Rodolfo Ferro
ferro@cimat.mx
https://rodolfoferro.xyz

Clubes de Ciencia México 2025
"Satélites y Neuronas: Explorando lo Invisible con AI"

Julio, 2025

Introducción a CV

Día 4

Instructor del club

Rodolfo Ferro (ferro@cimat.mx)

- Sr. SWE (Data Engineer) @ Bisonic México
- Tutor de Ciencia de Datos en Código Facilito y en el Diplomado en Ciencia de Datos de la ENES UNAM León
- Profesor de AI & Coordinador del TomorrowLab @ EdgeHub School of Innovation (Ags.)
- **Formación:** BMath @ UG, CSysEng @ UVEG, StatMethods Specialist @ CIMAT
- **Experiencia:** ML Engineer @ Vindoo.ai (España), Sherpa Digital en IA @ Microsoft México, AI Research Assistant @ CIMAT & AI Research Intern @ Harvard.

ferro@cimat.mx

Clubes de Ciencia México

Tabla de contenidos

1 Visión computacional profunda
- Introducción a imágenes
- Convoluciones & Pooling
- Redes neuronales convolucionales
- Clasificadores de imágenes (LeNet, VGG16, etc.)
- Trabajos relacionados y avances recientes

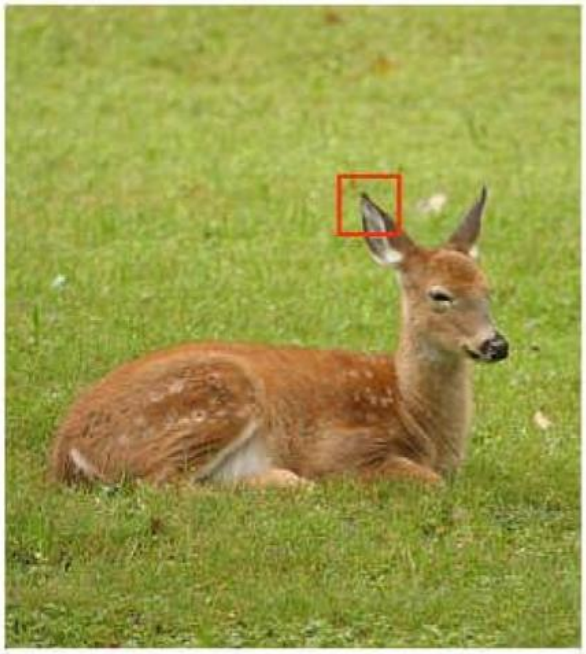
ferro@cimat.mx

Clubes de Ciencia México

Section 1: **Visión computacional profunda**

¿Qué es una imagen?

Visión computacional profunda

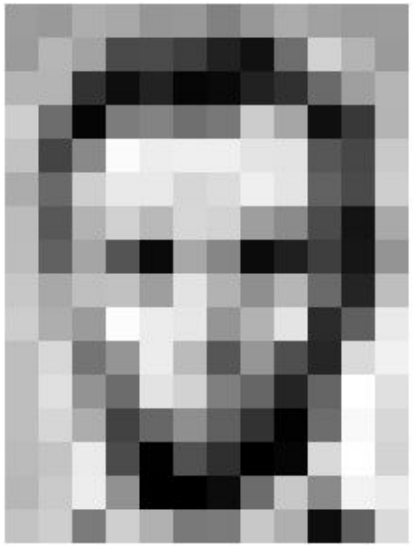


ferro@cimat.mx

Clubes de Ciencia México

¿Qué es una imagen?

Visión computacional profunda



157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

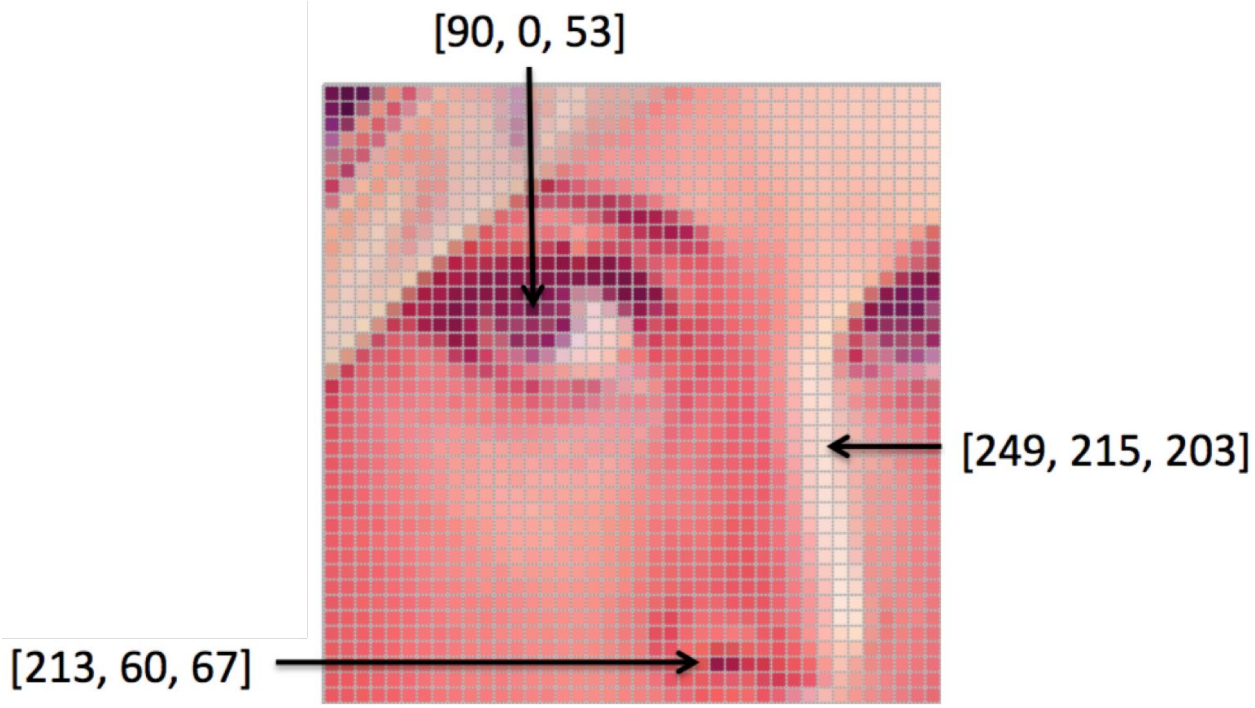
ferro@cimat.mx

Clubes de Ciencia México

"Tutorial 1: Image Filtering." AI Stanford. https://ai.stanford.edu/~syyeung/cvweb/tutorial1.html

¿Qué es una imagen?

Visión computacional profunda



ferro@cimat.mx

Clubes de Ciencia México

"Tutorial 1: Image Filtering." AI Stanford. https://ai.stanford.edu/~syyeung/cvweb/tutorial1.html

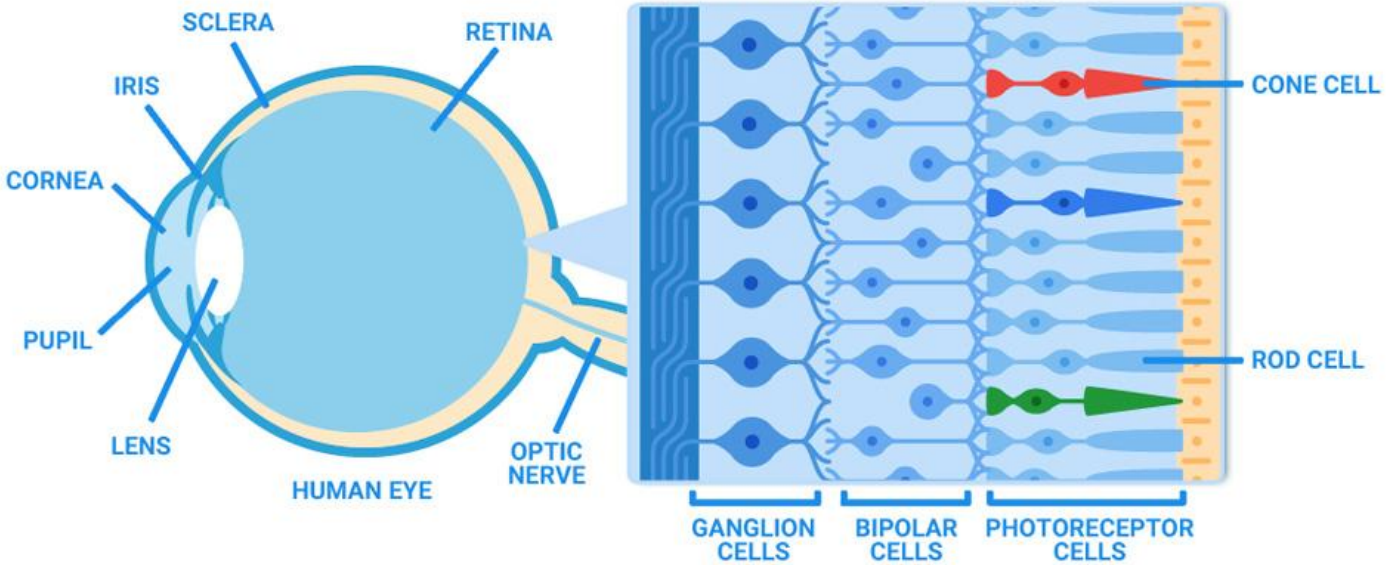
¿Qué es una imagen?

