

Introducción a la visión

Visión por Computador, curso 2024-2025

Silvia Martín Suazo, silvia.martin@u-tad.com

19 de septiembre de 2024

U-tad | Centro Universitario de Tecnología y Arte Digital



La visión física y biológica

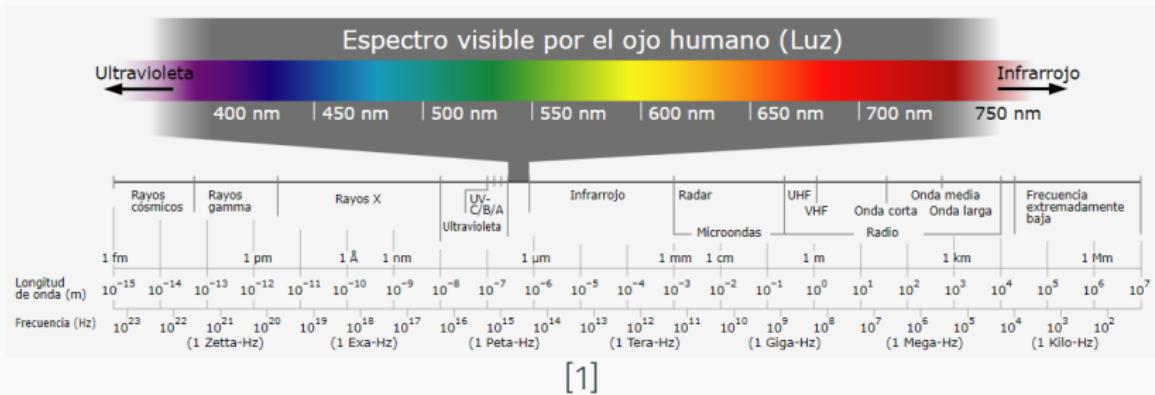
¿Qué es la visión?

La visión se define como la capacidad de captar el entorno que nos rodea gracias a la **captación de rayos de luz**. Este sentido es vital para que el ser humano se relacione con su entorno de manera adecuada.

- Se considera que al menos **un tercio** del cerebro está dedicado a procesar el sentido de la vista.
- El procesamiento cerebral de la información visual juega un papel **crucial** en la tarea de la visión.

El espectro electromagnético visible

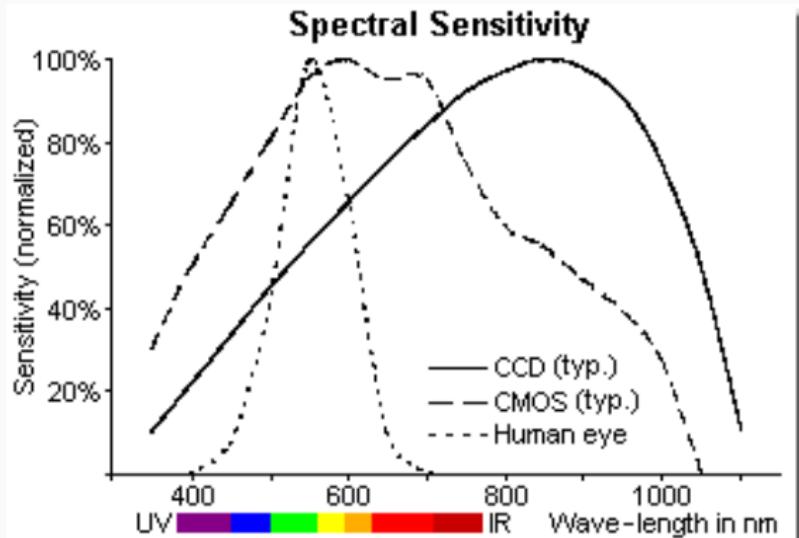
El ser humano solo es **capaz** de ver las longitudes de onda entre **380** y **750 nm**.



Sin embargo esto no es igual para todos los **seres vivos** ni todos los mecanismos para **capturar luz...**

El espectro electromagnético visible

Por ejemplo, un sensor **digital** que captura imágenes en una **cámara fotográfica** es capaz de capturar frecuencias fuera del **espectro visible**, especialmente en el **rango infrarrojo**.



