

# Visión en Paralelo: Un Análisis Comparativo de las Arquitecturas Perceptivas Humanas y Artificiales

## I. Introducción: Las Dos Arquitecturas de la Visión

La capacidad de percibir e interpretar el mundo visual es una piedra angular tanto de la inteligencia biológica como de la artificial. Históricamente, se ha recurrido a la analogía del sistema visual humano como una "cámara biológica" para simplificar su comprensión. Sin embargo, un análisis riguroso revela que esta comparación, aunque útil superficialmente, enmascara las profundas diferencias fundamentales en arquitectura, principio y propósito entre los dos sistemas.

El sistema de visión artificial, ejemplificado por una cámara digital, es fundamentalmente un sistema *pasivo de medición*. Su objetivo principal es capturar y registrar una proyección bidimensional de la luz (una imagen) con la mayor fidelidad fotométrica posible a la escena incidente.<sup>1</sup> La "verdad" de un sistema artificial se mide por su precisión en el registro de fotones.

Por el contrario, el sistema de visión humano es un sistema *activo, inferencial y adaptativo*. Su propósito evolutivo no es registrar la luz con precisión, sino construir un *modelo tridimensional útil* del entorno que permita la supervivencia y guíe el comportamiento. El sistema biológico descarta activamente la información redundante y prioriza la inferencia sobre la fidelidad.

Este informe presentará un análisis comparativo exhaustivo, deconstruyendo ambos sistemas desde sus componentes ópticos *front-end* hasta sus arquitecturas de procesamiento *back-end*. Se argumentará que la visión artificial, si bien se ha inspirado en la biología (notablemente en el desarrollo de Redes Neuronales Convolucionales), opera bajo principios fundamentalmente distintos. El núcleo de este análisis refutará la noción de que el ojo humano es una cámara biológica, demostrando que la "imagen" formada en la retina no es una "imagen acabada"<sup>3</sup>, sino una señal sensorial comprimida y pre-procesada. El verdadero logro de la visión humana no reside en la perfección del *sensor* (el ojo), sino en la inmensa capacidad computacional y predictiva del *procesador* (el cerebro).

## II. El Sistema Óptico Front-End: Captura de Fotones

La primera etapa de cualquier sistema de visión es la captura y el enfoque de la luz. Tanto el ojo como la cámara deben resolver este problema óptico fundamental, pero sus soluciones divergen significativamente en complejidad y adaptabilidad.

## A. La Óptica Biológica: El Ojo Humano como Sistema Adaptativo

El ojo humano utiliza un sistema de lentes compuestas para enfocar la luz en el plano sensorial (la retina). Este sistema está dominado por dos componentes refractivos:

1. **La Córnea:** Es la superficie frontal transparente y abombada del ojo. La córnea es responsable de la mayor parte del poder de enfoque del ojo, aportando aproximadamente el 70% de su refracción. Sin embargo, su poder de enfoque es *fijo*.<sup>4</sup>
2. **El Cristalino:** Situado detrás del iris y la pupila, el cristalino aporta el 30% restante del poder de enfoque.<sup>4</sup> Su característica crucial es que su forma es *variable*.

Esta variabilidad es la base del **mecanismo de acomodación**. El cristalino está suspendido por fibras (las zónulas) conectadas a los músculos ciliares. Para enfocar objetos cercanos, estos músculos se contraen, relajando la tensión sobre el cristalino y permitiéndole adoptar una forma más esférica y potente.<sup>5</sup> Este es un sistema de autoenfoque biológico, fluido e integrado.<sup>4</sup>

Además, el **iris**, la parte coloreada del ojo, actúa como un diafragma biológico. Controla el tamaño de la pupila para regular la cantidad de luz que llega a la retina.<sup>4</sup> Esta función es análoga al diafragma de una cámara.<sup>7</sup> No obstante, la pupila humana no solo reacciona a la luminancia; su diámetro también está modulado por estados cognitivos y emocionales, como la atención, el esfuerzo mental o la excitación.<sup>5</sup>

## B. La Abstracción Artificial: El Modelo de Cámara Pinhole (Estenopeica)

En el dominio de la visión por computador, el modelo óptico predominante no es una lente compleja, sino una abstracción geométrica idealizada: el modelo de cámara *pinhole* (estenopeica).<sup>2</sup> Su valor no reside en su fidelidad física, sino en su simplicidad matemática. El modelo *pinhole* describe la cámara como una caja a prueba de luz con una apertura infinitesimal (el *pinhole*) en un lado y un plano de imagen en el opuesto.<sup>8</sup> La luz de un punto de la escena viaja en línea recta a través de esta apertura y proyecta un punto en el plano de la imagen.<sup>2</sup>