- Una imagen es un arreglo de pixeles, la cual puede tener 1 o más canales de color. Usualmente:
  - 1 canal de color → Escala de grises
  - 3 canales de color → Escala RGB
  - 4 canales de color → Escala RGBA
- Un pixel puede ser visto como un objeto 5-dimensional (x, y, r, g, b).

ferro@cimat.mx

La biología humana

Visión computacional profunda

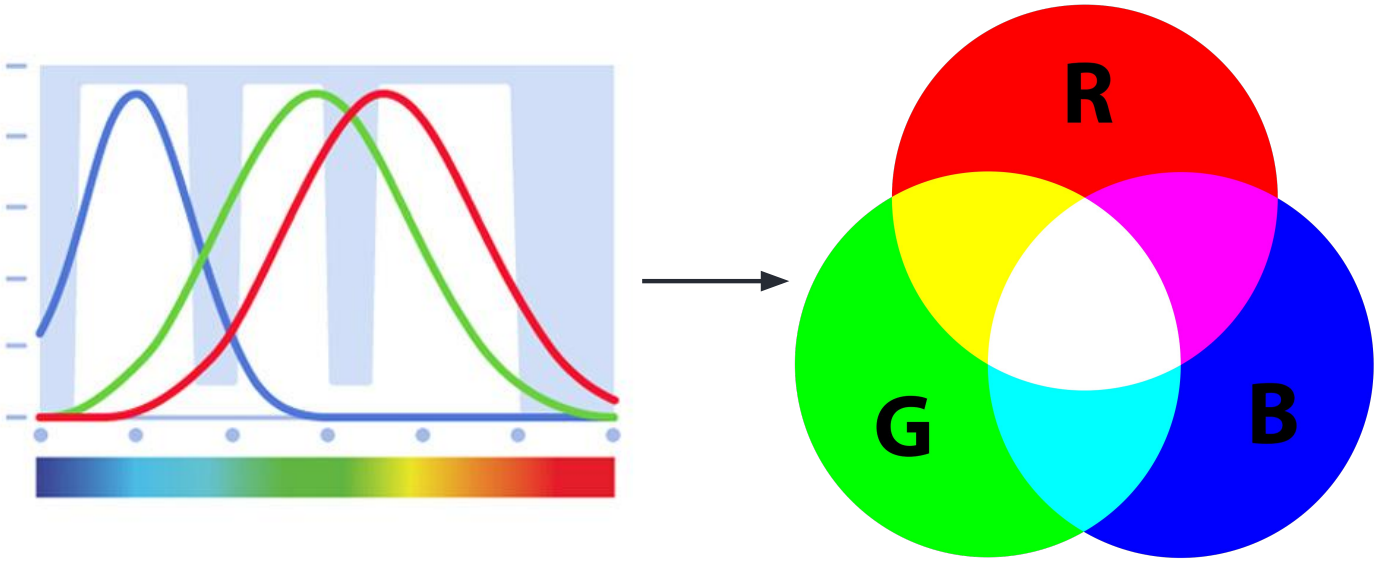


Clubes de Ciencia México

"How EnChroma Color Blind Glasses Work." EnChroma. http://enchroma.com/technology/

La biología humana

Visión computacional profunda



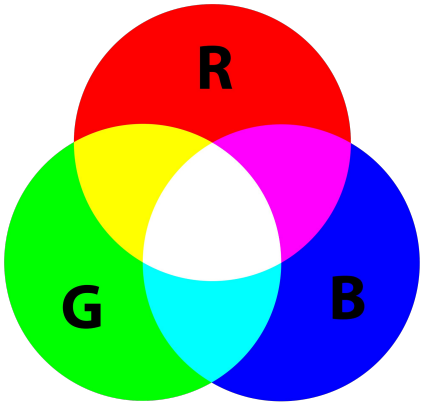
ferro@cimat.mx

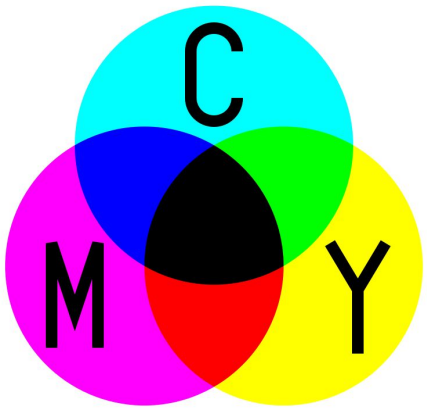
Clubes de Ciencia México

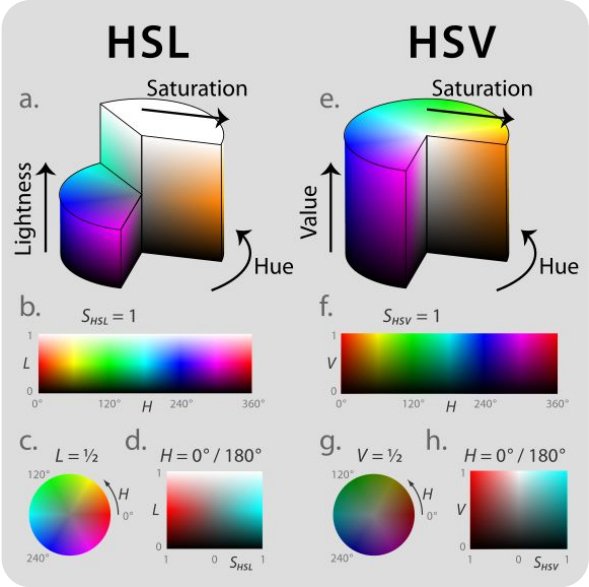
"How EnChroma Color Blind Glasses Work." EnChroma. http://enchroma.com/technology/