[2]

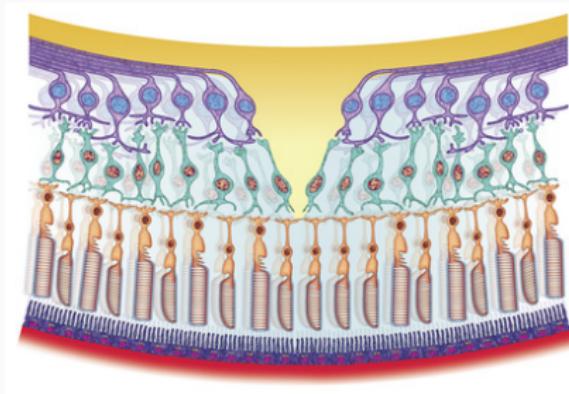
Espectro infrarrojo



La visión en el ser humano

El ser humano es capaz de **recibir y procesar** la información de los **rayos de luz** a través de células sensibles a este **espectro electromagnético**:

- **Conos:** Se encargan de la visión de **alta intensidad** (visión diurna).
- **Bastones:** Se encargan de la visión de **baja intensidad** (visión nocturna).

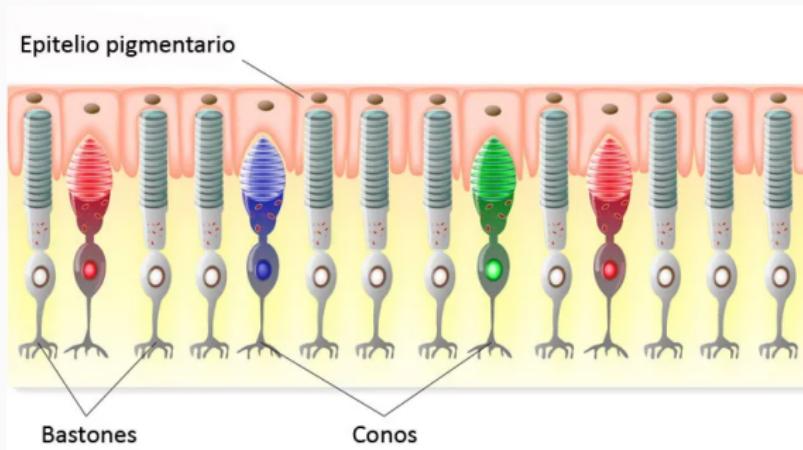


[3]

Visión a color

Para poder captar los distintos **colores** que una persona puede ver, los **conos** se dividen en 3 tipos:

- **Conos L:** Encargados de capturar el espectro **rojo**.
- **Conos S:** Encargados de capturar el espectro **azul**.
- **Conos M:** Encargados de capturar el espectro **verde**.



[4]

Visión a color

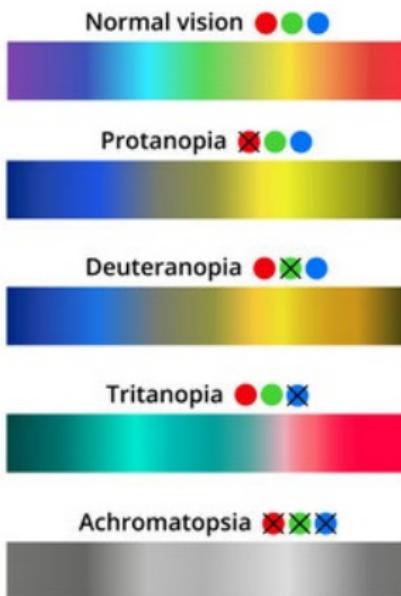
No todos los humanos son capaces de capturar todo el espectro visible. Existen ciertas personas que cuentan con daltonismo, lo que les impide diferenciar correctamente ciertos rangos del espectro.

Por ejemplo las personas que cuentan con tritanopia son insensibles al espectro azul, debido a no tener conos S.