Geométricamente, la relación entre un punto 3D en el mundo,  $P = (X, Y, Z)$ , y su proyección 2D en la imagen,  $p = (x, y)$ , se define por la distancia focal  $f$  (la distancia entre el *pinhole* y el plano de la imagen) y la división perspectiva:

$$x = f \frac{X}{Z}$$

$$y = f \frac{Y}{Z}$$

Debido a que los rayos de luz se cruzan en el *pinhole* (el centro de proyección), la imagen resultante en el plano de la imagen está *invertida* tanto vertical como horizontalmente.<sup>2</sup> En la práctica computacional, a menudo se utiliza un "plano de imagen virtual" situado delante del *pinhole* a una distancia  $f$  para evitar el signo negativo y trabajar con una imagen no invertida.<sup>2</sup>

## C. La Realidad de la Ingeniería y la Paradoja del Pinhole

La simplicidad del modelo *pinhole* lo hace ideal para la geometría de la visión por computador, pero como dispositivo físico, es fundamentalmente defectuoso. Un *pinhole* lo suficientemente pequeño como para crear una imagen nítida (limitando cada punto de la escena a un solo rayo) no puede recoger suficiente luz para una exposición razonable.<sup>2</sup> Si se agranda el *pinhole* para recoger más luz, cada punto de la escena proyecta un "círculo de confusión" borroso en la imagen, destruyendo la nitidez.<sup>9</sup>

Este modelo geométrico idealizado también presenta dos propiedades que son opuestas a las de los sistemas de visión reales:

1. **Profundidad de Campo (DoF):** Un *pinhole* ideal tiene una profundidad de campo casi infinita. Dado que solo un rayo de cada punto pasa a través de la apertura, los objetos están teóricamente enfocados a cualquier distancia.<sup>10</sup>
2. **Aberraciones:** El modelo *pinhole* es puramente geométrico e ignora los fenómenos de la óptica física. Por lo tanto, no sufre de aberraciones ópticas (como la esférica o la cromática).<sup>11</sup>

Irónicamente, un *pinhole* físico *real* está limitado no por aberraciones, sino por la *difracción*.<sup>12</sup> A medida que la apertura se hace extremadamente pequeña, la propia naturaleza ondulatoria de la luz hace que esta se disperse, introduciendo un límite fundamental de borrosidad.<sup>12</sup> Las cámaras reales (y el ojo) resuelven la paradoja del *pinhole* (luz vs. nitidez) mediante el uso de **lentes**.<sup>13</sup> Una lente permite una apertura mucho mayor, recogiendo una cantidad de luz significativamente superior, al tiempo que utiliza la refracción para enfocar todos esos rayos de un solo punto de la escena en un solo punto del sensor.<sup>2</sup>

Sin embargo, al introducir lentes, se introducen inevitablemente *aberraciones* y *distorsiones* (p. ej., distorsión radial o de barril) que el modelo *pinhole* no describe.<sup>14</sup> Por lo tanto, la óptica del ojo humano *no* debe compararse con el modelo *pinhole*. La óptica ocular es un sistema de lentes real y complejo que, al igual que una lente de cámara, sufre de aberraciones ópticas significativas.<sup>12</sup>

En la visión por computador, la "calibración de la cámara" es, en gran medida, el proceso de medir y modelar matemáticamente las distorsiones introducidas por la lente<sup>14</sup>, con el fin de *corregir* la imagen capturada para que se *aproxime* al modelo *pinhole* ideal, que es el que se utiliza para los cálculos geométricos. El cerebro humano realiza una tarea análoga, aprendiendo a "calibrar" e interpretar la señal óptica imperfecta que recibe de su propio

sistema de lentes biológicas.

### III. El Plano de Detección: La Retina vs. El Sensor Digital

Tras el sistema óptico, la luz enfocada debe ser convertida en una señal eléctrica. Aquí, la divergencia entre el diseño biológico y el de ingeniería es quizás la más profunda.

#### A. El "Sensor" Biológico: La Retina como Red Neuronal Preliminar

La analogía fundamental (retina = sensor) es profundamente errónea. La retina no es un sensor pasivo; es una red neuronal de tres capas, laminada y altamente sofisticada, que realiza un preprocesamiento y una compresión de datos masiva antes de que la señal salga del ojo.

