que perciba un ser humano. En la evaluación con pinturas conocidas, se comprobó que las estructuras de las com-

posiciones estaban representadas en los mapas del tejido interno. Esto demuestra que es posible obtener información sobre la composición sin tener que extraer las características visuales.

Como resumen, las contribuciones más relevantes a nivel de computación son:

- La representación neuronal de la composición de la imagen. Los mapas retinotópicos, por su excentricidad y anisotropía, facilitan los dos focos de atracción del marco estructural de Arnheim (central y externo) y la preponderancia de la región inferior izquierda.
- El cálculo del peso visual a partir de la actividad neuronal.
- La posibilidad de procesar los pesos visuales tanto en la vía ON como en la OFF dentro de la misma composición, lo cual permite detectar el contraste positivo-negativo en una vía y en la otra.
- La relación entre el escaneo de la imagen basado en el movimiento de ojos y los procesos de agudización y nivelación.
- La detección de regiones prominentes en la imagen, tanto por su nivel de actividad como de inactividad, diferenciando su peso visual (valor local) de su nivelación con el resto de la imagen (valor global).
- La representación del tejido interno en un mapa de prominencia sin y con discriminación ON y OFF, y el trazado del proceso de escaneo.
- El mapa de prominencia del tejido interno, al ser una representación por debajo de las características visuales, permite el procesamiento de la composición con independencia de los estilos, formatos y temáticas.
- La aplicación del mapa de prominencia del tejido interno tanto en el proceso creativo como en el análisis de imágenes ya creadas.

Con la finalidad de evaluar el mapa del tejido interno, se crearon dos casos de uso: la clasificación del tipo de composición y la búsqueda de imágenes usando el mapa de prominencia del tejido interno como criterio. En el caso de la clasificación, se solucionó un problema complejo de resolver a partir de la extracción de características visuales, ya que existe bastante subjetividad en el análisis del tipo de composición. El uso del tejido interno simplificó esta tarea, ya que utiliza un solo mapa de prominencia y no un conjunto de mapas de características visuales. Los resultados de los experimentos realizados demostraron que la clasificación se ajustaba al tipo de composición. En el caso de la búsqueda, el mapa del tejido interno facilitó la recuperación de imágenes con composiciones parecidas con independencia de los elementos visuales utilizados, lo que es algo muy complejo de obtener desde una imagen y su composición a partir de las características visuales.

Los mapas del tejido interno, al representar las tensiones y al no estar relacionados con el estilo, las formas o el contenido, pueden ser aplicados en diferentes escenarios. Des-

de la detección de áreas que puedan ser relevantes para otros procesos, lo cual optimiza el procesamiento, hasta su aplicación directa en tareas de visión, como hemos visto en los casos de uso. Y no sólo con imágenes que hayan sido creadas con criterios compositivos, sino también con cualquier imagen que necesite ser interpretada por seres humanos, ya que Inner Fabric está basado en el sistema de percepción.

Las imágenes viven a pesar de nosotros, una pintura del paleolítico ha estado durante milenios en lo más profundo de una cueva. Sin embargo, no cobra vida hasta que alguien baja hasta allí, enciende su antorcha y observa la pintura. La composición que utilizó el lejano pintor está presente ante los ojos del espectador, y todas esas fuerzas detrás de los elementos visuales están ahí. La realidad visual de ambos seres humanos es la misma, aunque les separen miles de años, pero también lo es la necesidad de situar todos esos elementos visuales según un orden que depende de un objetivo. Es probable que, si ambos estuvieran en la misma habitación, no se entendieran a través de las palabras, pero lo que sí que es seguro, es que la composición de la pintura y su estructura subyacente sería comprensible.

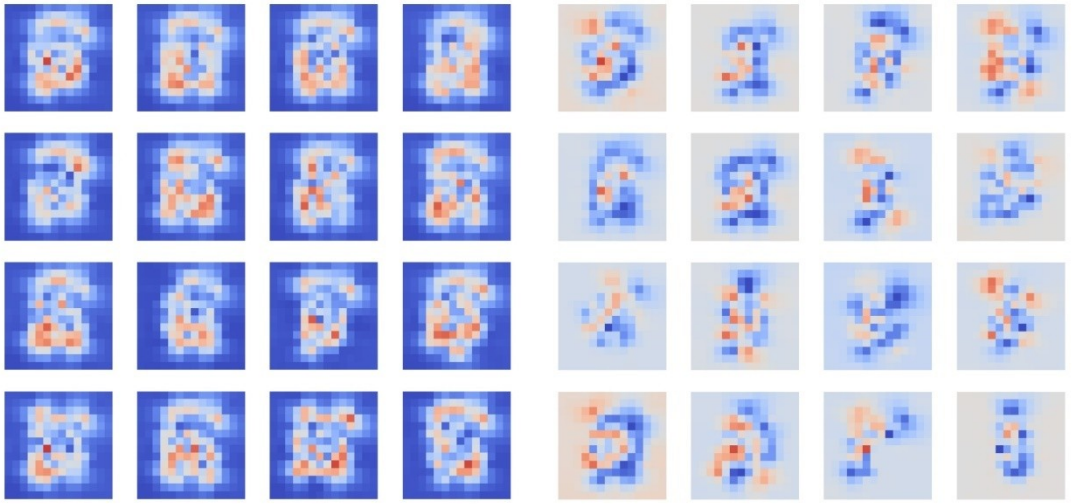
Anexos

Anexo 1. Resultados de la búsqueda de imágenes por el tejido interno

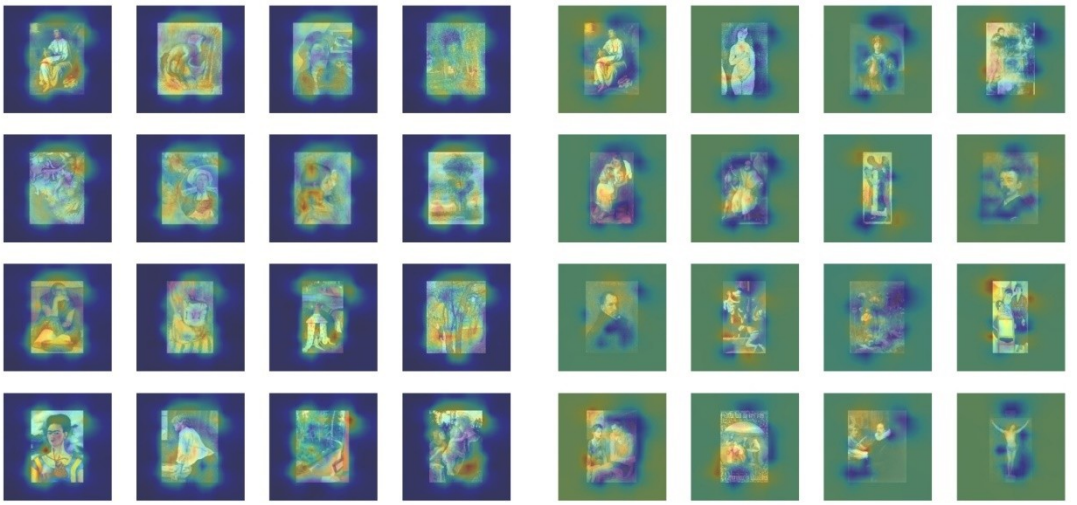
Sin disrciminación de ON y OFF

Con discriminación de ON y OFF

mapas de prominencia del tejido interno



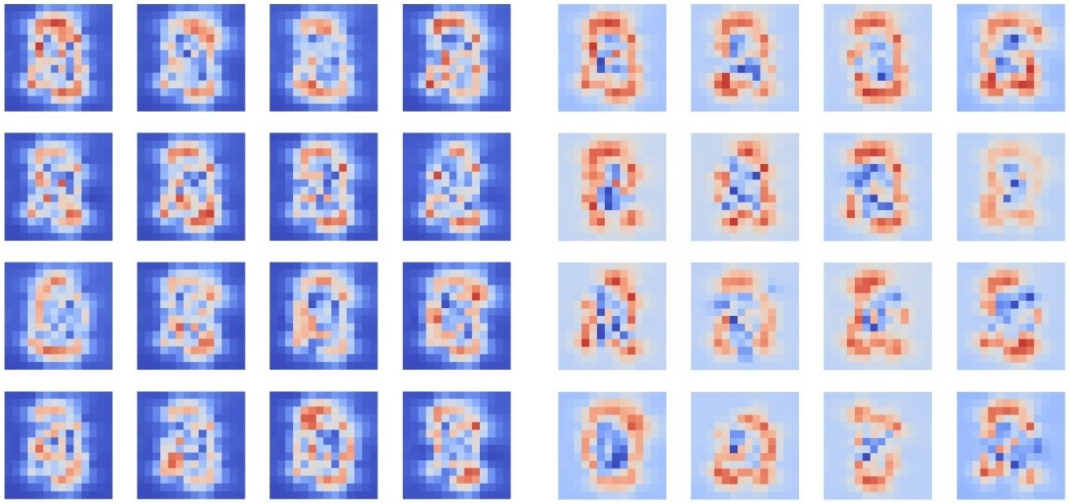
mapas de prominencia del tejido interno superpuestos a la imagen