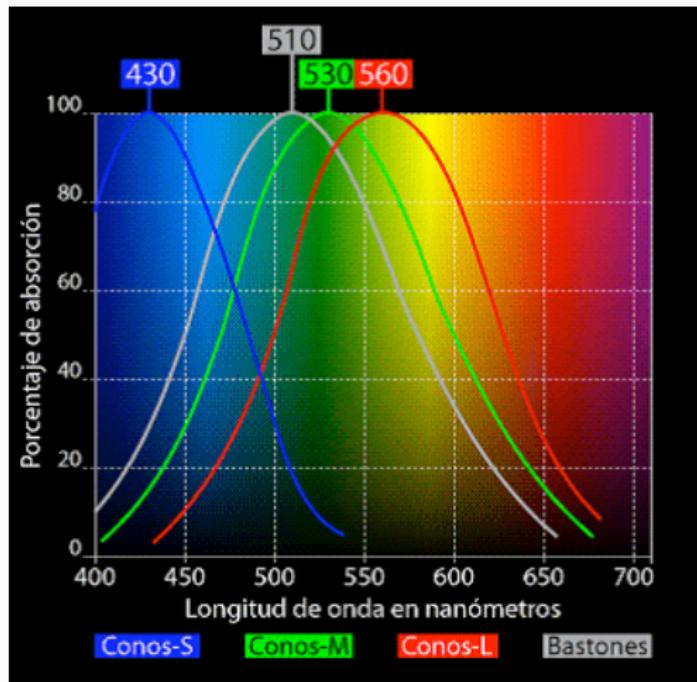
[5]

Colorblindness



Rango del color percibido

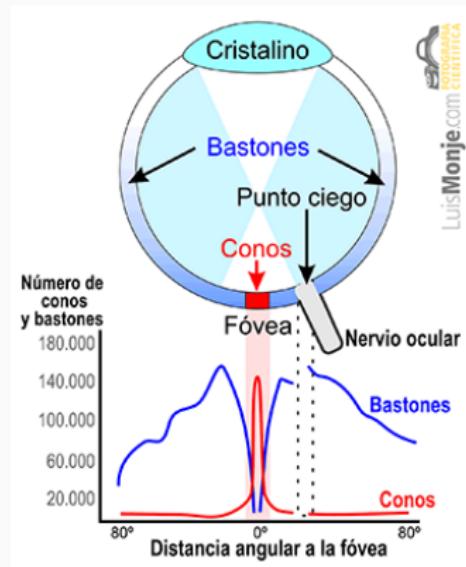
Sin embargo, **no todos** los colores se captan con la misma **intensidad**.



[6]

Fóvea

La **fóvea** es el punto de **mayor sensibilidad** del ojo humano. Es el punto donde se **enfocan** los rayos de luz para la formación de una imagen nítida.



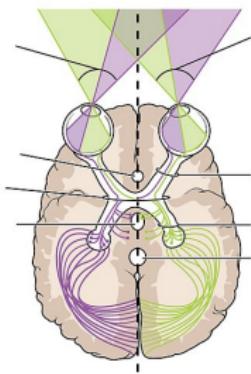
[7]

La visión neuronal

El cerebro en la visión

La visión no sólo es **captar la luz**, sino que el procesamiento de estas **señales** juega un papel crucial en la formación de imágenes.

El cerebro se encarga de **dar sentido** a los estímulos capturados por las células sensibles. El **hardware** se encarga de **captar la señal** mientras que el **software** la procesa adecuadamente.

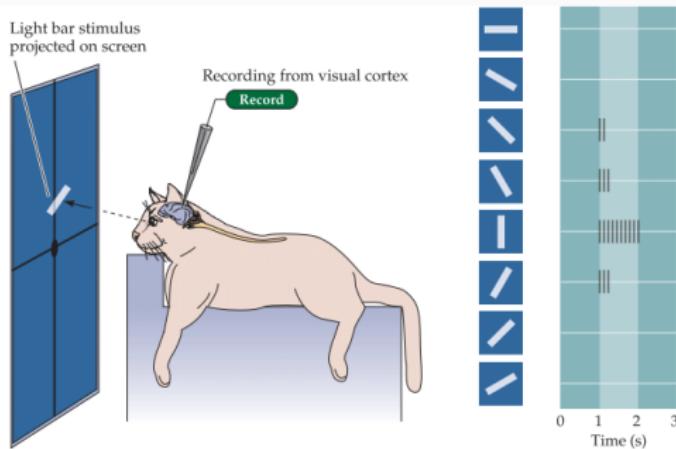


[8]

Las neuronas en la visión

La captación de las imágenes se realiza **primeramente** en la **corteza visual primaria**, donde las neuronas procesan la información recibida por los conos y bastones.