La diferencia más notable es su arquitectura *no uniforme*.<sup>15</sup> A diferencia de la cuadrícula uniforme de un sensor de cámara <sup>16</sup>, la retina tiene una distribución de fotorreceptores radicalmente desigual:

- **Conos:** Existen aproximadamente 6 millones de conos.<sup>15</sup> Son responsables de la visión del color (fotópica) y requieren buena iluminación.<sup>4</sup> crucialmente, están concentrados de forma masiva en la *fóvea* <sup>4</sup>, una pequeña depresión en el centro de la retina responsable de la visión central nítida y detallada. El ojo humano utiliza tres tipos de conos, cada uno sensible a diferentes longitudes de onda (rojo, verde, azul), permitiendo la visión tricromática.<sup>16</sup>
- **Bastones:** Existen aproximadamente 100 millones de bastones.<sup>15</sup> Son extremadamente sensibles a la luz, permitiendo la visión nocturna (escotópica). Sin embargo, solo detectan luminancia (blanco y negro) y movimiento.<sup>4</sup> Dominan la *periferia* de la retina.

#### B. La Falacia de los "Megapíxeles" Humanos

Esta arquitectura no uniforme invalida cualquier comparación directa de "megapíxeles".<sup>15</sup> La cifra de "576 megapíxeles" <sup>17</sup>, citada a menudo, es un mito basado en una extrapolación incorrecta: calcula la agudeza teórica de la fóvea y la aplica a todo el campo visual humano.<sup>15</sup> La realidad es que la visión humana es de *muy alta resolución* en un área diminuta (la fóvea, que cubre solo 1-2 grados del campo visual) y de *muy baja resolución* en toda la periferia.<sup>15</sup> El cerebro crea la *ilusión* de una imagen uniformemente nítida mediante movimientos oculares extremadamente rápidos e inconscientes (sacadas) para apuntar la fóvea a los puntos de interés.<sup>15</sup> El cerebro luego "une" estas instantáneas de alta resolución en un modelo mental

coherente.<sup>17</sup>

## C. La Compresión Retiniana: La Prueba del Procesamiento

La prueba definitiva del procesamiento retiniano reside en su ratio de compresión. La retina contiene más de 106 millones de fotorreceptores (conos + bastones).<sup>15</sup> Sin embargo, la "salida" de la retina, el nervio óptico (que es, en efecto, el "cable de datos" hacia el cerebro), contiene solo unos 1.2 millones de axones (de las células ganglionares).<sup>18</sup>

Esto representa un ratio de compresión de aproximadamente 100:1. La retina no está "enviando la imagen" al cerebro. Está enviando una señal *altamente procesada y comprimida*, que ya codifica características esenciales como bordes, cambios de contraste y movimiento, descartando el 99% de la información espacial y temporal redundante. Es, en efecto, la primera capa de una red neuronal profunda.

## D. El Sensor Artificial: La Arquitectura de Matriz Uniforme

Los sensores de imagen digitales, en cambio, se basan en una cuadrícula (matriz) uniforme de fotodetectores (píxeles). Su objetivo es la fidelidad total.<sup>20</sup> Históricamente, han existido dos tecnologías principales:

1. **CCD (Charge-Coupled Device):** En un sensor CCD, las cargas eléctricas acumuladas en cada píxel (proporcionales a la luz recibida) se *desplazan* (shuttled) a través de la matriz, fila por fila, hasta un único convertidor analógico-digital (ADC) de alta calidad, generalmente externo al sensor.<sup>20</sup>
2. **CMOS (Complementary Metal-Oxide Semiconductor):** En un sensor CMOS, cada píxel (o cada columna de píxeles) tiene su propia circuitería de amplificación y, a menudo, su propio ADC *integrado en el chip*.<sup>1</sup>

## E. El Dominio de CMOS

Aunque los sensores CCD ofrecían inicialmente una calidad de imagen superior (mejor rango dinámico y menor ruido)<sup>20</sup>, la tecnología CMOS domina casi por completo el mercado actual. La razón es la *integración*.

La arquitectura de CMOS permite que el procesamiento de la señal (amplificación, conversión A/D) ocurra *dentro del propio sensor*.<sup>1</sup> Esto reduce drásticamente el costo de fabricación, el tamaño del sistema y el consumo de energía.<sup>20</sup> Además, permite una velocidad de lectura mucho mayor (ya que los píxeles se pueden leer individualmente o en paralelo), lo cual es esencial para el vídeo de alta velocidad y la reducción de artefactos como el *rolling shutter*.<sup>21</sup> Las mejoras continuas en la fabricación de semiconductores han *subsano* en gran medida

las desventajas iniciales de ruido y calidad que tenía el CMOS.<sup>20</sup>

## F. Análisis Comparativo de Métricas de Rendimiento

La comparación directa de métricas de rendimiento subraya las diferentes filosofías de diseño entre la biología y la ingeniería.