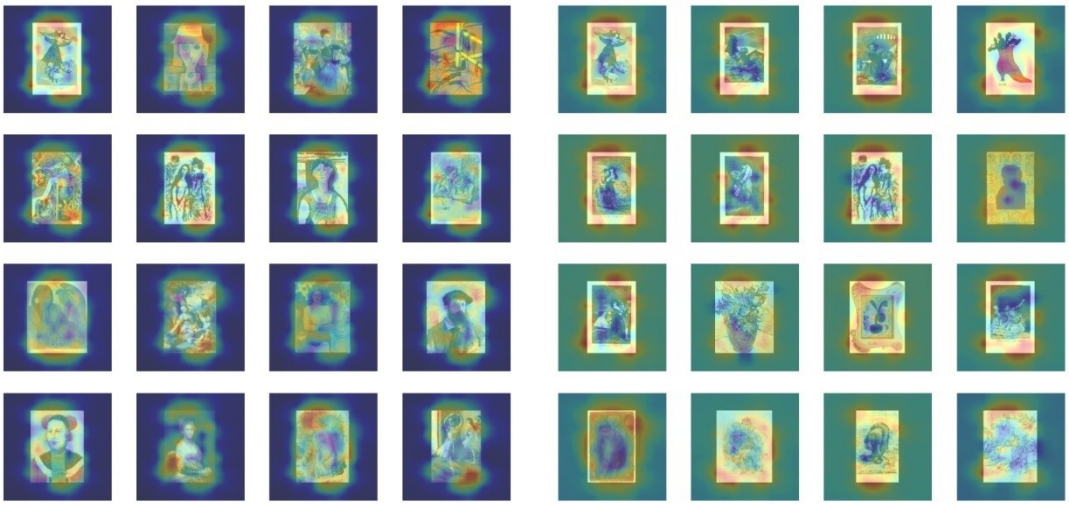
Sin disrciminación de ON y OFF

Con discriminación de ON y OFF

mapas de prominencia del tejido interno



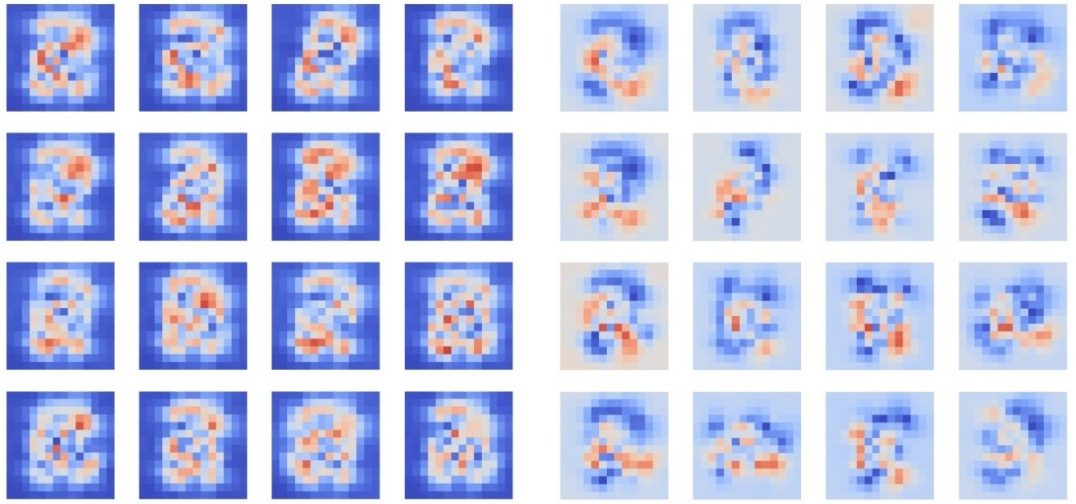
mapas de prominencia del tejido interno superpuestos a la imagen



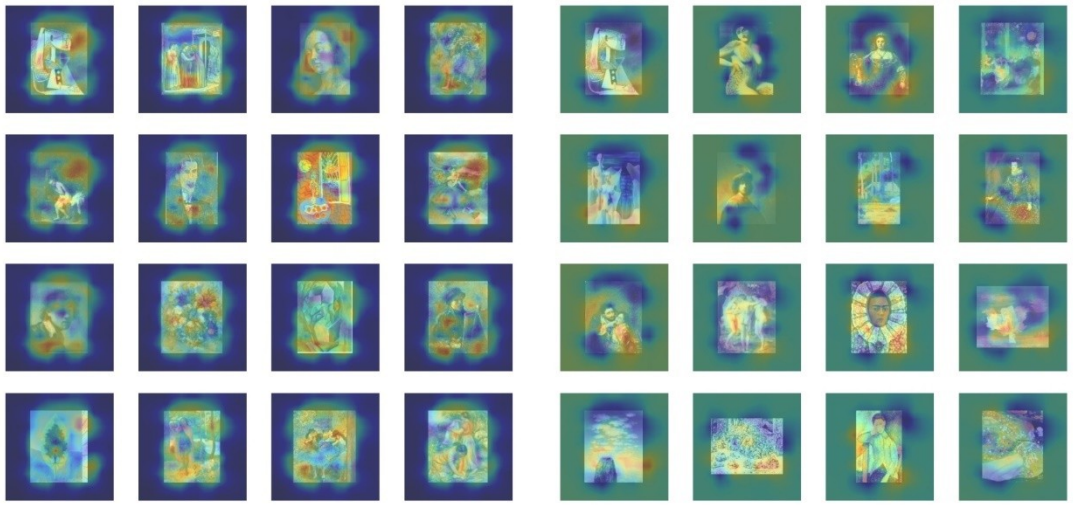
Sin disrciminación de ON y OFF

Con discriminación de ON y OFF

mapas de prominencia del tejido interno

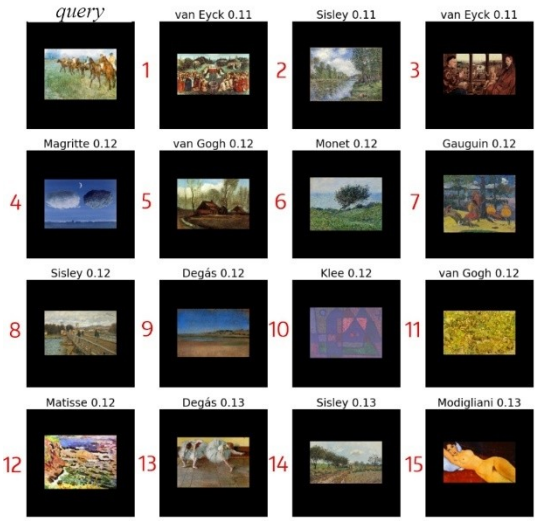


mapas de prominencia del tejido interno superpuestos a la imagen



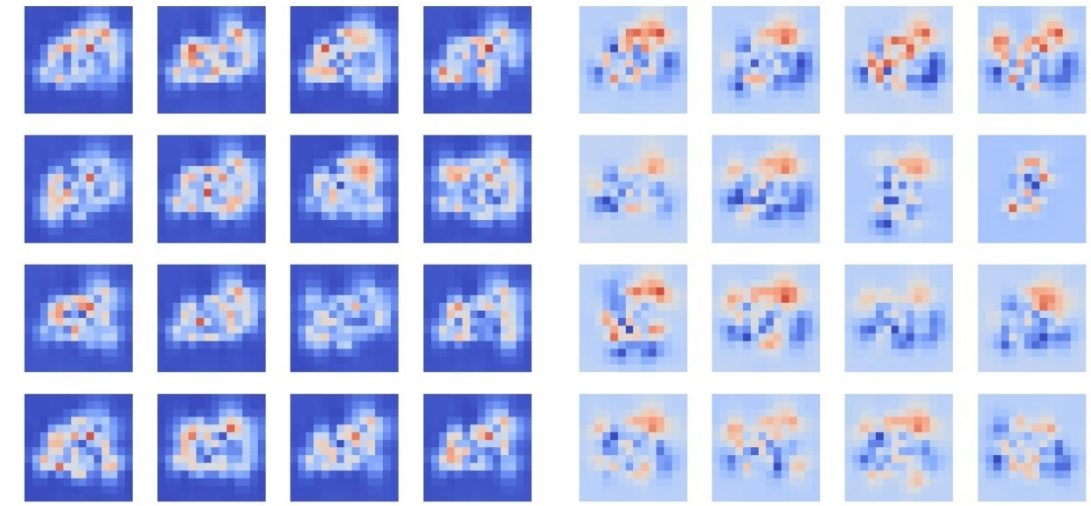

Sin disrciminación de ON y OFF
Con discriminación de ON y OFF
query
Bosch 0,06
Bosch 0,08
Picasso 0,09
query
Bosch 0,09
Magritte 0,09
Modigliani 0,10
1
2
3
1
2
3
Van Eyck 0,09
Klimt 0,11
Bosch 0,11
Renoir 0,11
Gauguin 0,10
Vrubel 0,10
Matisse 0,10
Rivera 0,10
4
5
6
7
4
5
6
7
Van Eyck 0,11
Degas 0,11
Giotto 0,11
Bosch 0,12
Bosch 0,11
Van Gogh 0,11
Degas 0,11
Degas 0,11
8
9
10
11
8
9
10
11
El Greco 0,12
Rublev 0,12
Rembrandt 0,12
Bosch 0,12
Picasso 0,11
Modigliani 0,11
Picasso 0,11
Degas 0,11
12
13
14
15
12
13
14
15
mapas de prominencia del tejido interno
mapas de prominencia del tejido interno superpuestos a la imagen