Espacios de color

Visión computacional profunda



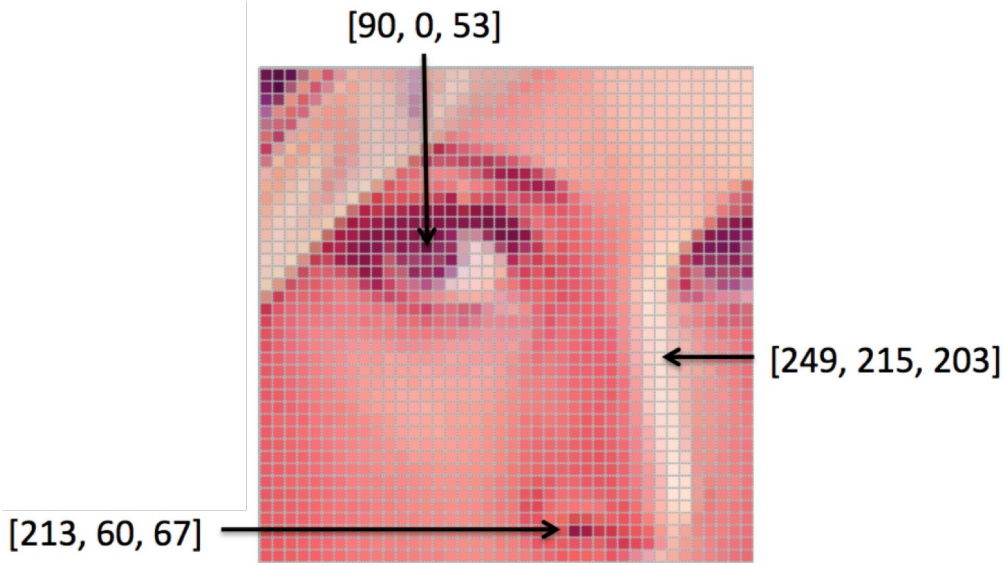




ferro@cimat.mx

Clubes de Ciencia México

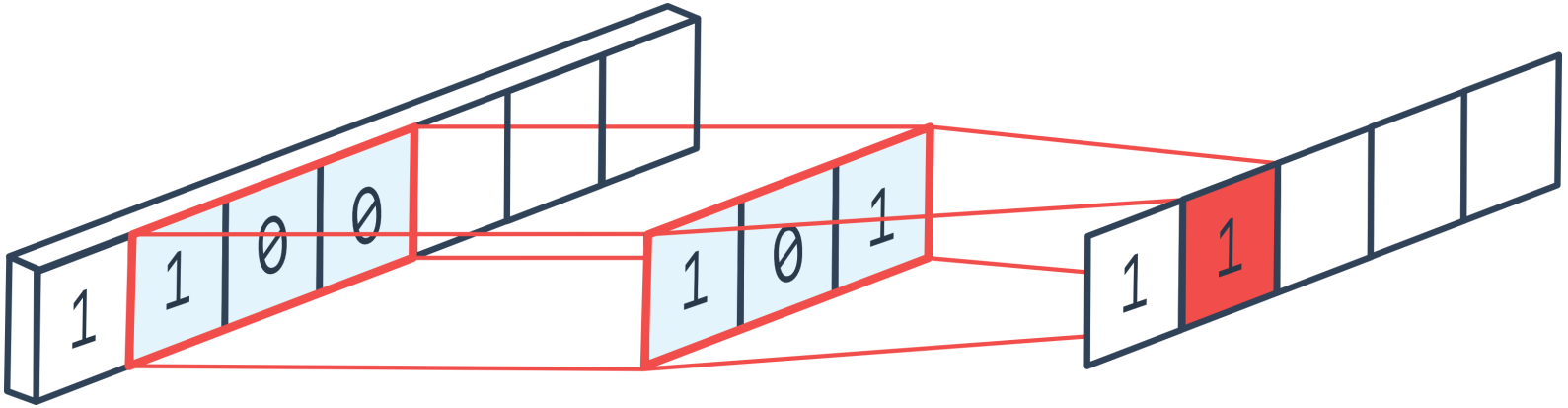
Ejercicio: Introducción a imágenes



ferro@cimat.mx

¿Qué es una convolución?

Visión computacional profunda



ferro@cimat.mx

Clubes de Ciencia México

¿Qué es una convolución?

Visión computacional profunda

$$
\begin{array}{ccccccc}
0 & 1 & 1 & 1_{\times 1} & 0_{\times 0} & 0_{\times 1} & 0 \\
0 & 0 & 1 & 1_{\times 0} & 1_{\times 1} & 0_{\times 0} & 0 \\
0 & 0 & 0 & 1_{\times 1} & 1_{\times 0} & 1_{\times 1} & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0
\end{array}
\quad * \quad
\begin{array}{ccc}
1 & 0 & 1 \\
0 & 1 & 0 \\
1 & 0 & 1
\end{array}
\quad = \quad
\begin{array}{ccccc}
1 & 4 & 3 & 4 & 1 \\
1 & 2 & 4 & 3 & 3 \\
1 & 2 & 3 & 4 & 1 \\
1 & 3 & 3 & 1 & 1 \\
3 & 3 & 1 & 1 & 0
\end{array}
$$

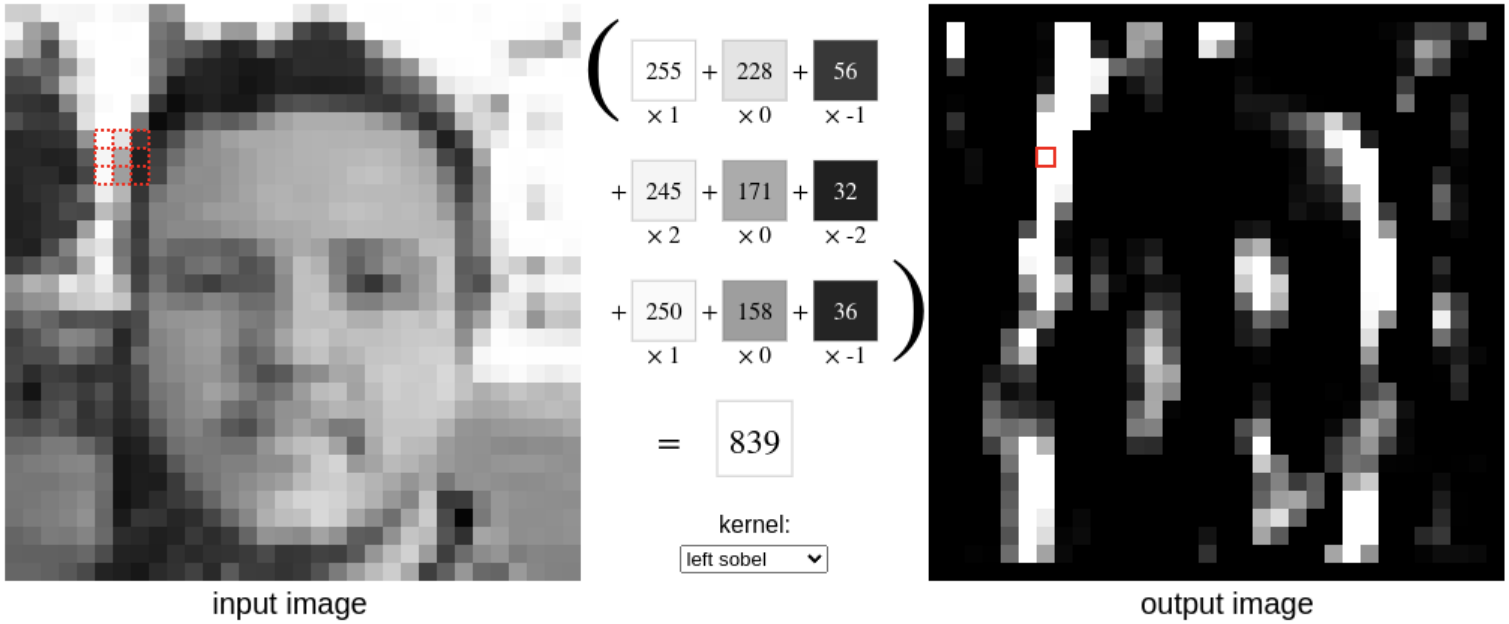
$$\mathbf{I} \qquad\qquad \mathbf{K} \qquad\qquad \mathbf{I} * \mathbf{K}$$

ferro@cimat.mx

Clubes de Ciencia México

¿Qué es una convolución?