**Tabla 1: Comparativa de Rendimiento: Ojo Humano vs. Cámara Digital (Sensor CMOS de Alta Gama)**

Métrica	Sistema de Visión Humano (Ojo + Cerebro)	Sistema de Visión Artificial (Cámara CMOS)	Notas Clave (Fuentes Relevantes)
Resolución	No uniforme. Extremadamente alta en la fovea (centro), muy baja en la periferia.	Uniforme. Millones de píxeles (p.ej., 24-100 MP) distribuidos equitativamente.	El ojo <i>compensa</i> con movimiento (sacadas) y reconstrucción cerebral. La métrica "megapíxel" no es aplicable. <sup>15</sup>
Rango Dinámico	~21 stops (Adaptativo).	~15 stops (Instantáneo, en una sola exposición).	El ojo tiene un rango <i>instantáneo</i> más bajo (~10-14 stops), pero su rango <i>adaptativo</i> (químico y neural) es vasto. La cámara requiere <i>bracketing</i> (HDR) para simular esto. <sup>17</sup>
Sensibilidad (ISO)	Enorme rango adaptativo. Visión escotópica (bastones) altamente sensible.	Limitado. El ISO alto introduce ruido digital.	El ojo cambia de "sensor" (conos a bastones) para adaptarse a la oscuridad. <sup>4</sup>
Campo de Visión	~150° binocular, ~100° monocular temporal. <sup>16</sup>	Dependiente de la lente. Desde "ojo de pez" ( $>180^\circ$ ) hasta teleobjetivo ( $<2^\circ$ ). <sup>16</sup>	El ojo humano equivale aproximadamente a una lente de 22-24 mm <sup>24</sup> para el campo visual general, pero a una de ~85 mm <sup>16</sup> para el campo de atención central.

<b>Enfoque</b>	Acomodación (flexión del cristalino). Continuo y biológico. <sup>4</sup>	Mecánico/Electrónico. Movimiento de elementos de la lente.	<sup>4</sup>
<b>"Tasa de Fotogramas"</b>	No aplicable. Procesamiento continuo de flujo.	Fijo (p.ej., 24, 60, 120 fps).	El ojo detecta movimiento <sup>25</sup> , no fotogramas. El <i>rolling shutter</i> <sup>21</sup> en CMOS es un artefacto que el ojo no tiene.
<b>Procesamiento</b>	Masivo en tiempo real (Retina + Cerebro). Pre-procesamiento <i>in-sensor</i> (retina).	Mínimo en el sensor (ADC en CMOS <sup>20</sup> ). Principalmente en un procesador de imagen (ISP) externo o en postproducción.	<sup>3</sup>

## IV. El Procesamiento de la Señal: La Vía Visual y la Arquitectura de IA

La diferencia más profunda entre los sistemas radica en lo que sucede *después* de que la luz se convierte en una señal eléctrica. En la visión artificial, la imagen de 24 MP es un producto final. En la visión humana, la señal comprimida de 1.2 MP del nervio óptico es solo el *comienzo* de un proceso computacional de una complejidad asombrosa.

### A. La Vía Visual Humana: Un Sistema Jerárquico y Paralelo

La señal de 1.2 millones de axones de cada nervio óptico<sup>18</sup> abandona el ojo y se dirige al cerebro.<sup>19</sup>

Del Quiasma Óptico al NGL:

La primera estructura clave es el Quiasma Óptico. Aquí ocurre una decusación (cruce) parcial fundamental para la visión binocular.<sup>19</sup>

- Las fibras de la *hemirretina nasal* (el lado de la retina más cercano a la nariz) de *ambos* ojos cruzan al hemisferio cerebral opuesto.<sup>19</sup>
- Las fibras de la *hemirretina temporal* (el lado de la retina más cercano a la sien) *no* cruzan y permanecen en el mismo hemisferio.<sup>19</sup>

La consecuencia de esta re-organización es profunda: *todo* el campo visual *izquierdo* (visto por la retina nasal derecha y la retina temporal izquierda) es enviado para su procesamiento al hemisferio *derecho* del cerebro. Y todo el campo visual *derecho* es enviado al hemisferio

izquierdo.<sup>19</sup> Esta es la solución de "cableado" del cerebro para fusionar las dos vistas ligeramente diferentes de los ojos en una única percepción 3D.