Sin disrciminación de ON y OFF

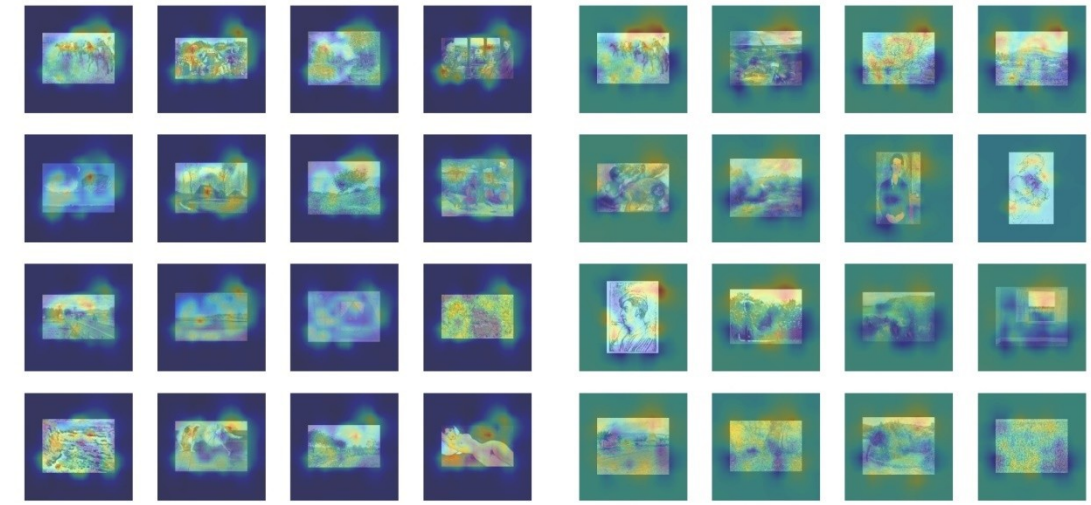


Con discriminación de ON y OFF

mapas de prominencia del tejido interno



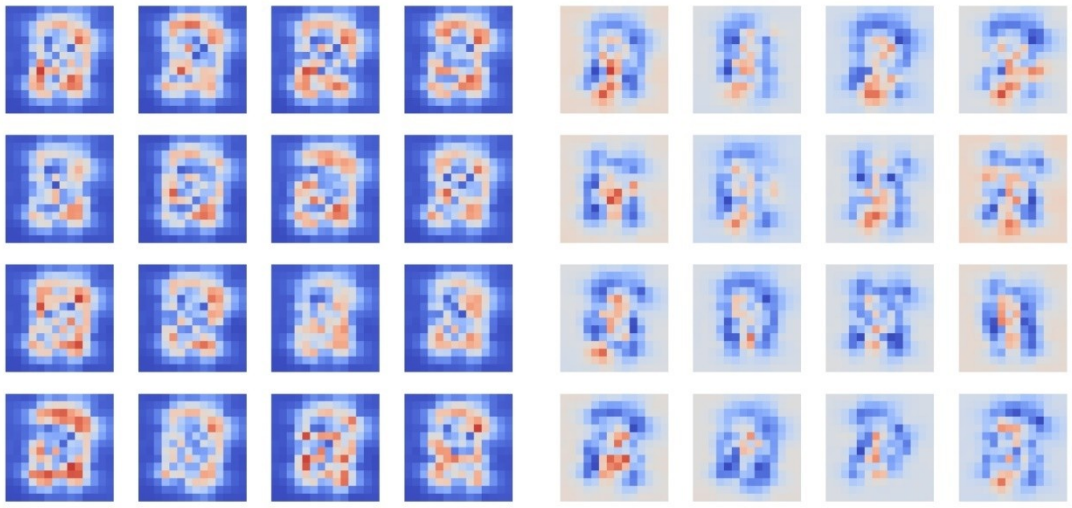
mapas de prominencia del tejido interno superpuestos a la imagen



Sin disrciminación de ON y OFF

Con discriminación de ON y OFF

mapas de prominencia del tejido interno



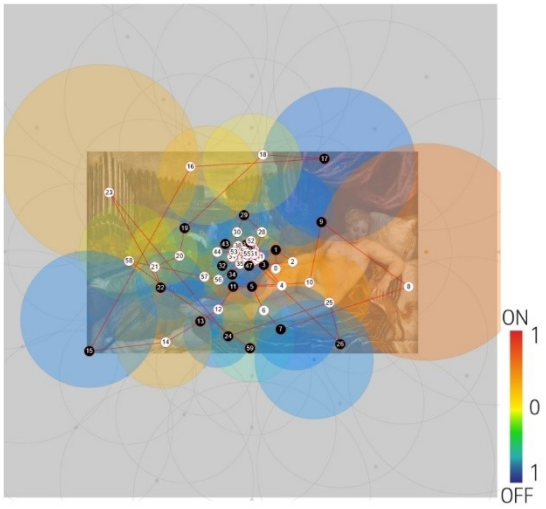
mapas de prominencia del tejido interno superpuestos a la imagen

Anexo 2. Mapas de prominencia del tejido interno de pinturas del Museo de Prado

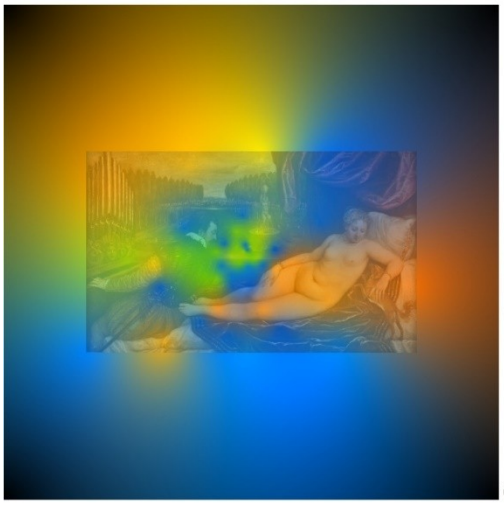
Las pinturas pertenecen a la Colección del Museo del Prado. Las imágenes utilizadas por Inner Fabric para obtener el mapa de prominencia del tejido internas son originales del catálogo digital del Museo (https://www.museodelprado.es/coleccion/obras-de-arte) y tienen derecho de uso sólo para fines académicos. Su reproducción fuera de este documento no está permitida.

imagen en el espacio visual

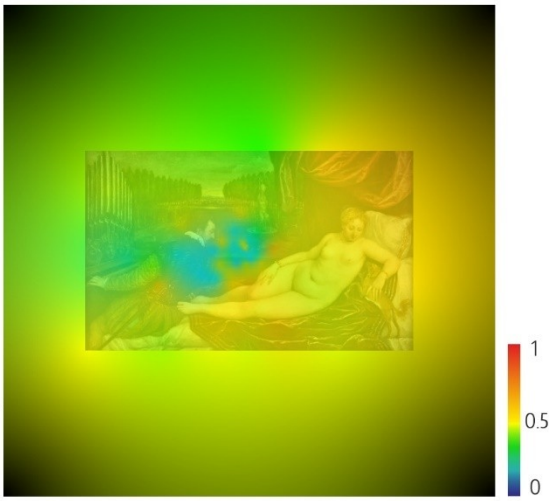
escaneo



mapa de prominencias del tejido interno con discriminación ON y OFF

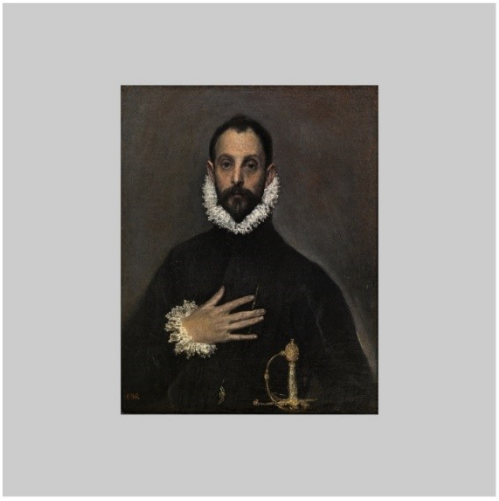


mapa de prominencias del tejido interno sin discriminación ON y OFF

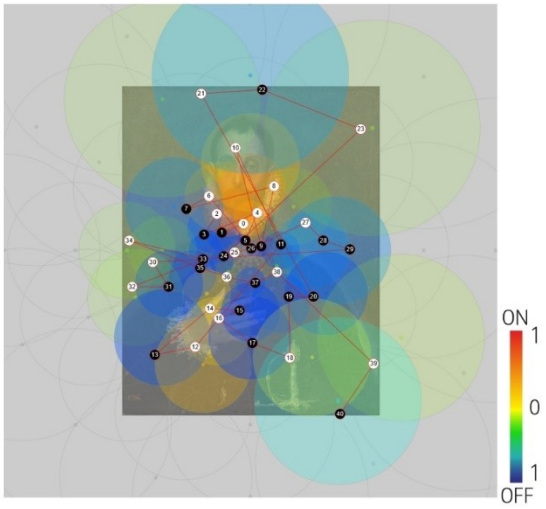


«Venus recreándose en la Música» por Veccellio di Gregorio Tiziano, (1550)

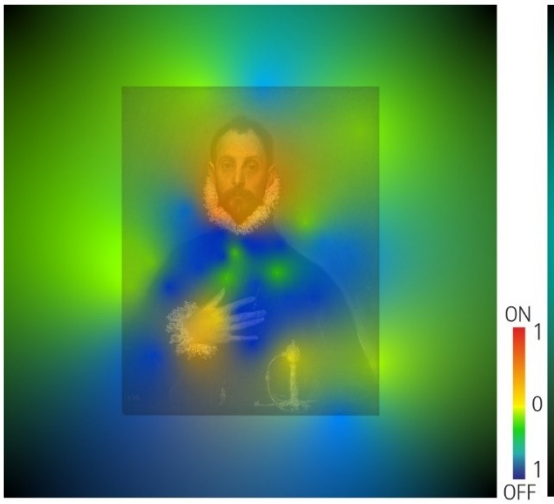
imagen en el espacio visual



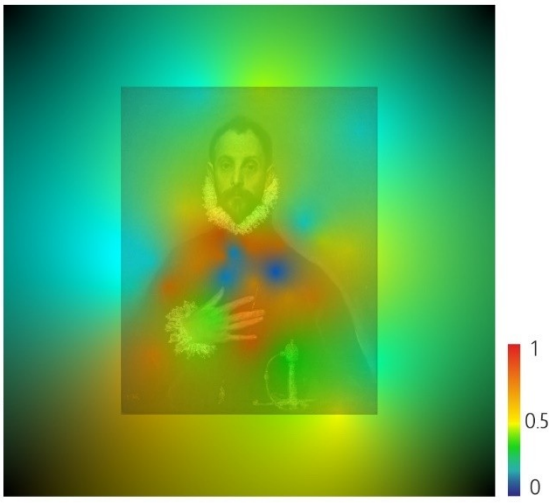
escaneo



mapa de prominencias del tejido interno con discriminación ON y OFF



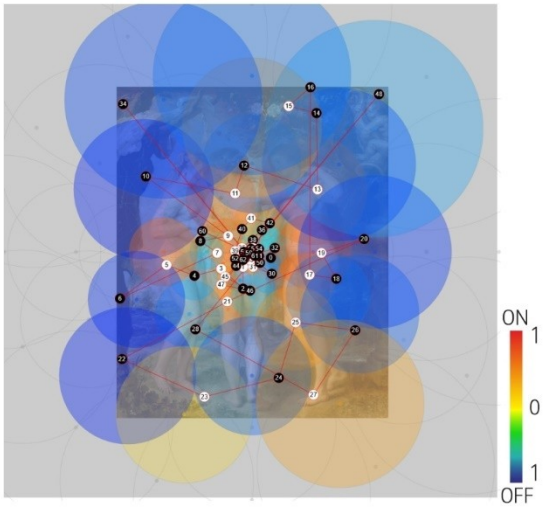
mapa de prominencias del tejido interno sin discriminación ON y OFF