Visión computacional profunda



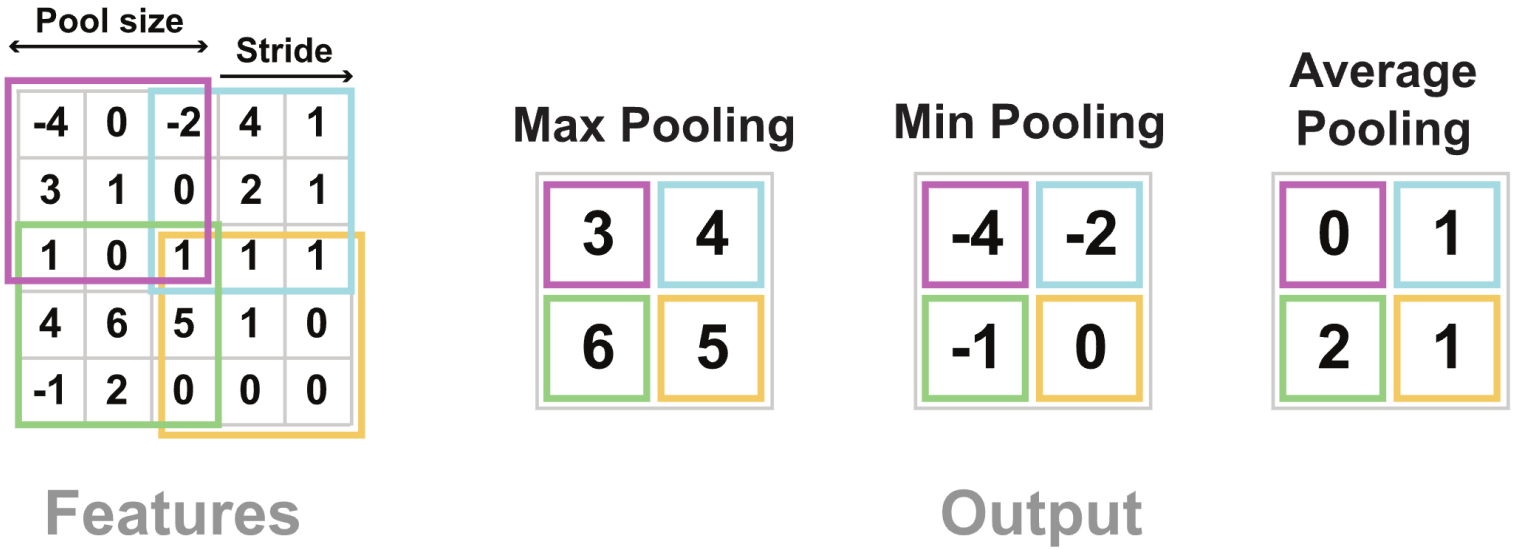
ferro@cimat.mx

Clubes de Ciencia México

"Image Kernels." Victor Powell. https://setosa.io/ev/image-kernels/

Pooling

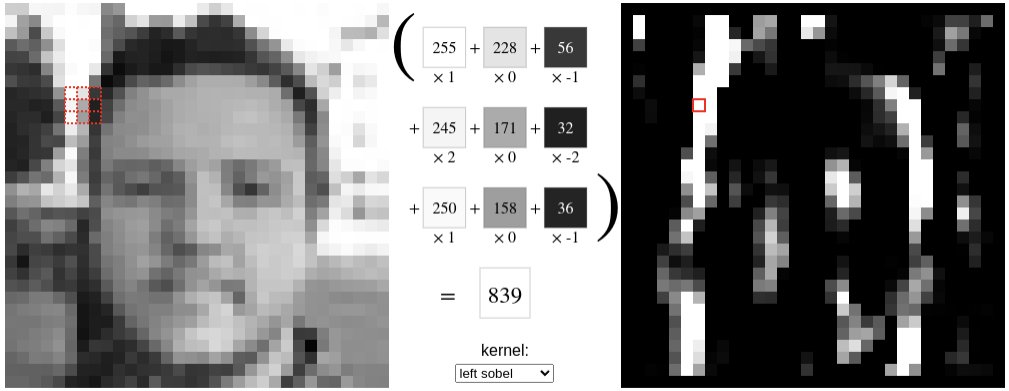
Visión computacional profunda



ferro@cimat.mx

Clubes de Ciencia México

Ejercicio: Convoluciones & Pooling



input image

output image

ferro@cimat.mx