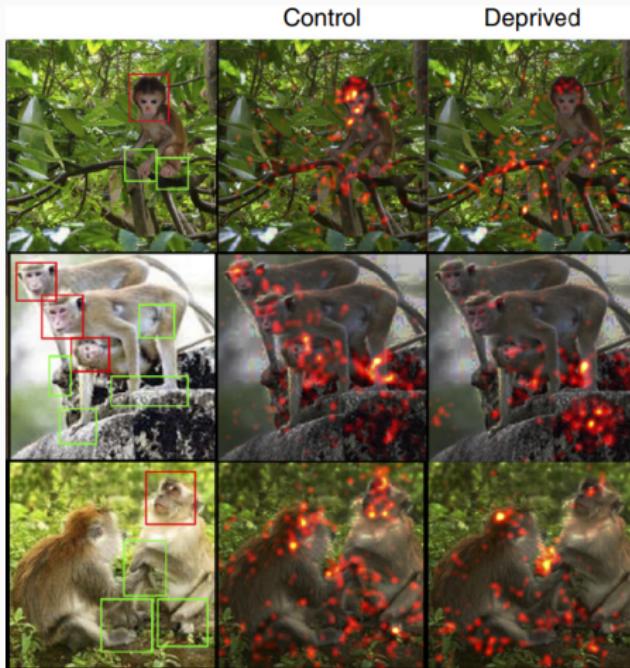
Las **neuronas** que primero procesan la información se encargan de reconocer **formas sencillas**, a medida que se traspasa la información el **cerebro** es capaz de reconocer estructuras **más complejas**[9].



[10]

Visión aprendida vs. visión innata

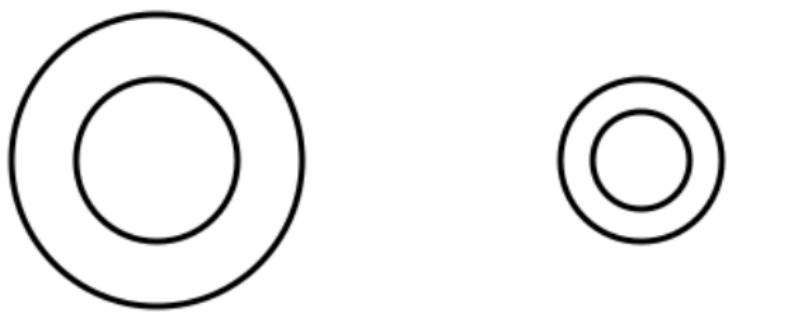
Existen ciertos estudios que sugieren que ciertas **habilidades visuales** se aprenden.



[11]

Visión aprendida vs. visión innata

Sin embargo, es **innegable** que ciertos aspectos de la visión son **innatos** a los **seres vivos**.

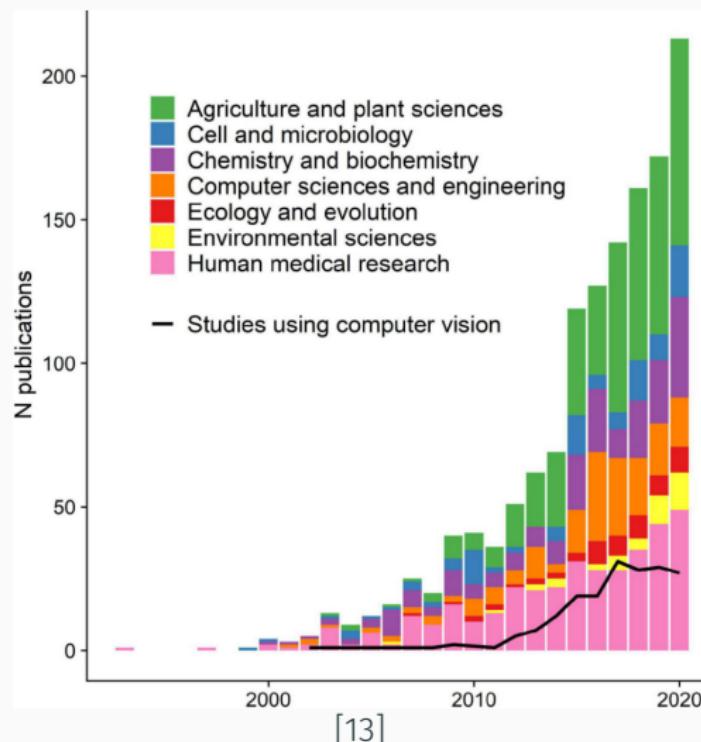


[12]