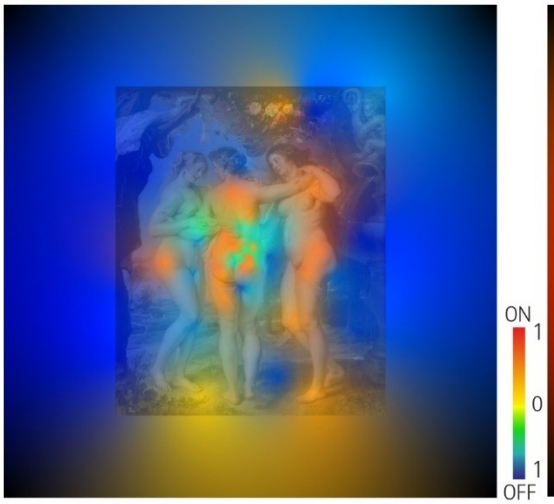
«El caballero de la mano en el pecho» por El Greco, (1580)

imagen en el espacio visual

escaneo



mapa de prominencias del tejido interno con discriminación ON y OFF

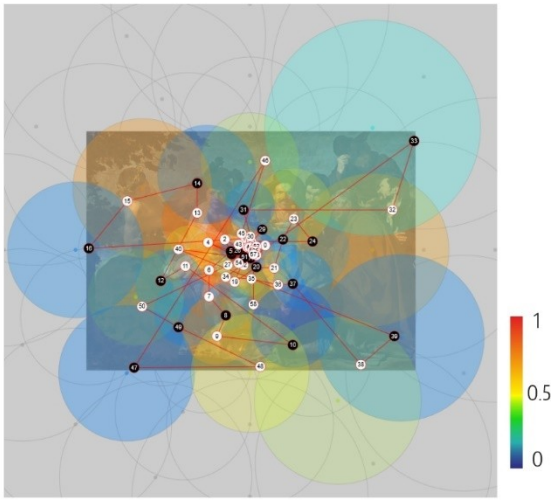


mapa de prominencias del tejido interno sin discriminación ON y OFF

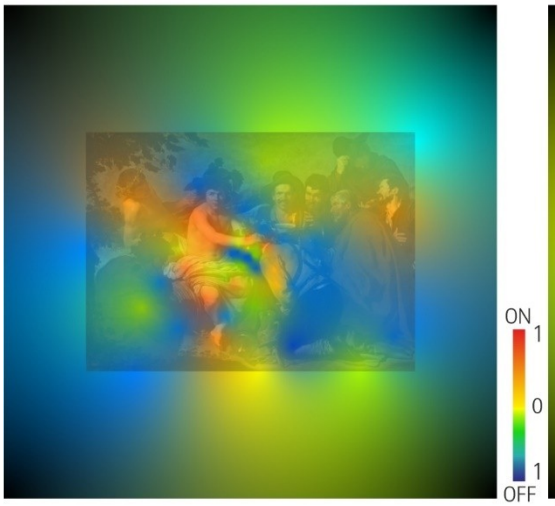
«Las tres Gracias» por Pedro Pablo Rubens, (1630-1635)

imagen en el espacio visual

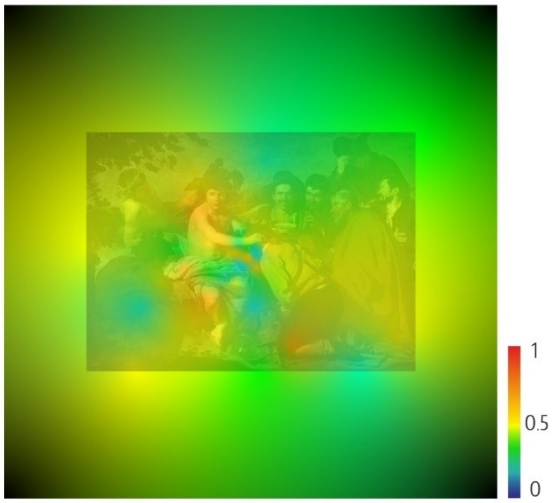
escaneo



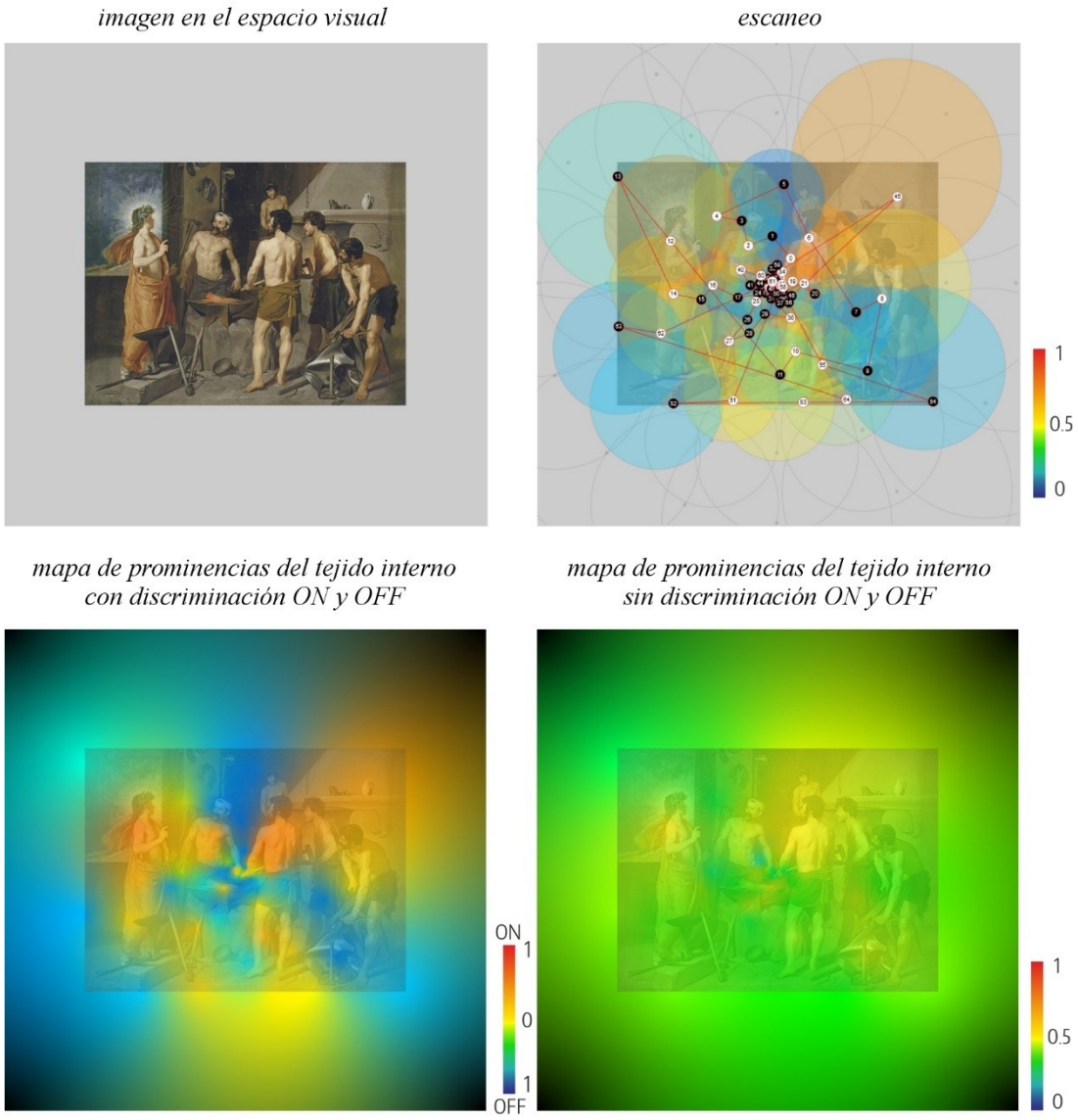
mapa de prominencias del tejido interno con discriminación ON y OFF



mapa de prominencias del tejido interno sin discriminación ON y OFF



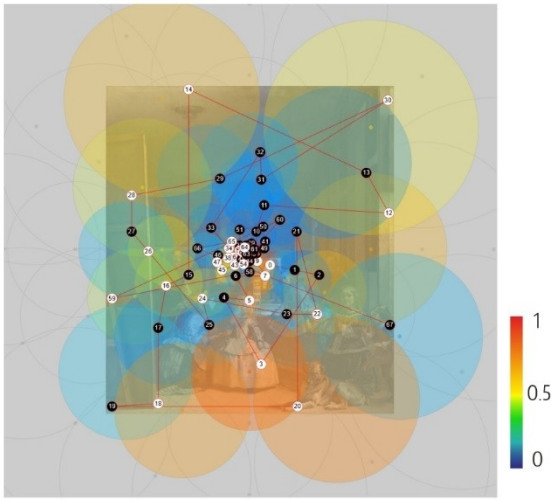
«Los borrachos», o «El triunfo de Baco» por Diego Rodríguez de Silva y Velázquez, (1628 - 1629)