Redes neuronales convolucionales

Visión computacional profunda

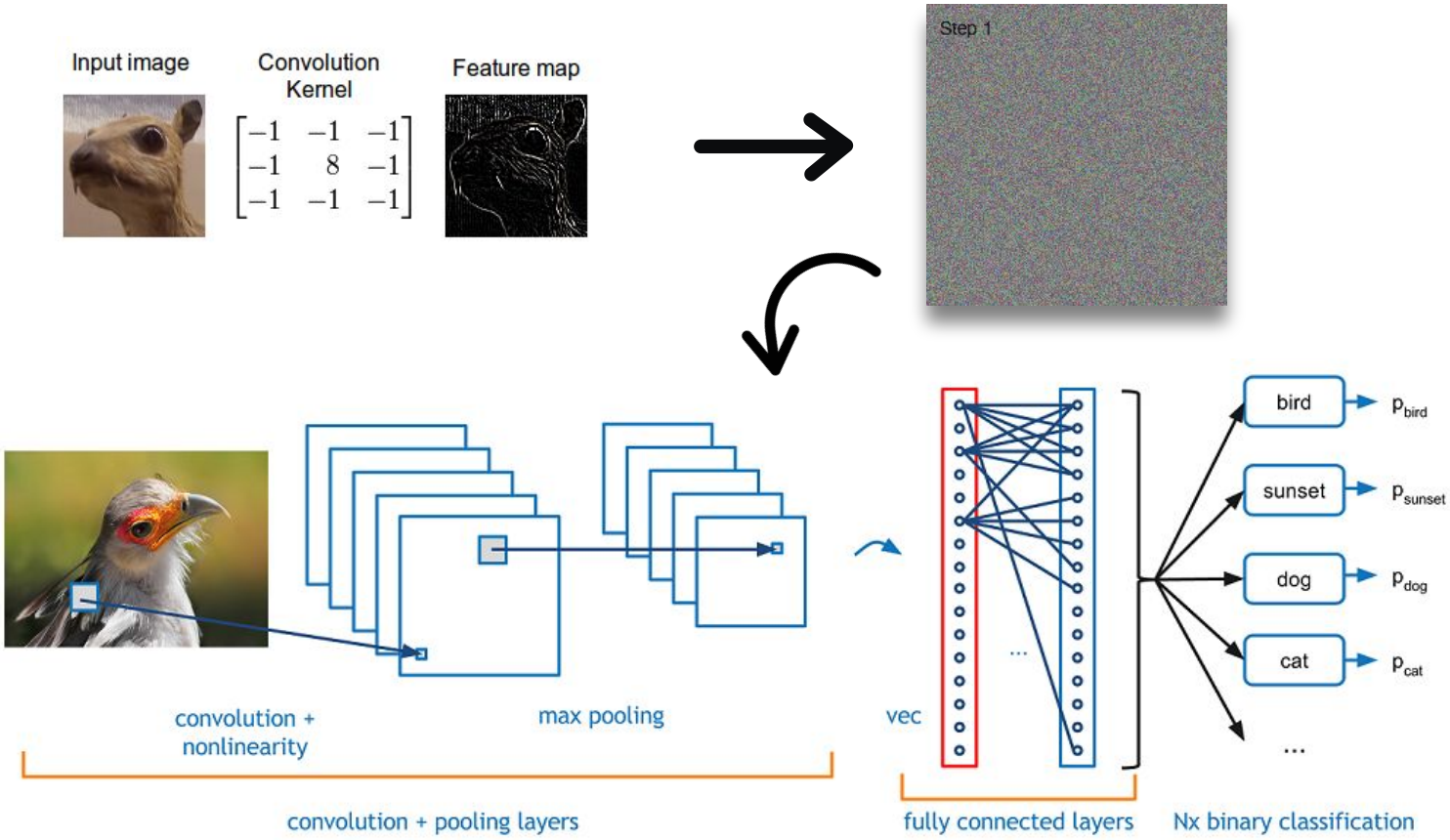
Figure: Yann LeCun

ferro@cimat.mx

Clubes de Ciencia México

Redes neuronales convolucionales

Visión computacional profunda

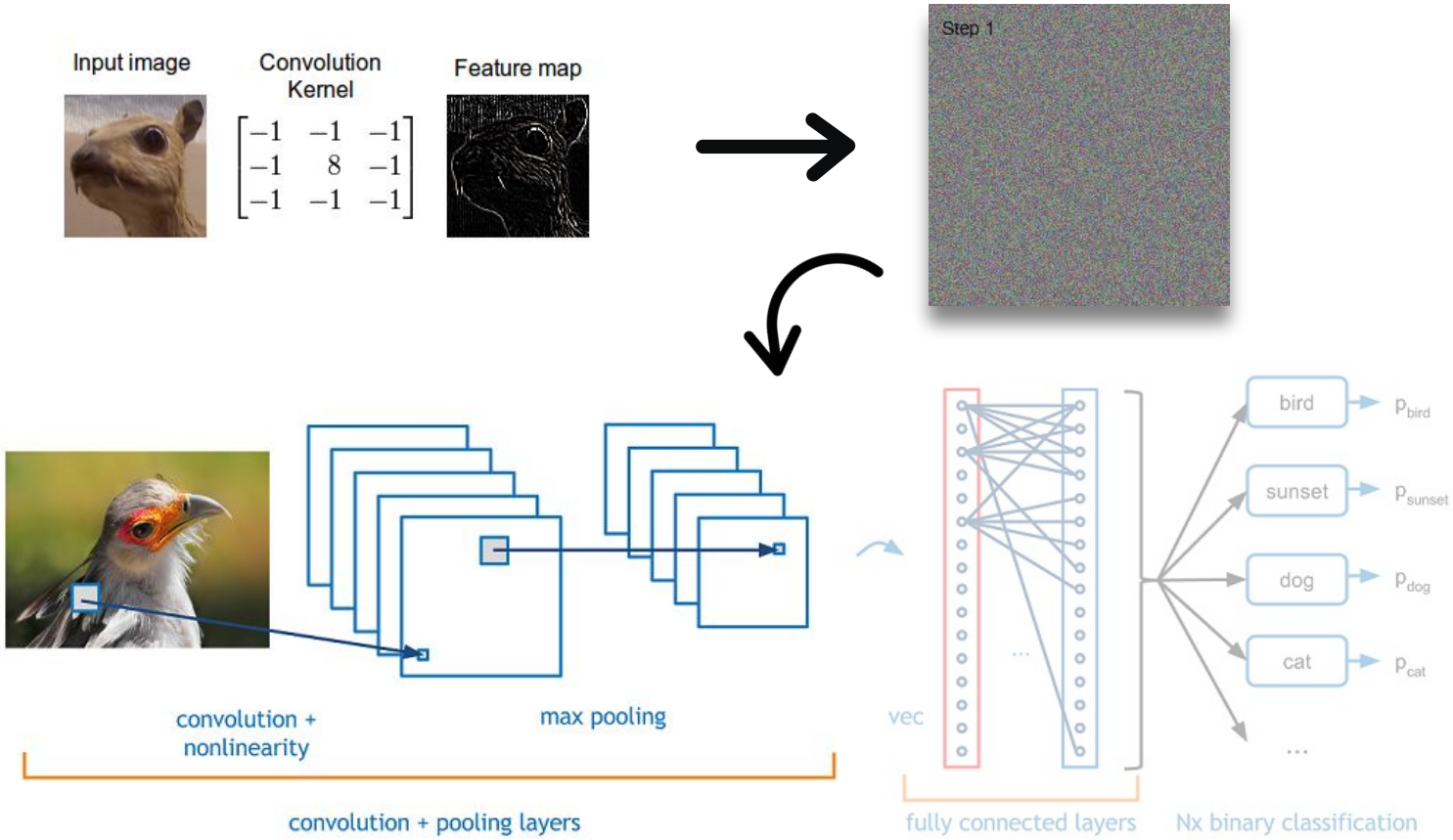


ferro@cimat.mx

Clubes de Ciencia México

Redes neuronales convolucionales

Visión computacional profunda