Después del quiasma, las fibras (ahora llamadas tractos ópticos) hacen sinapsis en el **Núcleo Geniculado Lateral (NGL)** del tálamo.<sup>19</sup> El NGL no es solo una estación de relevo; es un *portero* (gatekeeper) que organiza la información en capas separadas (por ejemplo, separando las señales de la vía magnocelular, rápida y de movimiento, de las de la vía parvocelular, lenta y de color/detalle<sup>18</sup>) y modula qué información sensorial se transmite a la corteza.

La Corteza Visual Primaria (V1):

Desde el NGL, la información viaja a través de las radiaciones ópticas hasta la Corteza Visual Primaria (V1), ubicada en el lóbulo occipital.<sup>19</sup> V1 es el primer destino cortical y funciona como un "lienzo" computacional. Mantiene un mapa retinotópico: neuronas adyacentes en V1 responden a puntos adyacentes en la retina.<sup>27</sup> Sin embargo, aquí la imagen ya no existe como tal. V1 descompone la señal en sus componentes más básicos: detectores de bordes, orientaciones, colores y frecuencias espaciales.<sup>28</sup>

Las Dos Vías Corticales: Procesamiento Paralelo:

Desde V1, la información se divide y se envía a dos corrientes de procesamiento funcionalmente distintas, una de las ideas más importantes de la neurociencia visual moderna 26:

1. **La Vía Ventral (la vía del "Qué"):** Esta corriente fluye desde V1, a través de V2 y V4, hasta la **Corteza Temporal Inferior**.<sup>29</sup> Está especializada en el *reconocimiento y la identificación*. Es la vía que procesa la *forma*, el *color* y la *textura* para permitirle identificar un objeto, como un plátano.<sup>26</sup>
2. **La Vía Dorsal (la vía del "Dónde/Cómo"):** Esta corriente fluye desde V1, a través de V2 y V5/MT (área medial temporal), hasta la **Corteza Parietal Posterior**.<sup>29</sup> Está especializada en el *espacio y la acción*. Procesa la *localización*, el *movimiento* y la *profundidad* de los objetos, y es fundamental para guiar las acciones motoras, como *alcanzar y agarrar* ese plátano.<sup>26</sup>

## B. La Inspiración Biológica: Las Redes Neuronales Convolucionales (CNN)

La arquitectura de las Redes Neuronales Convolucionales (CNN), que revolucionó la visión por computador en la década de 2010<sup>28</sup>, no surgió de la nada. Es una inspiración computacional directa de la jerarquía de la corteza visual humana, específicamente del trabajo de Hubel y Wiesel en los años 60.<sup>30</sup>

Hubel y Wiesel descubrieron dos tipos principales de células en V1 de gatos<sup>30</sup>:

- **Células Simples:** Responden a barras (bordes) de una orientación específica en una ubicación específica del campo visual.
- **Células Complejas:** También responden a una orientación específica, pero lo hacen en



cualquier lugar dentro de su campo receptivo (son invariantes a la posición). Descubrieron que estas células lograban esta invarianza al *agrupar* (pooling) las respuestas de múltiples células simples.

Esta arquitectura biológica tiene un mapeo directo y elegante con los componentes de una CNN moderna <sup>28</sup>:

- 1. **Capas Convolucionales (Células Simples):** Un *filtro* (o *kernel*) en una capa convolucional es un detector de características (como un detector de bordes). La operación de "convolución" consiste en deslizar este filtro por toda la imagen <sup>31</sup>, imitando funcionalmente una cuadrícula de células simples que buscan el mismo patrón en todas partes.<sup>30</sup>
- 2. **Capas de Pooling (Células Complejas):** Una operación de *Max-Pooling* toma una región de la capa anterior y emite solo la respuesta más fuerte.<sup>31</sup> Esto imita exactamente cómo una célula compleja agrupa las entradas de las células simples para lograr la invarianza a la posición local.<sup>30</sup>

Esta pila de capas convolucionales y de pooling crea una *jerarquía de características*. Las primeras capas aprenden a detectar características simples (bordes, colores).<sup>28</sup> Las capas intermedias combinan estos bordes para detectar formas más complejas (círculos, texturas), análogo al procesamiento en V2 y V4.<sup>28</sup> Finalmente, las capas profundas (totalmente conectadas) combinan estas formas para identificar objetos semánticos (p. ej., "cara", "coche", "plátano"), análogo a la Corteza Temporal Inferior.<sup>26</sup>