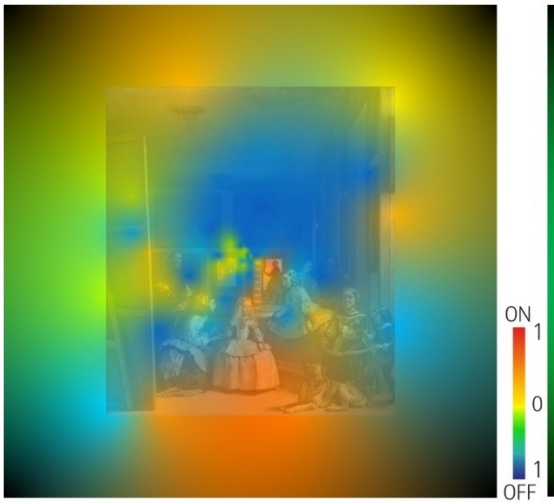
«La Fragua de Vulcano» por Diego Rodríguez de Silva y Velázquez, (1630)

imagen en el espacio visual

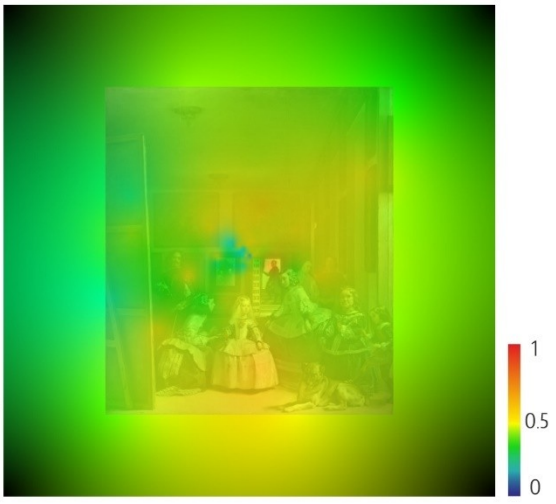
escaneo



mapa de prominencias del tejido interno con discriminación ON y OFF



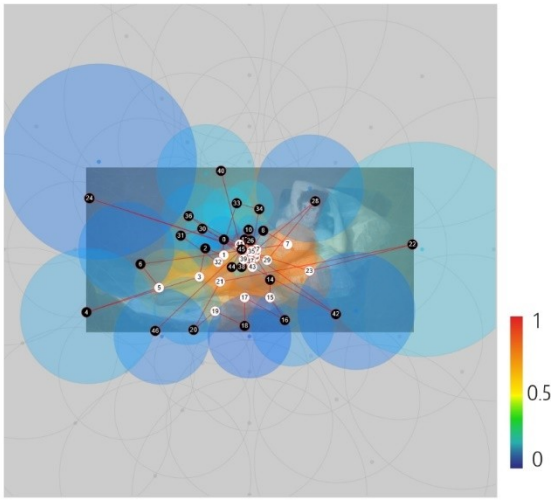
mapa de prominencias del tejido interno sin discriminación ON y OFF



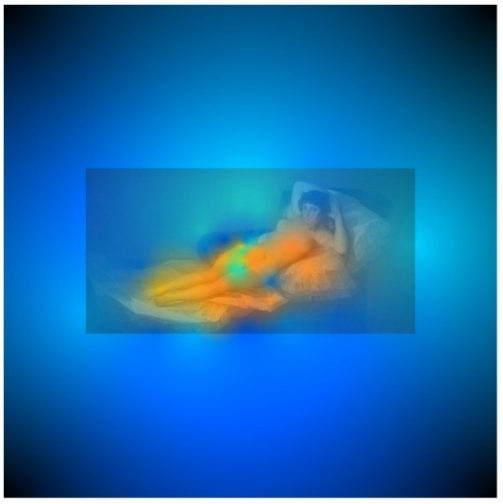
«Las meninas» por Diego Rodríguez de Silva y Velázquez, (1656)

imagen en el espacio visual

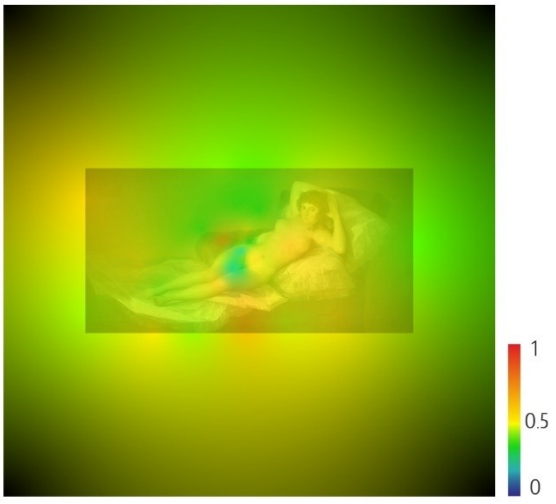
escaneo



mapa de prominencias del tejido interno con discriminación ON y OFF



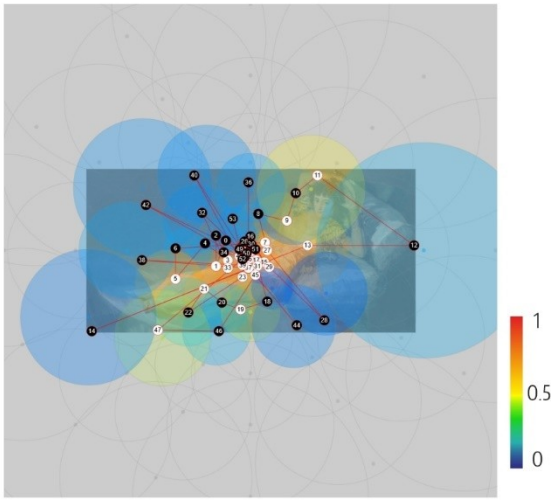
mapa de prominencias del tejido interno sin discriminación ON y OFF



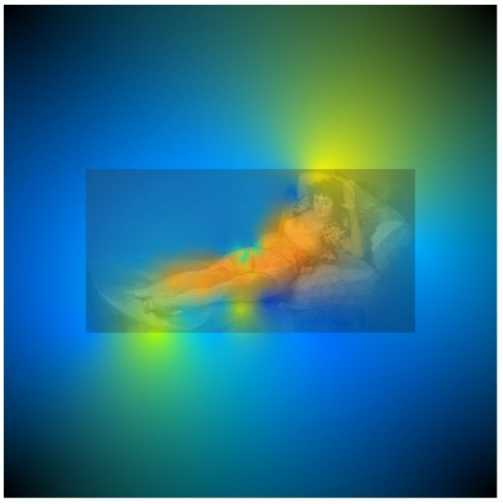
«La maja desnuda» por Francisco de Goya y Lucientes, (1785-1800)

imagen en el espacio visual

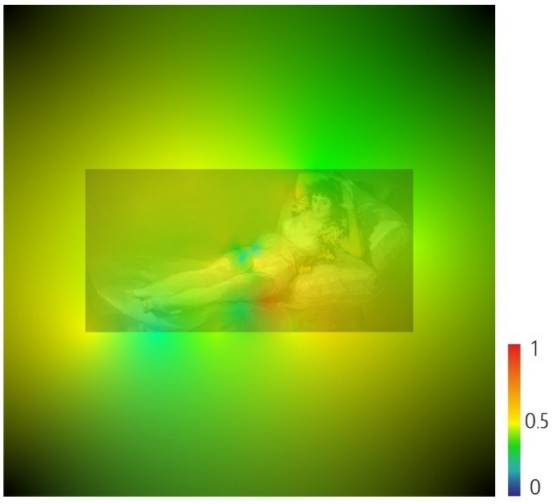
escaneo



mapa de prominencias del tejido interno con discriminación ON y OFF



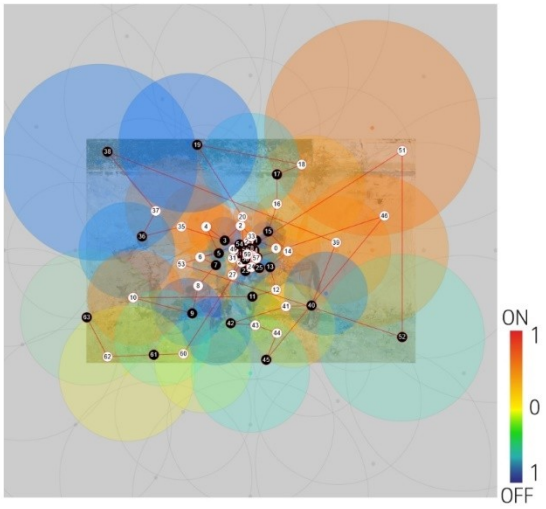
mapa de prominencias del tejido interno sin discriminación ON y OFF