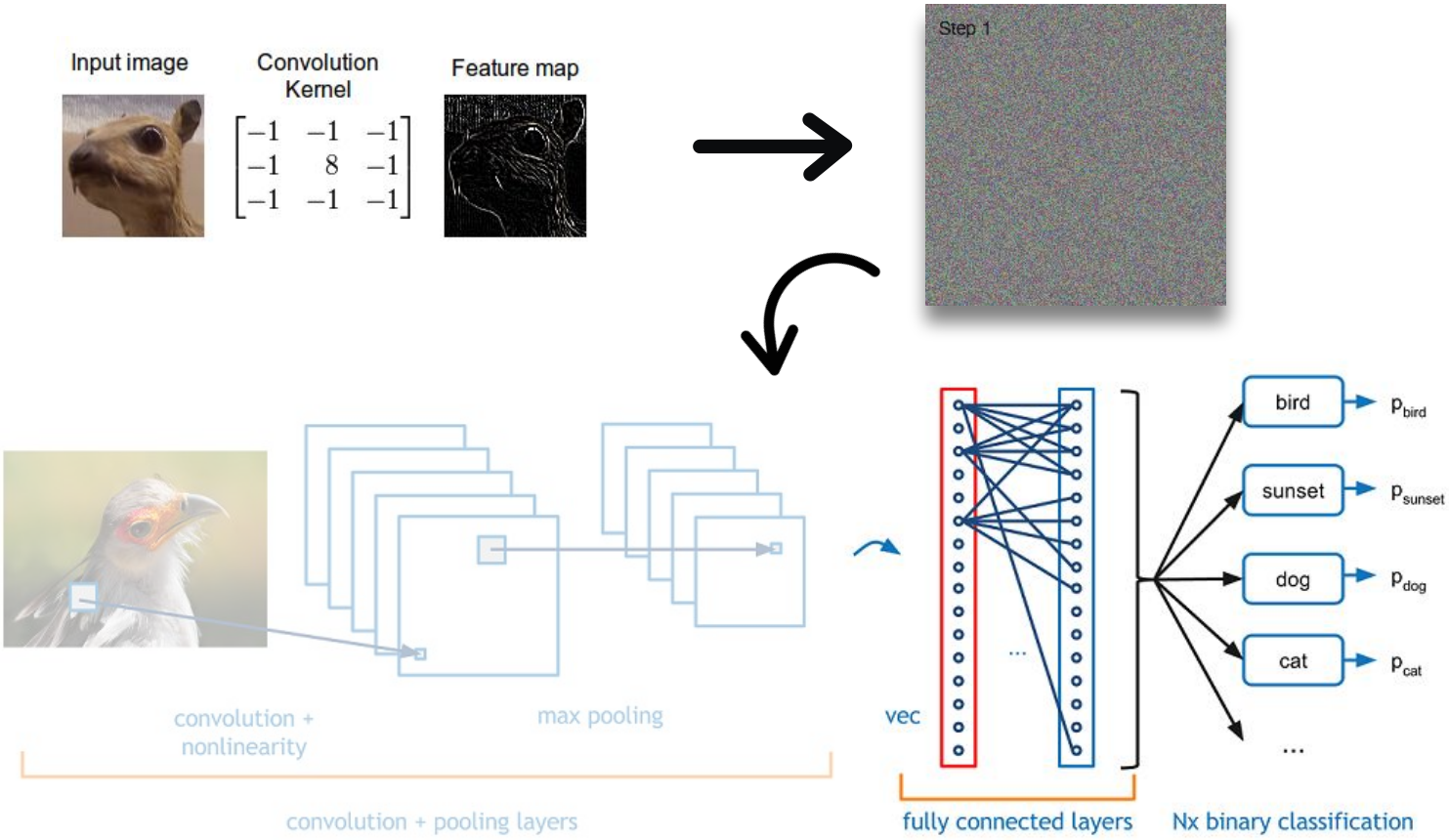


ferro@cimat.mx

Clubes de Ciencia México

Redes neuronales convolucionales

Visión computacional profunda



ferro@cimat.mx

Clubes de Ciencia México

LeNet-5

Visión computacional profunda

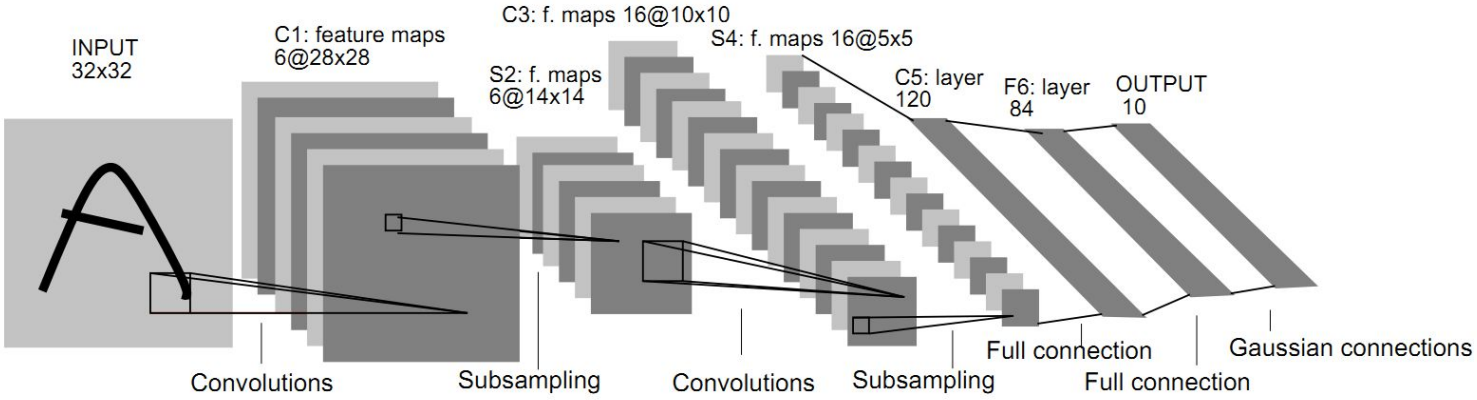
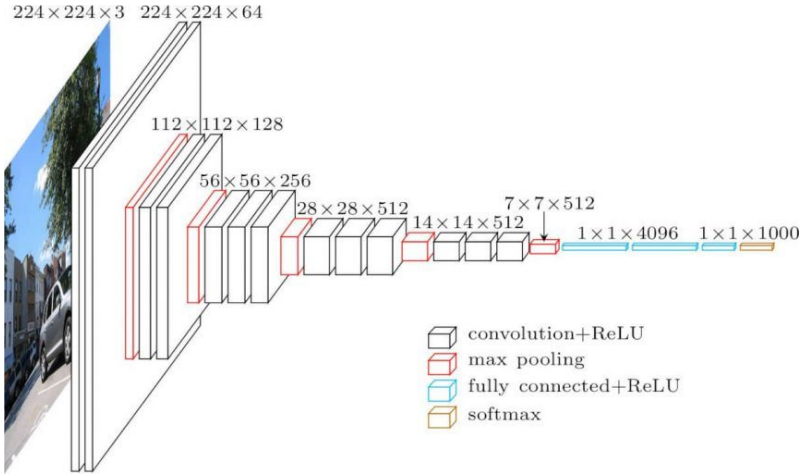


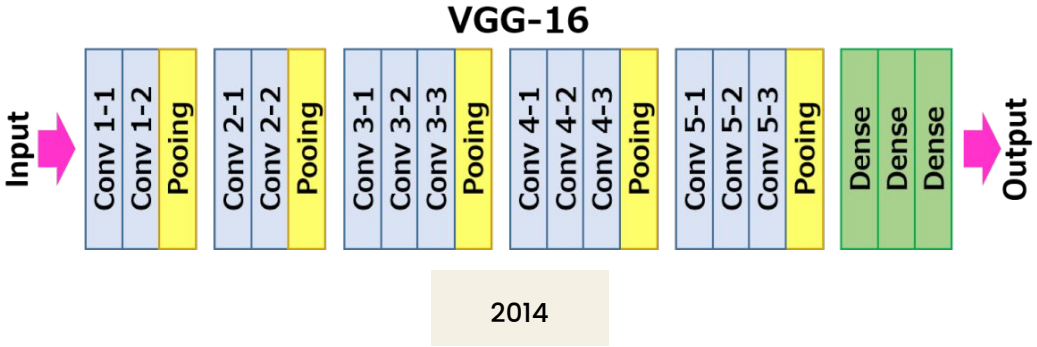
Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