**Tabla 2: Analogías Arquitectónicas: Corteza Visual Ventral vs. CNN**

Concepto Biológico (Vía Ventral)	Componente Computacional (CNN)	Función	Fuentes Relevantes
Retina / NGL	Capa de Entrada (Input Layer)	Recepción de la señal de píxeles.	<sup>19</sup>
Células Simples (V1)	<b>Capa Convolucional</b> (Filtros/Kernels)	Detección de características locales (bordes, orientaciones).	<sup>28</sup>
Células Complejas (V1/V2)	<b>Capa de Pooling</b> (p.ej., Max-Pooling)	Agregación de características, creación de invarianza local.	<sup>30</sup>
Jerarquía V1 → V2 → V4	<b>Pila de Capas</b> (Deep Network)	Construcción de características cada vez más complejas (de bordes a texturas, a partes de objetos).	<sup>26</sup>
Corteza Temporal Inferior	Capas Totalmente Conectadas (FCN)	Clasificación semántica	<sup>26</sup>

		(identificación del objeto).	
Campo Receptivo (Neurona)	Tamaño del Filtro/Kernel	La porción de la imagen que la "neurona" (filtro) "ve".	<sup>28</sup>

Sin embargo, esta analogía es incompleta. Las CNN estándar son predominantemente *feed-forward* (la información fluye en una dirección) y modelan casi exclusivamente la **Vía Ventral** (el reconocimiento de "Qué").<sup>26</sup> Carecen de tres componentes críticos del cerebro:

1. **La Vía Dorsal:** Las CNN son notoriamente deficientes en el razonamiento espacial, la navegación y la comprensión de las interacciones físicas ("Dónde/Cómo").<sup>26</sup>
2. **Conexiones de Retroalimentación (Top-Down):** El cerebro tiene más conexiones que van *hacia atrás* (p. ej., de V4 a V1) que *hacia adelante*. Estas conexiones de retroalimentación permiten que el contexto y las expectativas influyan en la percepción, algo de lo que carecen las CNN simples.
3. **Robustez:** Esta arquitectura limitada hace que la IA sea *frágil*. Son susceptibles a ataques adversariales y fallan catastróficamente ante oclusiones, distorsiones o condiciones de iluminación novedosas que un humano manejaría sin esfuerzo.<sup>32</sup>

## V. Síntesis y Perspectivas Futuras: Más Allá de la Analogía

El análisis desmiente la noción de la visión humana como una cámara. Lo que "vemos" no es la imagen que llega a la retina. Es una *construcción* cerebral, un *modelo* 3D inferido.<sup>3</sup>

La visión humana es un proceso de *generación de hipótesis*. El cerebro actúa como un motor de inferencia bayesiana, utilizando la señal sensorial (que, como hemos visto, es pobre, comprimida y de baja resolución en la periferia<sup>15</sup>) como *evidencia* para confirmar o negar sus *predicciones* internas sobre el mundo. Fenómenos como el "llenado" del punto ciego (donde el nervio óptico<sup>19</sup> crea un agujero en la retina, que el cerebro simplemente "pinta"), la eliminación del desenfoque de movimiento de las sacadas oculares<sup>17</sup>, y la increíble robustez al ruido y la oclusión<sup>32</sup>, demuestran que la visión no es una recepción pasiva, sino una *creación* activa.

La inteligencia artificial actual (IA), incluidas las CNN, ha superado a los humanos en tareas *restringidas* de la vía ventral (p. ej., la clasificación de imágenes en bases de datos).<sup>28</sup> Sin embargo, carece de la *generalidad*, *robustez* y *comprensión contextual* de un niño.<sup>32</sup>

El camino a seguir para la visión artificial no es solo construir sensores con más megapíxeles<sup>15</sup> o rangos dinámicos más altos.<sup>22</sup> El camino es imitar la *arquitectura de procesamiento* superior del cerebro. Esto incluye:

1. Integrar modelos computacionales de la *vía dorsal*<sup>26</sup> para el razonamiento espacial y la interacción.

2. Implementar mecanismos de *retroalimentación* (top-down) para la inferencia contextual y la predicción.
3. Modelar la *adaptación* y el *aprendizaje* en tiempo real, en lugar de los modelos estáticos entrenados que usamos hoy.

El sistema visual humano es la prueba de que se puede lograr una percepción robusta e inteligente a partir de un hardware (neuronas) que es lento, ruidoso e imperfecto, y un sensor (la retina) que está radicalmente comprometido. El secreto no reside en el hardware, sino en la arquitectura de software (la red neuronal biológica). La visión por computador ha pasado la última década imitando con éxito V1 con las Redes Neuronales Convolucionales. Su próximo gran avance provendrá de la imitación del resto de la corteza.