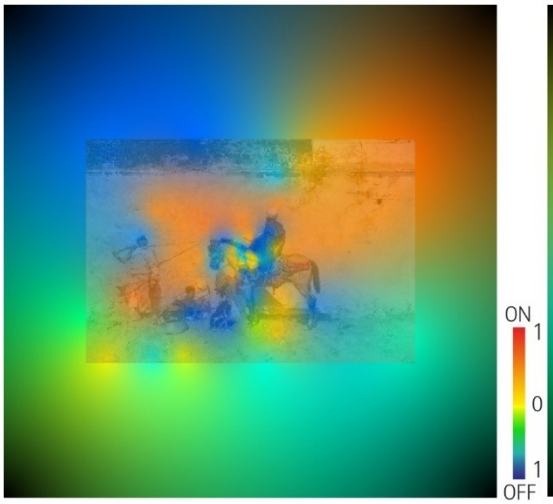
«La maja vestida» por Francisco de Goya y Lucientes, (18001807)

imagen en el espacio visual

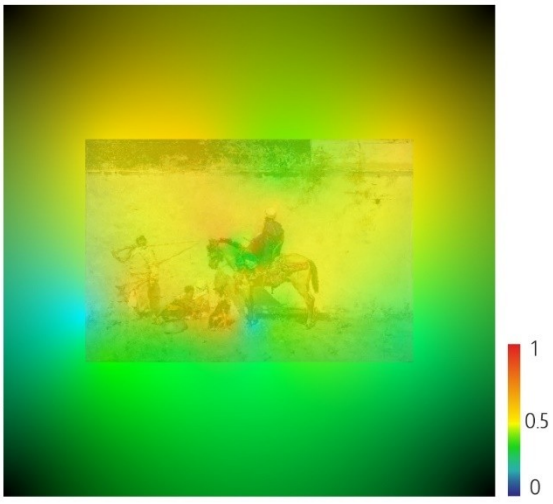
escaneo



mapa de prominencias del tejido interno con discriminación ON y OFF



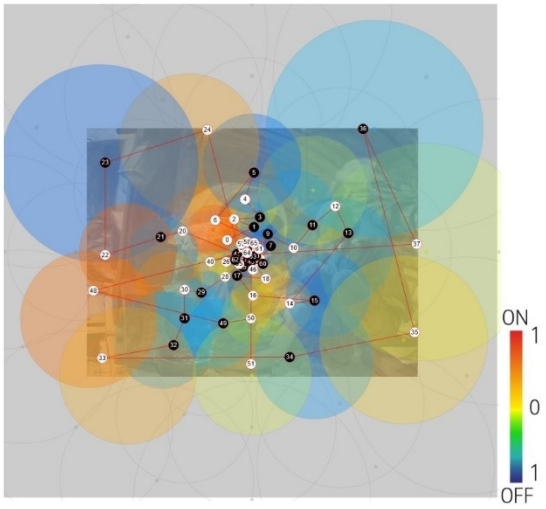
mapa de prominencias del tejido interno sin discriminación ON y OFF



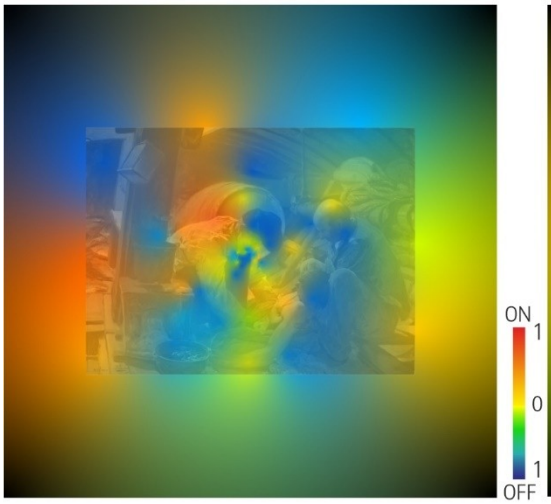
«Marroquíes» por Mariano Fortuny y Marsal, (1872-1874)

imagen en el espacio visual

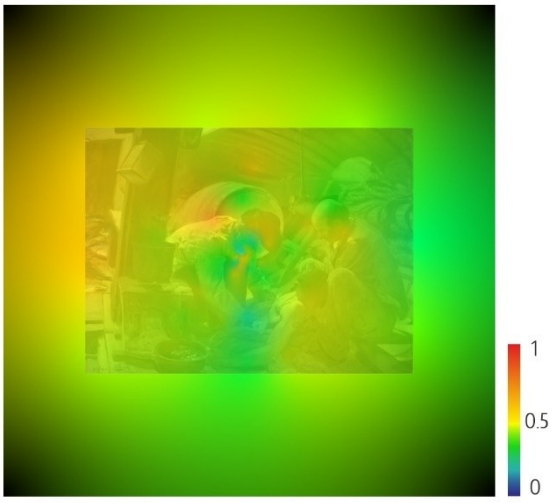
escaneo



mapa de prominencias del tejido interno con discriminación ON y OFF



mapa de prominencias del tejido interno sin discriminación ON y OFF



«¡Aún dicen que el pescado es caro!» por Joaquín Sorolla y Bastida, (1894)

Referencias

Agarwal, R., & Sarma, S. V. (2011). An analytical study of relay neuron's reliability: Dependence on input and model parameters. *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2426-2429.

Alitto, H. J., & Usrey, W. M. (2003). Corticothalamic feedback and sensory processing. *Current opinion in neurobiology*, 440-445.

Aloimonos, Y., & Shulman, D. (1990). *Integration of Visual Modules. An extension of the Marr Paradigm.* Academic Press, London.

Amunts, K., Armstrong, E., Malikovic, A., Homke, L., Mohlberg, H., Schleicher, A., & Zilles, K. (2007). Gender-specific left--right asymmetries in human visual cortex. *Journal of Neuroscience, 27*, 1356-1364.

Arnheim, R. (1956). *Art and visual perception: A psychology of the creative eye.* Univ of California Press.

Arnheim, R. (1969). *Visual thinking.* Univ of California Press.

Arnheim, R. (1983). *The power of the center: A study of composition in the visual arts.* Univ of California Press.

Arrigo, A., Calamuneri, A., Mormina, E., Gaeta, M., Quartarone, A., Marino, S., . . . Aragona, P. (2016). New insights in the optic radiations connectivity in the human brain. *Investigative Ophthalmology & Visual Science, 57*, 1–5.

Babenko, A., Slesarev, A., Chigorin, A., & Lempitsky, V. (2014). Neural codes for image retrieval. *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, (pp. 584–599).

Baingio, P. (2013). Foreword - On the Meaning of Visual Meanings. *Gestalt Theory, 34*, 237-258.

Bala, R., & Eschbach, R. (2004). Spatial color-to-grayscale transform preserving chrominance edge information. *Color and Imaging Conference, 1*, 82-86.

Bala, R., & Eschbach, R. (2008, 6). Color to grayscale conversion method and apparatus. Google Patents.

Baldrati, A., Bertini, M., Uricchio, T., & Del Bimbo, A. (2022). Effective conditioned and composed image retrieval combining clip-based features. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (pp. 21466–21474).

Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human factors, 34*, 513–533.

Berlin, B., & Kay, P. (1991). *Basic color terms: Their universality and evolution.*

Berlyne, D. E. (1973). Aesthetics and psychobiology.

Blackburn, M. R. (1993). A simple computational model of center-surround receptive fields in the retina. *NAVAL COMMAND CONTROL AND OCEAN SURVEILLANCE CENTER RDT AND E DIV SAN DIEGO CA.*

Boehm, G. (1994). *Die Wiederkehr der Bilder. Was istein Bild?* Munich: Munich: Fink.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence, 35*, 185-207.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of cognitive neuroscience, 9*, 27-38.

Bratkova, M., Boulos, S., & Shirley, P. (2009). oRGB: a practical opponent color space for computer graphics. *IEEE computer graphics and applications, 29*, 42-55.

Brea, J. (2006). Estética, historia del arte y estudios visuales. *Revista de estudios visuales, 3*.

Bredekamp, H. (2004). Dehmomente- Merkmale und Ansprüche des Iconic Turn. *Iconic Turn: Die Neue Macth der Bilder.*, 15-26.

Broadbent, D. E. (2013). *Perception and communication.* Elsevier.

Brown, C. M. (1984). Computer vision and natural constraints. *Science, 224*, 1299-1305.

Busswell, G. T. (1935). How people look at pictures: A study of the psychology of perception in art. Chicago: University of Chicago Press.

Calabrese, O. (1985). *Il linguaggio dell'arte.*

Calì, G. (2013). Gestalt Models for Data Decomposition and Functional Architecture in Visual Neuroscience. *Gestalt Theory, 35*, 227-264.

Chen, Q., Weidner, R., Weiss, P. H., Marshall, J. C., & Fink, G. R. (2012). Neural interaction between spatial domain and spatial reference frame in parietal–occipital junction. *Journal of cognitive neuroscience, 24*, 2223-2236.

Chen, W., Du, X., Yang, F., Beyer, L., Zhai, X., Lin, T.-Y., . . . others. (2021). A simple single-scale vision transformer for object localization and instance segmentation. *arXiv preprint arXiv:2112.09747*.

Clark, J. H. (1924). The Ishihara test for color blindness. *American Journal of Physiological Optics.*

Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience, 3*, 553-562.

Colby, C. L. (1998). Action-oriented spatial reference frames in cortex. *Neuron, 20*, 15-24.

Corbett, J. E. (2011). Visual performance fields: Frames of reference. *PloS one, 6*, 1-10.

Cox, D. D., & Dean, T. (2014). Neural Networks and Neuroscience-Inspired Computer Vision. *Current Biology, 24*, 921-929.

Crick, F. (1994). *The Astonishing Hypothesis, The Scientific Search for the Soul.*

Csurka, G., Dance, C., Fan, L., Willamowski, J., & Bray, C. (2004). Visual categorization with bags of keypoints. *Workshop on statistical learning in computer vision, ECCV*, *1*, pp. 1–2.

Dacey, D., Packer, O. S., Diller, L., Brainard, D., Peterson, B., & Lee, B. (2000). Center surround receptive field structure of cone bipolar cells in primate retina. *Vision research, 40*, 1801-1811.

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2006). Studying aesthetics in photographic images using a computational approach. *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part III 9*, (pp. 288–301).

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance.* Academic Press.

di Lenardo, I., Seguin, B. L., & Kaplan, F. (2016). *Visual patterns discovery in large databases of paintings.* Tech. rep.