1998

VGG16

Visión computacional profunda



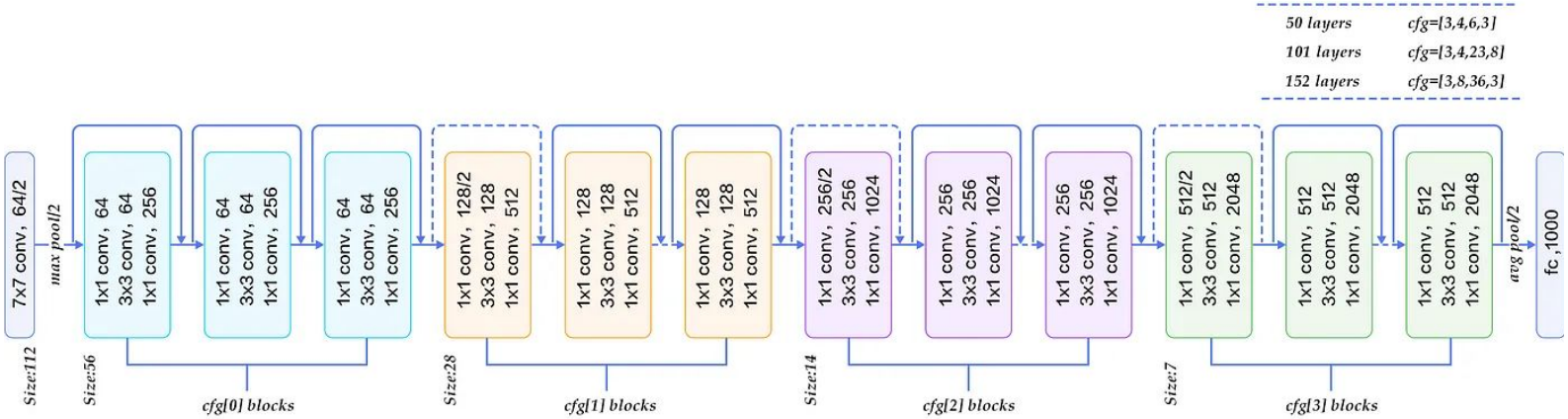


ferro@cimat.mx

Clubes de Ciencia México

Resnet50

Visión computacional profunda



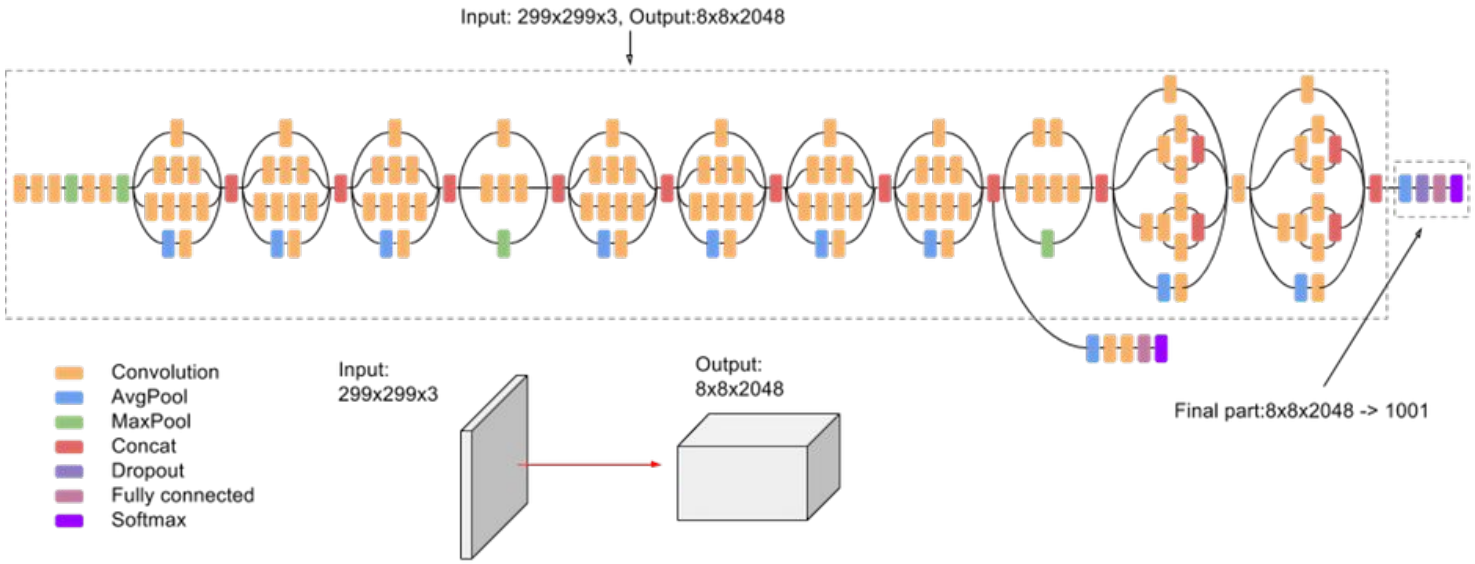
2015

ferro@cimat.mx

Clubes de Ciencia México

GoogLeNet

Visión computacional profunda



2015

ferro@cimat.mx

Clubes de Ciencia México

Ejercicio: Redes neuronales convolucionales

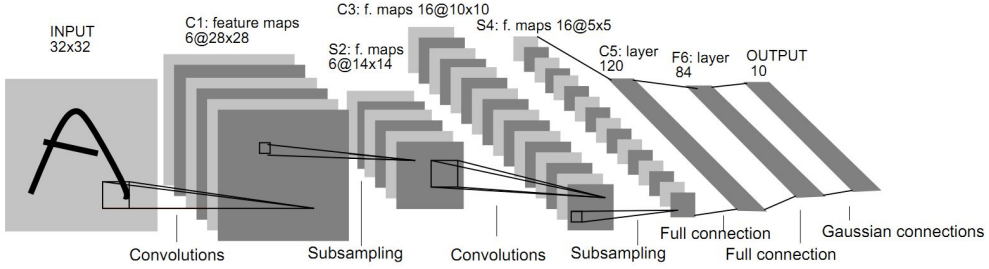


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

1998

Lecturas recomendadas

Visión computacional

- Tutorial 1: Image Filtering
- Image Kernels Explained Visually by Victor Powell
- Parameterized Pooling Layers by Hao Hao Tan
- TensorFlow Tutorials: Vision

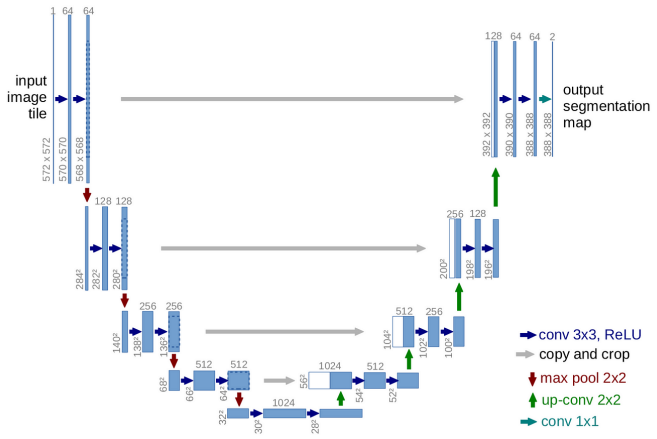
ferro@cimat.mx

Clubes de Ciencia México

Trabajos relacionados y avances recientes

Han habido varios trabajos de investigación y avances recientes que han contribuido al desarrollo de nuevas arquitecturas, técnicas de entrenamiento mejoradas y aplicaciones emergentes.

- **UNet:** Es ampliamente utilizada en el campo de la segmentación de imágenes, pero también se ha aplicado con éxito en tareas de denoising.



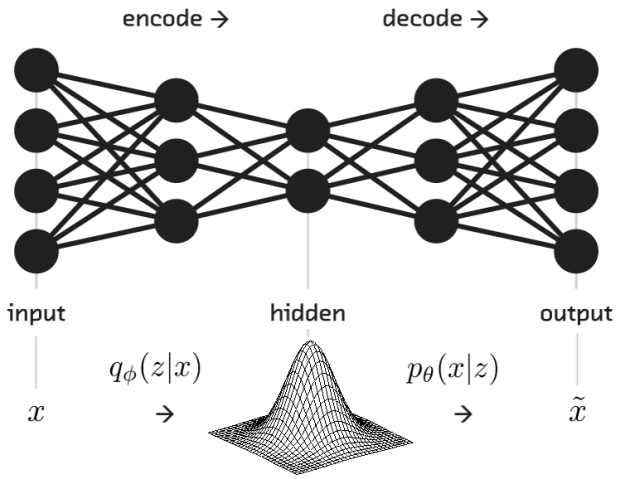
ferro@cimat.mx

Clubes de Ciencia México

Trabajos relacionados y avances recientes

Visión computacional profunda

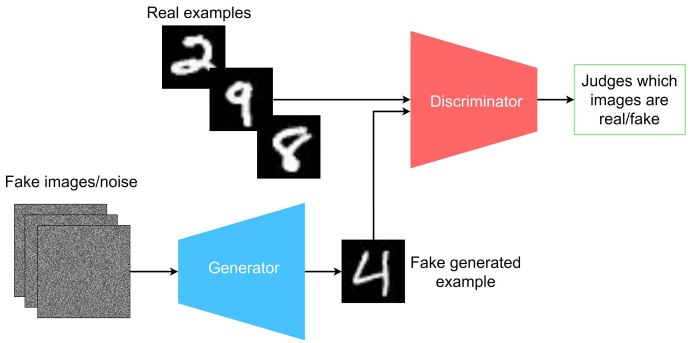
- **Variational Autoencoders (VAEs):** Los VAEs son una variante de los autoencoders que se utilizan para el aprendizaje de distribuciones latentes. Han demostrado ser efectivos en el denoising de imágenes al aprender representaciones latentes que siguen una distribución probabilística, lo que permite una generación más controlada y realista de imágenes limpias.



Trabajos relacionados y avances recientes

Visión computacional profunda

- **Generative Adversarial Networks (GANs):** Estos modelos aprovechan la capacidad de los GANs para generar imágenes realistas y para aprender representaciones latentes eficientes. Los GANs han demostrado ser efectivos en el denoising y la generación de imágenes de alta calidad, entre otros.



ferro@cimat.mx

Clubes de Ciencia México

Tareas en el campo de visión artificial

- **Clasificación de imágenes:** La tarea de clasificación de imágenes implica asignar una etiqueta o categoría a una imagen de entrada. Esto implica entrenar un modelo para reconocer y distinguir diferentes objetos, personas o escenas en una imagen.
- **Detección de objetos:** La detección de objetos implica localizar y clasificar múltiples objetos en una imagen. El objetivo es detectar la presencia y la ubicación de objetos específicos en una escena, a menudo utilizando cuadros delimitadores para delinear las regiones donde se encuentran los objetos.
- **Denoising o reconstrucción de imágenes:** Consiste en eliminar o reducir el ruido presente en una imagen, obteniendo una versión más limpia y clara. Esta tarea es relevante en áreas como la fotografía, la medicina y la seguridad.

ferro@cimat.mx

Tareas en el campo de visión artificial

- **Segmentación semántica:** La segmentación semántica implica asignar una etiqueta a cada píxel de una imagen para identificar y delimitar las diferentes regiones o objetos presentes. El objetivo es comprender la estructura y el contenido de una imagen a nivel de píxel.
- **Detección de rostros:** La detección de rostros es una tarea específica de la visión artificial que implica detectar y localizar los rostros en una imagen. Es ampliamente utilizado en aplicaciones de reconocimiento facial, análisis de emociones y sistemas de seguridad.
- **Reconocimiento y verificación facial:** El reconocimiento facial se refiere a la tarea de identificar y reconocer a una persona específica a partir de una imagen o secuencia de imágenes. La verificación facial se enfoca en verificar si una imagen de rostro coincide con una identidad específica.