Introducción a la visión por computador

Visión por computador

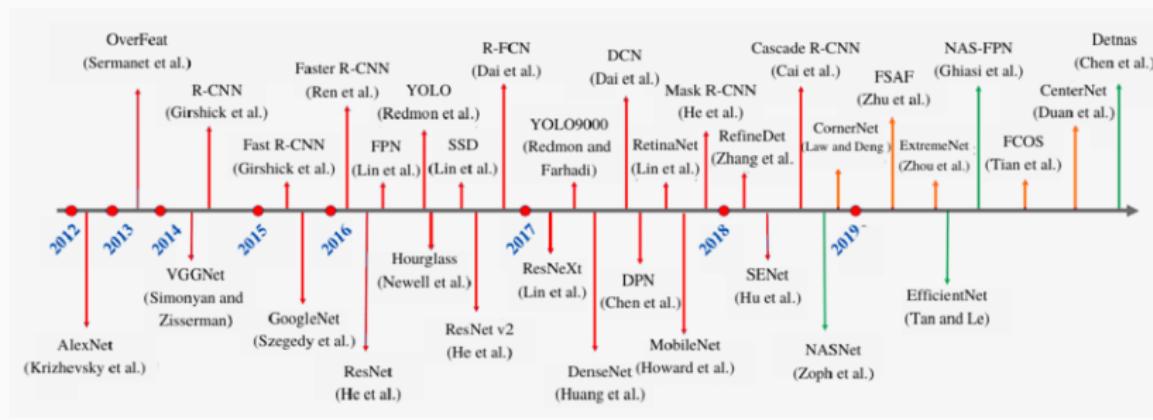
La **visión por computador** es la rama de la **inteligencia artificial** encargada del procesamiento de **imágenes digitales**.



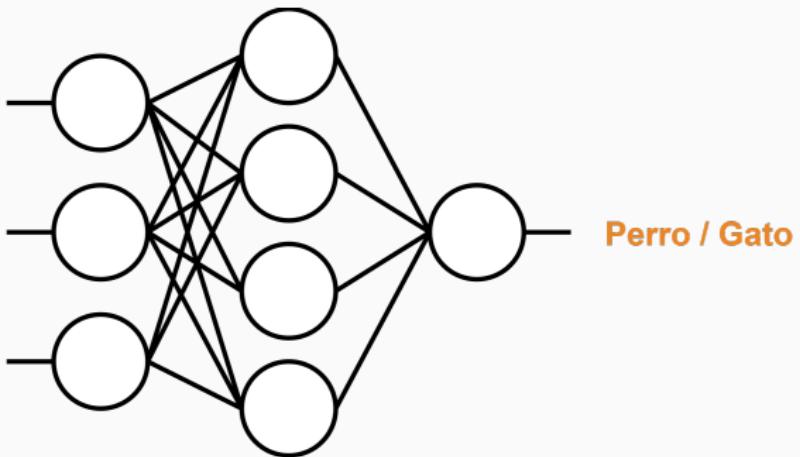
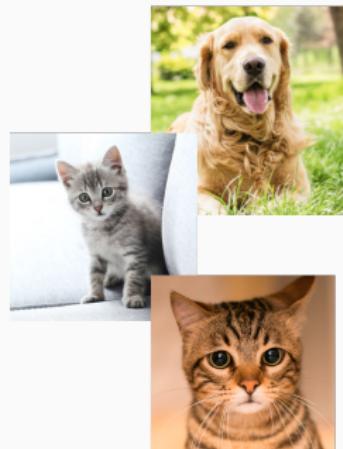
Visión por computador

Es un campo de estudio en presente **auge**, donde se están produciendo grandes **avances** en los últimos años, apoyados de mejoras en la técnica.

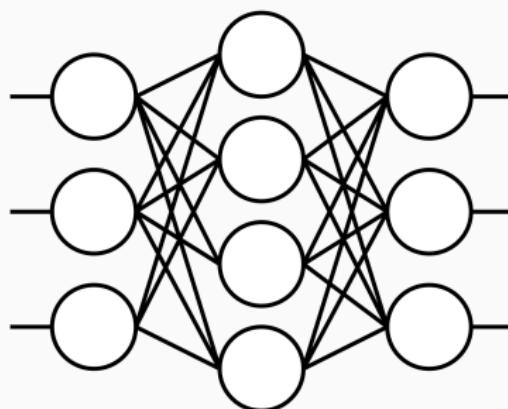
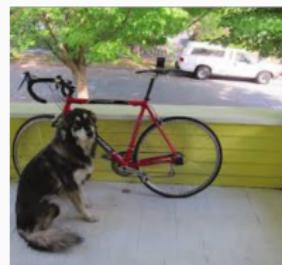
Por ejemplo, los grandes avances en **redes neuronales** durante los últimos años crean un **caldo de cultivo ideal** para la investigación en aplicaciones de la visión por computador.