Dondis, D. A. (1974). *A primer of visual literacy.* Mit Press.

Dorai, C., & Venkatesh, S. (2003). Bridging the semantic gap with computational media aesthetics. *IEEE multimedia, 10*, 15-17.

Draves, S. (2005). The electric sheep screen-saver: A case study in aesthetic evolution. *Workshops on applications of evolutionary computation*, (pp. 458–467).

Du, H., He, S., Sheng, B., Ma, L., & Lau, R. W. (2014). Saliency-guided color-to-gray conversion using region-based optimization. *IEEE Transactions on Image Processing, 24*, 434-443.

Ehrenzweig, A. (1967). *The hidden order of art: A study in the psychology of artistic imagination.* Univ of California Press.

Einevoll, G. T. (2003). Mathematical modelling in the early visual system: Why and how? *NATO SCIENCE SERIES SUB SERIES I LIFE AND BEHAVIOURAL SCIENCES*, 135-164.

Elkins, J. (2001). *The domain of images.* Cornell University Press.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological review, 112*, 777.

Fattori, P., & Pitzalis, C. (2009). The cortical visual area V6 in macaque and human brains. *Journal of Physiology-Paris, 103*, 88-97.

Fergus, R., Fei-Fei, L., Perona, P., & Zisserman, A. (2005). Learning object categories from Google's image search. *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, *2*, pp. 1816–1823.

Ferrera, V. P., Nealey, T. A., & Maunsell, J. H. (1994). Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN pathways. *Journal of Neuroscience, 14*, 2080-2088.

Fishman, R. S. (1997). Gordon Holmes, the cortical retina, and the wounds of war. *The seventh Charles B. Snyder Lecture Documenta Ophthalmologica, 93*, 8-28.

Frankel, C., Swain, M. J., & Athitsos, V. (1996). *Webseer: An image search engine for the world wide web.* Tech. rep., Technical Report 96-14, University of Chicago, Computer Science Department.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics, 36*, 193-202.

Galanter, P. (2012). Computational aesthetic evaluation: past and future. *Computers and creativity*, 255–293.

Galati, G., Pelle, G., Berthoz, A., & Committeri, G. (2010). Multiple reference frames used by the human brain for spatial perception and memory. *Experimental brain research, 206*, 109-120.

Ghosh, K., & Pal, S. K. (2010). Some insights into brightness perception of images in the light of a new computational model of figure--ground segregation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 40*, 758-766.

Goethe, J. W. (1840). *Theory of colours.* MIT press.

Gombrich, E. (1959). *Illusion and art. A study in the psychology of pictorial representation.* Oxford.

Gong, Y., Cosma, G., & Finke, A. (2023). Neural-based Cross-modal Search and Retrieval of Artwork. *arXiv preprint arXiv:2307.14244*.

Gooch, A. A., Olsen, S. C., Tumblin, J., & Gooch, B. (2005). Color2gray: salience-preserving color removal. *ACM Transactions on Graphics (TOG), 24*, 634-639.

Gravesen, J. (2015). The metric of colour space. *Graphical Models, 82*, 77-86.

Grossberg, S. (1990). Neural Facades: Visual Representations of Static and Moving Form-And-Color-And-Depth. *Mind \& Language, 5*, 411-456.

Grossberg, S. (2003). How does the cerebral cortex work? Development, learning, attention, and 3-D vision by laminar circuits of visual cortex. *Behavioral and Cognitive Neuroscience Reviews, 2*, 47-76.

Hadjidimitrakis, K., Breveglieri, R., Bosco, A., & Fattori, P. (2012). Three-dimensional eye position signals shape both peripersonal space and arm movement activity in the medial posterior parietal cortex. *Frontiers in integrative neuroscience, 6*, 37.

Hare, J. S., Sinclair, P. A., Lewis, P. H., Martinez, K., Enser, P. G., & Sandom, C. J. (2006). Bridging the semantic gap in multimedia information retrieval: Top-down and bottom-up approaches.

Hegel, G. W. (1845). *Vorlesungen über die Ästhetik, Werke.* Ge- sammelte Ausgabe IM 18 Bde.

Helmholtz, H. V. (1852). LXXXI. on the theory of compound colours. *Philosophical Magazine A, 4*, 519-534.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. En *Eye movements* (págs. 537--III). Elsevier.

Hering, E. (1885). *Ueber individuelle Verschiedenheiten des Farbensinnes.*

Holmes, G. (1945). Ferrier Lecture: the organization of the visual cortex in man. *Proceedings of the Royal Society of London. Series B-Biological Sciences, 132*, 348-261.

Horton, J. C., & Hoyt, W. F. (1991). The representation of the visual field in Human Striate Cortex. *Archives of Ophthalmology, 109*.

Hubel, D. (1988). *Eye, Brain and Vision. Scientific American Libarary.* Scientific American Libarary. New York, USA.

Hubel, D. H. (1995). *Eye, Brain and Vision.* New York: Scientific American Library/Scientific American.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of neurophysiology, 28*, 229–289.

Icaro. (2023, 9 23). *Best Artworks of All Time*. Best Artworks of All Time: https://www.kaggle.com/datasets/ikarus777/best-artworks-of-all-time

Itten, J. (1992). *El arte del color.* Limusa.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis \& Machine Intelligence*, 1254-1259.

Jain, R. C., & Binford, T. O. (1991). Ignorance, myopia, and naivete in computer vision systems.

Javal, É. (1878). *Essai sur la physiologie de la lecture* (Vol. 80). Annales d'Ocilistique.

Jeffries, A. M., Killian, N. J., & Pezaris, J. S. (2014). Mapping the primate lateral geniculate nucleus: a review of experiments and methods. *Journal of Physiology-Paris, 108*, 3-10.

Jégou, H., Douze, M., & Schmid, C. (2010). Improving bag-of-features for large scale image search. *International journal of computer vision, 87*, 316–336.

Jing, Y., & Baluja, S. (2008). Pagerank for product image search. *Proceedings of the 17th international conference on World Wide Web*, (pp. 307–316).

Joshi, D., Datta, R., Fedorovskaya, E., Luong, Q.-T., Wang, J. Z., Li, J., & Luo, J. (2011). Aesthetics and emotions in images. *IEEE Signal Processing Magazine, 28*, 94–115.

Kandinsky, V. (1969). *Punto y línea frente al plano.* Nueva Visión.

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. *Spatial cognition*, (pp. 1-17).

Klee, P. (1961). *Paul Klee: the thinking eye: The notebooks of Paul Klee* (Vol. 15). G. Wittenborn.

Koch, C., & Ullman, S. (1987). Shifts in selective visual attention: towards the underlying neural circuitry. *In Matters of intelligence*, 115-141.

Kofka, K. (1955). *Principles of Gestalt Psychology.* Routledge and Kegan Paul Ltd.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, (pp. 1097-1105).

Kueppers, H. (1982). *The basic law of color theory.* Barrons Educational Series Incorporated.

Kuk, J. G., Ahn, J. H., & Cho, N. I. (2010). A color to grayscale conversion considering local and global contrast. *Asian Conference on Computer Vision*, (pp. 513-524).

Kuk, J. G., Ahn, J. H., & Cho, N. I. (2011). A color to grayscale conversion considering local and global contrast. *Asian Conference on Computer Vision*, 513-524.

Land, E. H. (1977). The retinex theory of color vision. *Scientific America*, 2-17.

LDMTWO. (2023, 9 23). *Midjourney 2022 - 250k [CSV]*. https://www.kaggle.com/datasets/ldmtwo/midjourney-250k-csv

Le Meur, O., Le Callet, P., Barba, D., & Thoreau, D. (2006). A coherent computational approach to model bottom-up visual attention. *IEEE transactions on pattern analysis and machine intelligence, 28*, 802-817.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation, 1*, 541-551.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*, 2278-2324.

Lee, J.-T., Kim, H.-U., Lee, C., & Kim, C.-S. (2018). Photographic composition classification and dominant geometric element detection for outdoor scenes. *Journal of Visual Communication and Image Representation, 55*, 91–105.

Lehky, S. R., & Sereno, A. B. (2010). Population coding of visual space: modeling. *Frontiers in computational neuroscience, 4*.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE,, 47*, 1940-1951.

Levkwoitz, H. (1990). Integration of Visual Modules. An extension of the Marr Paradigm (book reviews). *Sigurt Bulletin, 2*.

Lin, Y., & Fang, Y. (2010). A Computational Model for Saliency Maps by Using Local Entropy. *AAAI*.

Loreto, V., Mukherjee, A., & Tria, F. (2012). On the origin of the hierarchy of color names. *Proceedings of the National Academy of Sciences, 109*, 6819-6824.

Ma, K., Zhao, T., Zeng, K., & Wang, Z. (2015). Objective quality assessment for color-to-gray image conversion. *IEEE Transactions on Image Processing, 24*, 4673-4685.

Maeda, S., Mukunoki, M., & Ikeda, K. (1999). A classification method of images based on composition and its application to image retrieval. *Proceedings IEEE International Conference on Multimedia Computing and Systems*, *2*, pp. 240–244.

Majewicz, P., & Smith, K. K. (2013, 11). Method and apparatus for converting a color image to grayscale. Google Patents.

Manzotti, R. (2017). A Perception-Based Model of Complementary Afterimages. *SAGE Open, 7*, 2158244016682478.

Marr, D. (1982). *Vision*. WH San Francisco: Freeman and Company.

Mathibela, B., Posner, I., & Newman, P. (2013). A roadwork scene signature based on the opponent colour model. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 4394–4400).

Medathati, N. K., Neumann, H., Masson, & Kornprobst, P. (2016). Bio-inspired computer vision: Towards a synergistic approach of artificial and biological vision. *Computer Vision and Image Understanding, 150*, 1-30.