## Fuentes citadas

1. Cómo funcionan los sensores de CMOS: una guía para principiantes - Sinoseen, acceso: noviembre 10, 2025, <https://www.sinoseen.com/es/how-cmos-sensors-work-a-beginners-guide>
2. Focal Length and Intrinsic Camera Parameters | Baeldung on ..., acceso: noviembre 10, 2025, <https://www.baeldung.com/cs/focal-length-intrinsic-camera-parameters>
3. Modelo de visión de Kepler 3, acceso: noviembre 10, 2025, <http://rsefalicante.umh.es/TemasVision/Luz-vision07.htm>
4. Anatomía del ojo: partes del ojo y como vemos - American Academy ..., acceso: noviembre 10, 2025, <https://www.aaao.org/salud-ocular/anatomia/partes-del-ojo>
5. Mecanismos de acomodación del ojo humano | secoir, acceso: noviembre 10, 2025, <https://secoir.org/wp-content/uploads/2022/09/2010-Cap-02-Mecanismos-de-acomodacion-del-ojo-humano.pdf>
6. Cristalino - Wikipedia, la enciclopedia libre, acceso: noviembre 10, 2025, <https://es.wikipedia.org/wiki/Cristalino>
7. ¿En qué se parece el ojo humano a una cámara de fotos? - Clínica Rahhal, acceso: noviembre 10, 2025, <https://rahhal.com/ojo-humano-camara-fotos/>
8. Pinhole camera - Wikipedia, acceso: noviembre 10, 2025, [https://en.wikipedia.org/wiki/Pinhole\\_camera](https://en.wikipedia.org/wiki/Pinhole_camera)
9. ¿Alguien me puede explicar por qué la profundidad de campo y la apertura están relacionadas? : r/photography - Reddit, acceso: noviembre 10, 2025, [https://www.reddit.com/r/photography/comments/lzz3y/can\\_someone\\_explain\\_w\\_hy\\_depth\\_of\\_field\\_and/?tl=es-419](https://www.reddit.com/r/photography/comments/lzz3y/can_someone_explain_w_hy_depth_of_field_and/?tl=es-419)
10. ¿Es necesario enfocar las cámaras estenopeicas? - Lomography, acceso: noviembre 10, 2025, <https://www.lomography.es/school/es-necesario-enfocar-las-camaras-estenopeicas-fa-qkeggjei>
11. Reconstrucción densa de modelos tridimensionales utilizando Visión Artificial, acceso: noviembre 10, 2025, <https://www.ehu.eus/documents/1545039/1570316/10umujika.pdf>
12. CALIDAD VISUAL EN EL RANGO DE ADAPTACIÓN MESÓPICO, acceso: noviembre

- 10, 2025,  
[https://ri.conicet.gov.ar/bitstream/handle/11336/90739/CONICET\\_Digital\\_Nro.78a7c2ed-daf7-49fb-bb04-3aeb3f6fdd16\\_A.pdf?sequence=2&isAllowed=n&date=2020-05-27](https://ri.conicet.gov.ar/bitstream/handle/11336/90739/CONICET_Digital_Nro.78a7c2ed-daf7-49fb-bb04-3aeb3f6fdd16_A.pdf?sequence=2&isAllowed=n&date=2020-05-27)
13. Modelos usados en los sistemas de visión | PPTX - Slideshare, acceso: noviembre 10, 2025,  
<https://es.slideshare.net/slideshow/modelos-usados-en-los-sistemas-de-visin/132930631>
  14. Distorted pinhole camera modeling and calibration - Optica Publishing Group, acceso: noviembre 10, 2025, [https://opg.optica.org/upcoming\\_pdf.cfm?id=412159](https://opg.optica.org/upcoming_pdf.cfm?id=412159)
  15. ¿Alguna cámara tiene mejor resolución que el ojo humano? - Quora, acceso: noviembre 10, 2025,  
<https://es.quora.com/Alguna-c%C3%A1mara-tiene-mejor-resoluci%C3%B3n-que-el-ojo-humano>
  16. Geijo N. El ojo y la cámara fotográfica. Semejanzas y diferencias., acceso: noviembre 10, 2025,  
<https://uvadoc.uva.es/bitstream/handle/10324/19108/TFG-G1799.pdf?sequence=1>
  17. Why do camera sensors struggle to recreate what the human eye can see so readily? : r/photography - Reddit, acceso: noviembre 10, 2025,  
[https://www.reddit.com/r/photography/comments/1ish6sg/why\\_do\\_camera\\_sensors\\_struggle\\_to\\_recreate\\_what/](https://www.reddit.com/r/photography/comments/1ish6sg/why_do_camera_sensors_struggle_to_recreate_what/)
  18. 1 Anatomía de la vía óptica y vías visuales, acceso: noviembre 10, 2025,  
<https://libroseodeoftalmologia.oftalmoseo.com/wp-content/uploads/2021/12/01-n euroftalmologia.pdf>
  19. Actualización desde la anatomía funcional y clínica del sistema ..., acceso: noviembre 10, 2025,  
[https://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S2604-12272022000200071](https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2604-12272022000200071)
  20. Sensores con tecnología CCD vs CMOS - Xataka Foto, acceso: noviembre 10, 2025,  
<https://www.xatakafoto.com/camaras/sensores-con-tecnologia-ccd-vs-cmos>
  21. Sensores digitales de nuestras cámaras fotográficas: CCD y CMOS - YouTube, acceso: noviembre 10, 2025, <https://www.youtube.com/watch?v=eanGailAJ4Y>
  22. If your eyes were cameras, what would the specs be? | Digital ..., acceso: noviembre 10, 2025,  
<https://www.digitalcameraworld.com/news/if-your-eyes-were-cameras-what-would-the-specs-be>
  23. What is the Dynamic Range of the Human Eye? - wolfcrow, acceso: noviembre 10, 2025, <https://wolfcrow.com/what-is-the-dynamic-range-of-the-human-eye/>
  24. El Ojo Humano y la Cámara Digital II - mstudio, acceso: noviembre 10, 2025,  
<https://mstudio.es/el-ojo-humano-y-la-camara-digital-2-parte/>
  25. 2.1. El ojo capta la imagen | Dibujo, Visión y Conocimiento: los engaños de la mente, acceso: noviembre 10, 2025,  
[https://edea.juntadeandalucia.es/bancorecursos/file/045bb4fb-0376-4225-b090-28d6c0936503/1/es-an\\_2017011712\\_9162415.zip/21\\_el\\_ojo\\_capta\\_la\\_imagen.html](https://edea.juntadeandalucia.es/bancorecursos/file/045bb4fb-0376-4225-b090-28d6c0936503/1/es-an_2017011712_9162415.zip/21_el_ojo_capta_la_imagen.html)