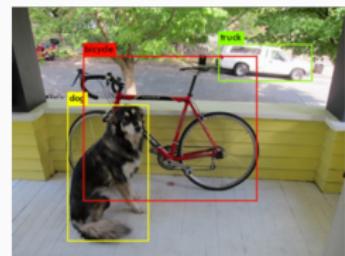
Clasificación



Detección de objetos



Demo youtube[15]



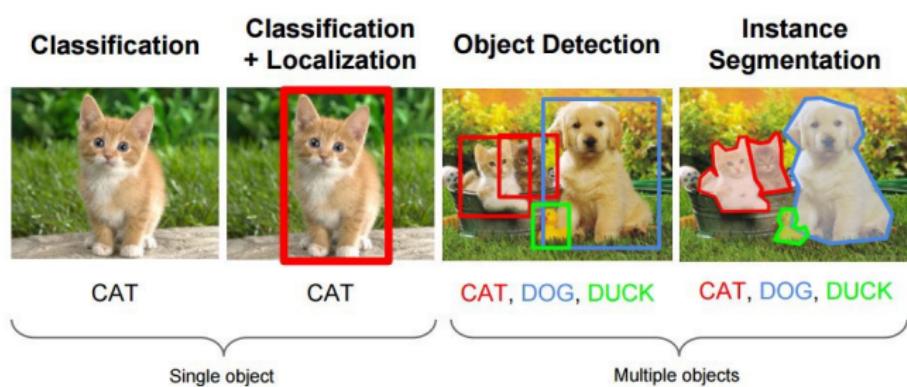
Segmentación de objetos



[16]