Meilinger, T. (2008). The network of reference frames theory: A synthesis of graphs and cognitive maps. *International Conference on Spatial Cognition*, (pp. 344-360).

Mel, B. W., Ruderman, D. L., & Archie, K. A. (1998). Translation-invariant orientation tuning in visual "complex" cells could derive from intradendritic computations. *The Journal of Neuroscience,, 18*, 4325-4334.

Milner, M. (1987). What is art? And what is genius in art? When I set out to prepare this lecture I intended to try to select out of writings on art, both by analysts and non-analysts, whatever might point the direction towards an answer to these questions. *Psycho-analysis and Contemporary Thought*, 77.

Mitchell, W. (1994). *The Pictorial turn. Pictorial theory: Essays on verbal and visual representations.* University of Chicago Press.

Mitchell, W. (2002). Sowing seeing: A critique of visual culture. *Art history, aesthectics, visual studies*, 231-50.

Moles, A. (1971). *Art et Ordinateur.* Casterman.

Moles, A. (1973). *Théoric de L´information et Perception esthétique. Paris, Denoël.* Denoël.

Munsell, A. H. (1915). *Atlas of the Munsell color system.* Wadsworth, Howland & Company, Incorporated, Printers.

Newton, I. (1730). *Optics: or a treatise of the refections, refractions, inflections and colours of light.* William Innys at the West-End of St. Paul's.

Ng, P. (2013, 1). Method and device for use in converting a colour image into a grayscale image. Google Patents.

Norheim, E. S., Wyller, J., Nordlie, E., & Einevoll, G. T. (2012). A minimal mechanistic model for temporal signal processing in the lateral geniculate nucleus. *Cognitive Neurodynamics, 6*, 259-281.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: a computational model of fixation durations in scene viewing. *Psychological Review, 117*, 382.

Panofsky, E. (1962). *Studies in Iconology.* Harper & Row.

Pawlik, J. (1976). *Theorie der Farbe.*

Pertzov, Y., Avidan, G., & Zohary, E. (2011). Multiple reference frames for saccadic planning in the human parietal cortex. *Journal of Neuroscience, 31*, 1059-1068.

Peters, R. J., & Itti, L. (2007). Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, (pp. 1-8).

Petscharnig, S., Lux, M., & Chatzichristofis, S. (2017). Dimensionality reduction for image features using deep learning and autoencoders. *Proceedings of the 15th international workshop on content-based multimedia indexing*, (pp. 1–6).

Posner, M. I. (1993). 15 Attention before and during the Decade. *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience, 14*, 343.

Pouget, A., & Sejnowski, T. J. (1997). *Lesion in a Basis Function Model of Parietal Cortex: Comparison with Hemineglect. Parietal Lobe Constribution in Orientation 3 D Space.* (P. Thier, & H. O. Karnath, Eds.) Springer-Verlag.

Pridmore, R. W. (2008). Chromatic induction: Opponent color or complementary color process? *Color Research \& Application, 33*, 77-81.

Przybyszewski, A. W. (1998). Vision: Does top-down processing help us to see? *Current Biology, 8*, R135--R139.

Puhalla, D. M. (2008). Perceiving hierarchy through intrinsic color structure. *Visual Communication, 7*, 199-228.

Qiu, A., Rosenau, B. J., Greenberg, A. S., Hurdal, M. K., Barta, P., Yantis, S., & Miller, M. I. (2006). Estimating linear cortical magnification in human primary visual cortex via dynamic programming. *Neuroimage, 31*, 125-138.

Raizada, R. D., & Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral Cortex, 13*, 100-113.

Ramirez-Quintana, J. A., Chacon-Murguia, M. I., & Ramirez-Alonso, G. M. (2018). Adaptive background modeling of complex scenarios based on pixel level learning modeled with a retinotopic self-organizing map and radial basis mapping. *Applied Intelligence, 48*, 4976–4997.

Rayner, K., Reichle, E. D., & Pollatsek, A. (1998). Eye movement control in reading: An overview and model. In *Eye guidance in reading and scene perception* (pp. 243-268). Elsevier.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences, 26*, 445-476.

Rock, I., & Palmer, S. (1990). The legacy of Gestalt psychology. *Scientific American, 263*, 84-91.

Rosch, E. (1975). The nature of mental codes for color categories. *Journal of experimental psychology: Human perception and performance, 1*, 303.

Rui, Y., Huang, T. S., Chang, S.-F., & others. (1999). Image retrieval: Past, present, and future. *Journal of Visual Communication and Image Representation, 10*, 1–23.

Runge, P. O. (2010). *Color Sphere.* Pricenton Architectural Press.

Sabry, E. S., Elagooz, S. S., Abd El-Samie, F. E., El-Shafai, W., El-Bahnasawy, N. A., El-Banby, G. M., . . . Ramadan, R. A. (2023). Image Retrieval Using Convolutional Autoencoder, InfoGAN, and Vision Transformer Unsupervised Models. *IEEE Access, 11*, 20445–20477.

Saleh, B., & Elgammal, A. (2015). Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *arXiv preprint arXiv:1505.00855*.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research, 1*, 201-220.

Sánchez Cesteros, Ó., & Rincón Zamorano, M. (2017). *España Patent No. 201831253*.

Sánchez, Ó., & Rincón, M. (2009). Image Equilibrium: A Global Image Property for Human-Centered Image Analysis. *Bioinspired Applications in Artificial and Natural Computation: Third International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2009, Santiago de Compostela, Spain, June 22-26, 2009, Proceedings, Part II 3*, (pp. 216–224).

Sanchez-Cesteros, O., Rincon, M., Bachiller, M., & Valladares-Rodriguez, S. (2023). A Long Skip Connection for Enhanced Color Selectivity in CNN Architectures. *Sensors, 23*, 7582.

Santos, I., Castro, l., Rodriguez-Fernandez, N., Torrente-Patino, A., & Carballal, ,. (2021). Artificial neural networks and deep learning in the visual arts: A review. *Neural Computing and Applications, 33*, 121-157.

Schiller, P. H., Sandell, J. H., & Maunsell, J. H. (1986). Functions of the ON and OFF channels of the visual system. *Nature, 322*, 824-825.

Schneider, K. A., Richter, M. C., & Kastner, S. (2004). Retinotopic organization and functional subdivisions of the human lateral geniculate nucleus: a high-resolution functional magnetic resonance imaging study. *The Journal of Neuroscience, 24*, 8975-8985.

Schopenhauer, A. (1816). *Ueber das Sehn und die Farben*. Leipzig: Hartknoch.

Seguin, B. (2018). The Replica Project: Building a visual search engine for art historians. *XRDS: Crossroads, The ACM Magazine for Students, 24*, 24–29.

Seguin, B., diLenardo, I., & Kaplan, F. (2017). Tracking Transmission of Details in Paintings. *DH*.

Seo, J.-W., & Kim, S. D. (2013). Novel pca-based color-to-gray image conversion. *2013 IEEE international conference on image processing*, (pp. 2279–2283).

Shah, M. (2002). Guest introduction: the changing shape of computer vision in the twenty-first century. Springer.

Sherman, S. M. (2005). Thalamic relays and cortical functioning. *Progress in brain research*, 107-126.

Sims, K. (1991). Artificial evolution for computer graphics. *Proceedings of the 18th annual conference on Computer graphics and interactive techniques*, (pp. 319–328).

Smeulders, A. W., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis \& Machine Intelligence*, 1349-1380.

Solomon, S. G., White, A. J., & Martin, P. R. (2002). Extraclassical receptive field properties of parvocellular, magnocellular, and koniocellular cells in the primate lateral geniculate nucleus. *The Journal of neuroscience, 22*, 338-349.

Soodak, R. E. (1986). Two-dimensional modeling of visual receptive fields using Gaussian subunits. *Proceedings of the National Academy of Sciences, 83*, 9259-9263.

Stratton, G. M. (1902). Eye-Movements and the AEsthetics of Visual Form. *Philos. Stud.*

Stratton, G. M. (1906). Symmetry, linear illusions, and the movements of the eye. *Psychological Review, 13*, 82.

Teller, D. (2014). Vision and the Visual System. In E. John Palmer (Ed.).

Underwood, G., Foulsham, T., Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology, 18*, 321-342.

Van Essen, D. C., Newsome, W. T., & Maunsell, J. H. (1984). The visual field representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability. *Vision Research, 24*, 429-448.

Wade, N. J. (2010). Pioneers of eye movement research. *i-Perception, 1*, 33-68.

Wandell, B. A., & Winawer, J. (2011). Imaging retinotopic maps in the human brain. *Vision Research, 51*, 718-737.

Yao, L., Suryanarayan, P., Qiao, M., Wang, J. Z., & Li, J. (2012). Oscar: On-site composition and aesthetics feedback through exemplars for photographers. *International Journal of Computer Vision, 96*, 353–383.

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and vision* (pp. 171-211). Springer.

Young, T. (1801). II. The Bakerian Lecture. On the mechanism of the eye. *Philosophical Transactions of the Royal Society of London, 91*, 23-88.

Zeki, S. (1993). The visual association cortex. *Current Opinion in Neurobiology, 3*, 155-159.

Zeki, S. (2005). The Ferrier Lecture 1995 behind the seen: the functional specialization of the brain in space and time. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360*, 1145-1183.

Zeki, S., Cheadle, S., Pepper, J., & Mylonas, D. (2017). The Constancy of Colored After-Images. *Frontiers in human neuroscience, 11*, 229.

Zhang, B., Niu, L., & Zhang, L. (2021). Image composition assessment with saliency-augmented multi-pattern pooling. *arXiv preprint arXiv:2104.03133*.

Zhao, R., & Grosky, W. I. (2002). Bridging the semantic gap in image retrieval. In *Distributed multimedia databases: Techniques and applications* (pp. 14-36). IGI Global.

MOBILIBVS
MOBILIORA
SAPIENTIA
OMNIBVS