[?temp.hn=true&temp.hb=true](#)

26. El sistema visual, acceso: noviembre 10, 2025,  
[http://www.psi.uba.ar/academica/carrerasdegrado/psicologia/sitios\\_catedras/obligatorias/048\\_neuro1/cursada/descargas/old/teorico\\_6.pdf](http://www.psi.uba.ar/academica/carrerasdegrado/psicologia/sitios_catedras/obligatorias/048_neuro1/cursada/descargas/old/teorico_6.pdf)
27. Actualización desde la anatomía funcional y clínica del sistema visual: énfasis en la vía y la corteza visual Funcional an - SciELO México, acceso: noviembre 10, 2025,  
<https://www.scielo.org.mx/pdf/rmof/v96n2/2604-1731-rmo-96-2-71.pdf>
28. Clasificación de imágenes mediante aprendizaje automático: Los ..., acceso: noviembre 10, 2025,  
<https://www.processmaker.com/es/blog/image-classification-using-machine-learning-ins-and-outs/>
29. Corteza visual - Wikipedia, la enciclopedia libre, acceso: noviembre 10, 2025,  
[https://es.wikipedia.org/wiki/Corteza\\_visual](https://es.wikipedia.org/wiki/Corteza_visual)
30. Inspiración Biológica de las Redes Neuronales Convolucionales ..., acceso: noviembre 10, 2025,  
<https://medium.com/soldai/inspiraci%C3%B3n-biol%C3%B3gica-de-las-redes-neuronales-convolucionales-c686f74b4723>
31. Introducción a las redes neuronales convolucionales (CNN) - DataCamp, acceso: noviembre 10, 2025,  
<https://www.datacamp.com/es/tutorial/introduction-to-convolutional-neural-networks-cnns>
32. Universidad Complutense de Madrid (UCM) Universidad Nacional ..., acceso: noviembre 10, 2025,  
[https://www.uned.es/universidad/facultades/dam/jcr:0794be38-fdc1-4dfb-8b92-ac74a9d50294/PFM\\_Esther\\_Gasco\\_Marco.pdf](https://www.uned.es/universidad/facultades/dam/jcr:0794be38-fdc1-4dfb-8b92-ac74a9d50294/PFM_Esther_Gasco_Marco.pdf)
33. Aplicaciones de la Inteligencia Artificial Generativa (IAG) en el Contexto de la Seguridad - O2 Repositori UOC, acceso: noviembre 10, 2025,  
<https://openaccess.uoc.edu/bitstream/10609/150603/1/salgadogbTFM0624memoria.pdf>