Comparación aplicaciones

Computer Vision Tasks

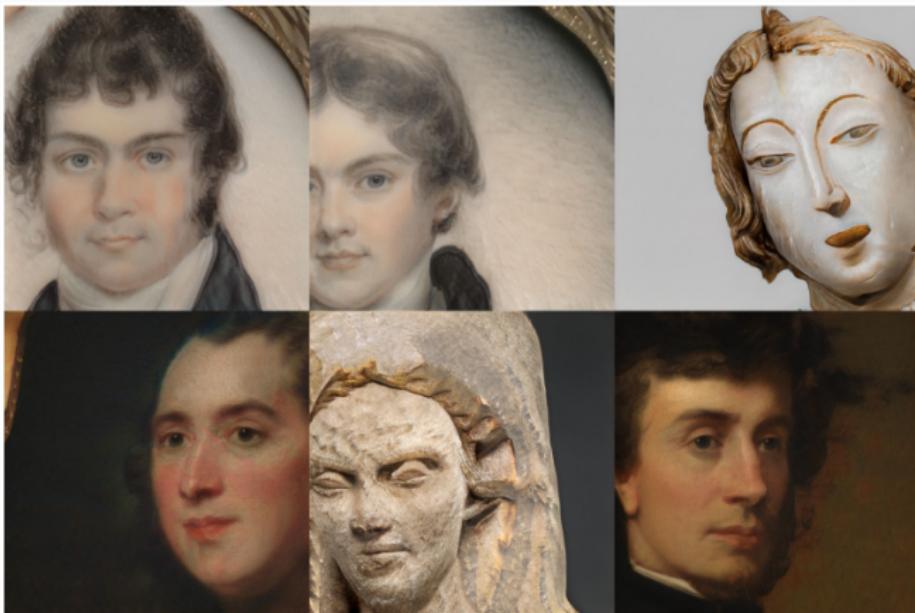


Generación de imágenes



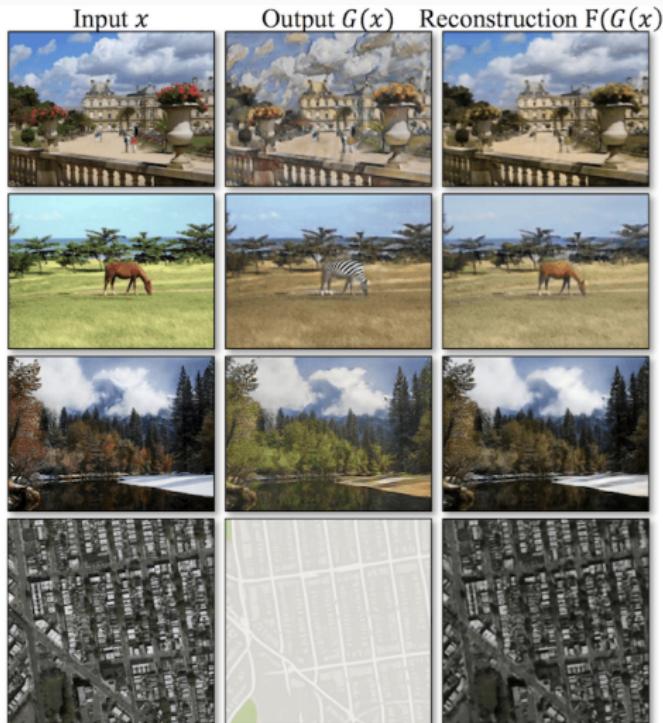
Vídeo youtube [17]

Generación de imágenes



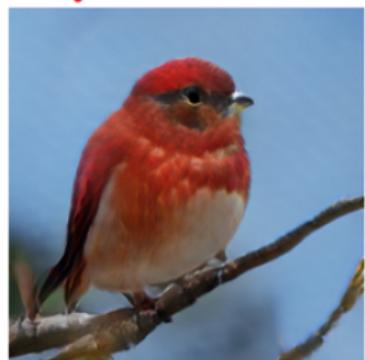
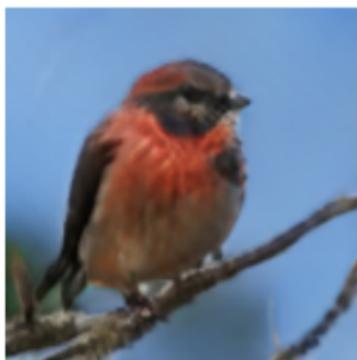
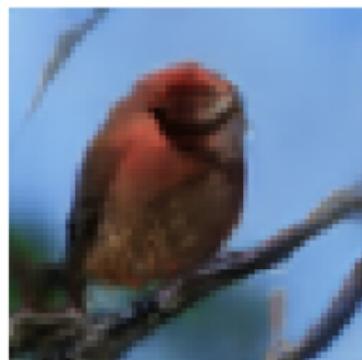
Web NVIDIA [18]

Image-to-image translation



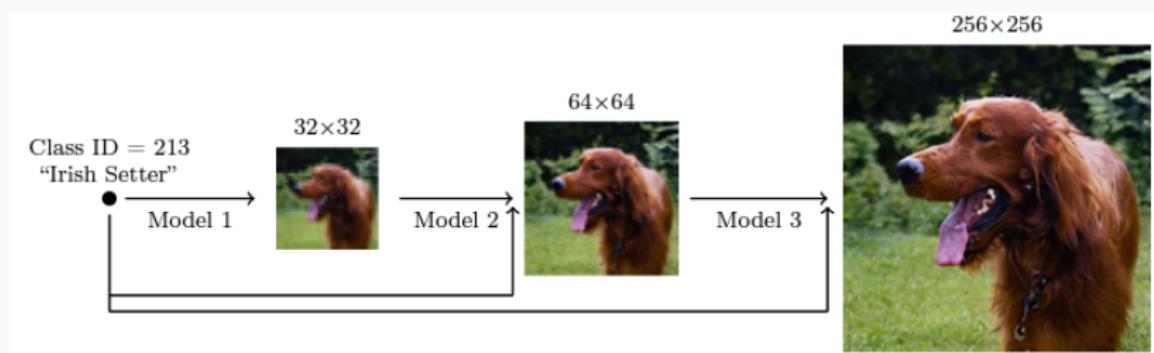
Text-to-image translation

this bird is red with white and has a very short beak



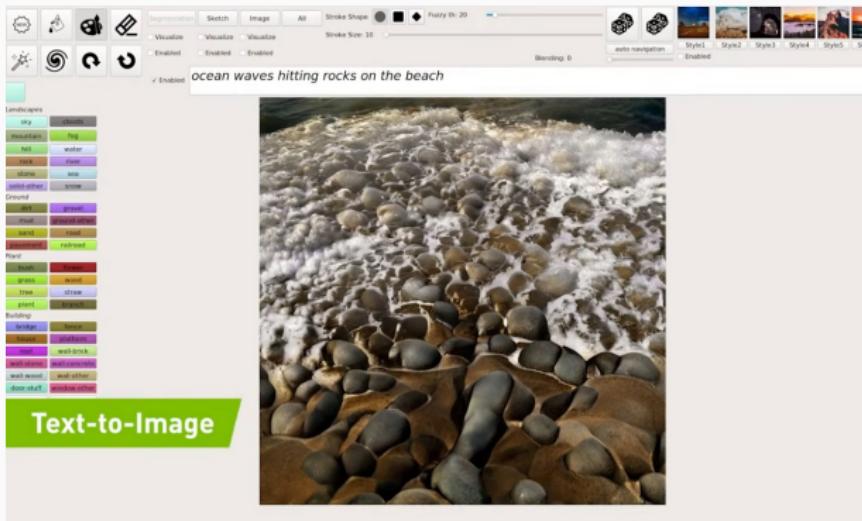
[20]

Text-to-image translation: superresolution



Web Google [21]

Text-to-image translation: *sketch*



YouTube example [22]

Text-to-image translation: *Midjourney*



People having picnic on the grass? ;surrealism;[23]

Referencias i

- [1] Wikipedia.
Visible spectrum image.
[Online; accessed August, 2022].
- [2] AMRR Bandara.
A music keyboard with gesture controlled effects based on computer vision.
PhD thesis, Thesis, University of Sri Jayewardenepura, 2011.
- [3] Opticos-Optometristas CNOO.
Retina image.
[Online; accessed August, 2022].
- [4] Arnaldo Rodríguez (Biblogteca).
Cones image.
[Online; accessed August, 2022].

Referencias ii

- [5] Wikipedia.
Tritanopia spectrum image.
[Online; accessed August, 2022].
- [6] Laboratorio Óptico Foucault.
Sensible range image.
[Online; accessed August, 2022].
- [7] Luis Monje.
Fovea image.
[Online; accessed August, 2022].
- [8] Wikipedia.
Brain hemispheres image.
[Online; accessed August, 2022].

Referencias iii

- [9] David H Hubel and Torsten N Wiesel.
Receptive fields of single neurones in the cat's striate cortex.
The Journal of physiology, 148(3):574, 1959.
- [10] David Berga Garreta.
Understanding eye movements: psychophysics and a model of primary visual cortex.
2019.
- [11] Michael J Arcaro, Peter F Schade, Justin L Vincent, Carlos R Ponce, and Margaret S Livingstone.
Seeing faces is necessary for face-domain formation.
Nature neuroscience, 20(10):1404–1412, 2017.
- [12] Wikipedia.
Optical illusion image.
[Online; accessed August, 2022].

- [13] Moritz D Lürig, Seth Donoughe, Erik I Svensson, Arthur Porto, and Masahito Tsuboi.
Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology.
Frontiers in Ecology and Evolution, 9:642774, 2021.
- [14] Megha Nain, Shilpa Sharma, and Sandeep Chaurasia.
Safety and compliance management system using computer vision and deep learning.
In *IOP Conference Series: Materials Science and Engineering*, volume 1099, page 012013. IOP Publishing, 2021.
- [15] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi.
You only look once: Unified, real-time object detection.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

Referencias v

- [16] Alexander Kirillov, Yuxin Wu, Kaiming He, and Ross Girshick.
Pointrend: Image segmentation as rendering.
In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9799–9808, 2020.
- [17] Tero Karras, Samuli Laine, and Timo Aila.
A style-based generator architecture for generative adversarial networks.
In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.
- [18] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila.
Alias-free generative adversarial networks.
Advances in Neural Information Processing Systems, 34:852–863, 2021.

- [19] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros.
**Unpaired image-to-image translation using cycle-consistent
adversarial networks.**
CoRR, abs/1703.10593, 2017.
[Online; accessed September, 2023].
- [20] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan,
Xiaolei Huang, and Xiaodong He.
**AttnGAN: Fine-grained text to image generation with attentional
generative adversarial networks.**
*In Proceedings of the IEEE conference on computer vision and
pattern recognition*, pages 1316–1324, 2018.

Referencias vii

- [21] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans.
Cascaded diffusion models for high fidelity image generation.
J. Mach. Learn. Res., 23:47–1, 2022.
- [22] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu.
Semantic image synthesis with spatially-adaptive normalization.
In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2337–2346, 2019.
- [23] Midjourney.
Recent showcases.
[Online; accessed September, 2023].