Tareas en el campo de visión artificial

- **Estimación de pose:** La estimación de pose se refiere a la tarea de determinar la posición y orientación de un objeto o persona en una imagen. Esto implica detectar y rastrear las articulaciones o puntos clave en una imagen para comprender la postura y el movimiento.
- **Estimación de profundidad:** La estimación de profundidad implica inferir la información de la distancia o la profundidad de los objetos en una imagen. Es útil en aplicaciones de realidad virtual, conducción autónoma y sistemas de navegación.
- **Super-resolución:** La super-resolución se refiere a aumentar la resolución o la calidad de una imagen de baja resolución. El objetivo es generar una versión de alta resolución que capture más detalles y claridad.

ferro@cimat.mx

Lecturas recomendadas

Visión computacional

- From Autoencoder to Beta-VAE
- Building Autoencoders in Keras
- Autoencoders: explicación y tutorial en Python
- Autoencoder For Denoising Images
- Medical image denoising using convolutional denoising autoencoders

ferro@cimat.mx

Clubes de Ciencia México

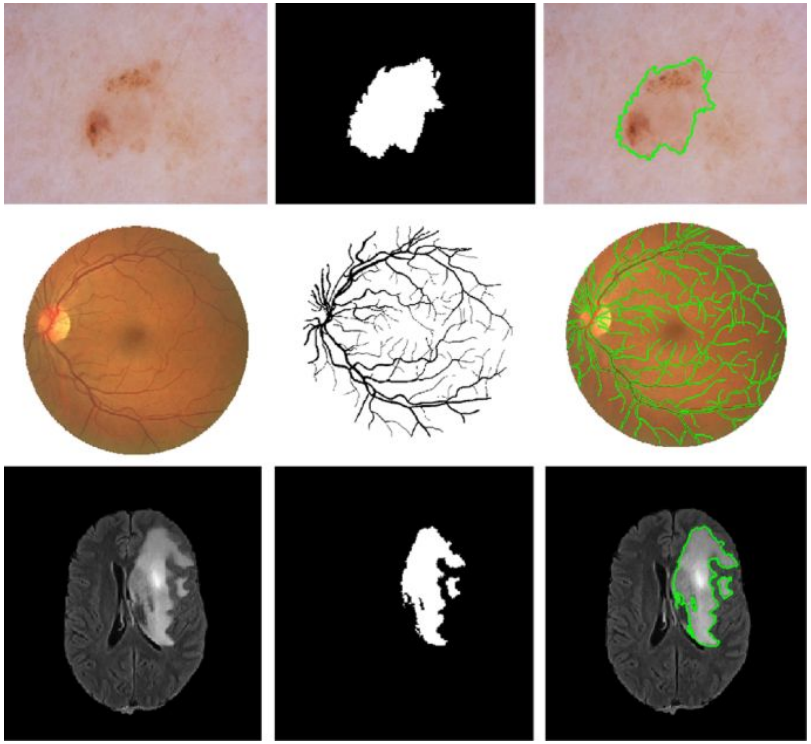
Segmentación de imágenes médicas

ferro@cimat.mx

Olaf
Ronneberger
et al.,
2015

ferro@cimat.mx

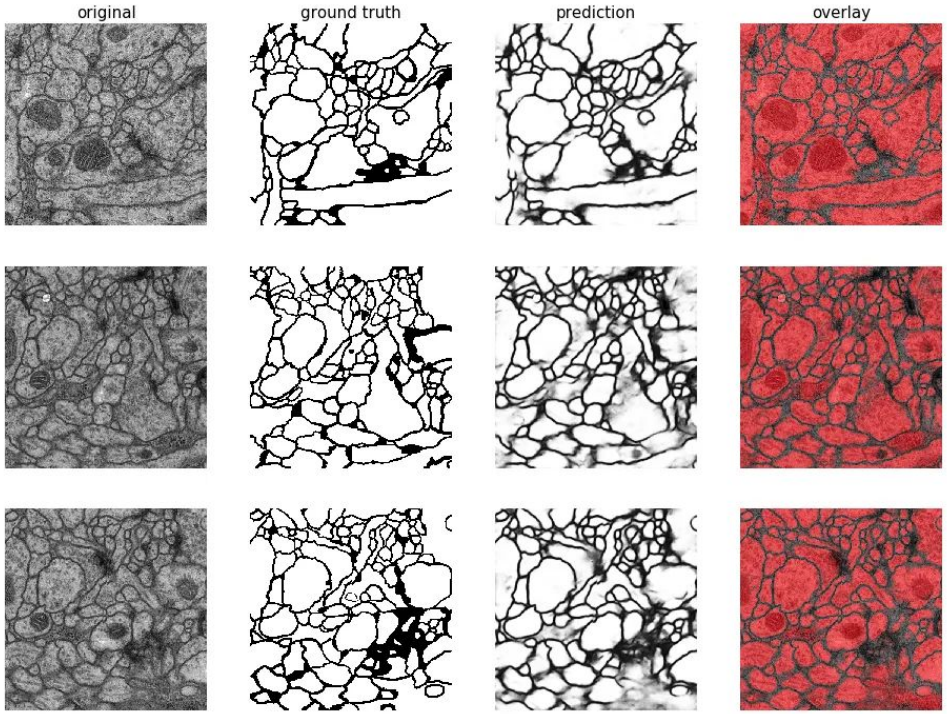
Segmentación de imágenes



https://www.nature.com/articles/s41598-021-89686-3

ferro@cimat.mx

Segmentación con U-Net



https://medium.com/@venkateshtata9/semantic-segmentation-on-medical-images-3ba8264cda5e

Segmentación con U-Net

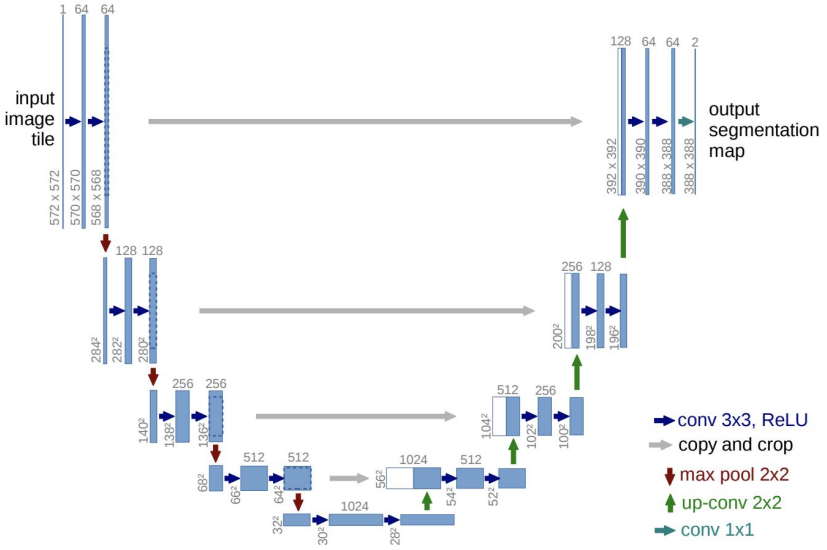
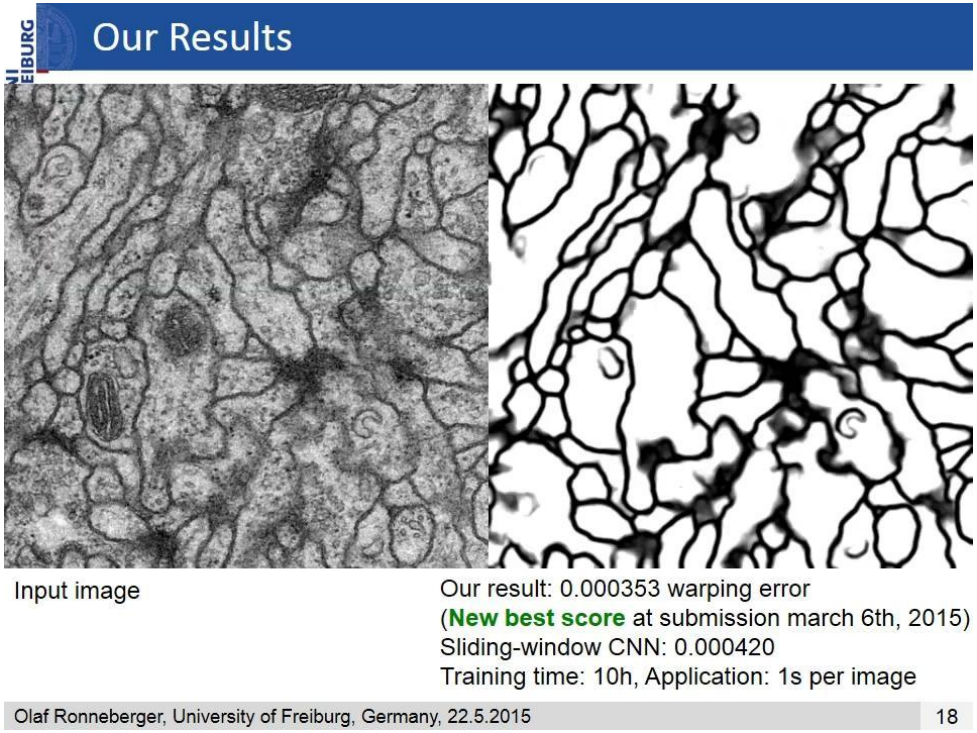


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

ferro@cimat.mx

Segmentación con U-Net



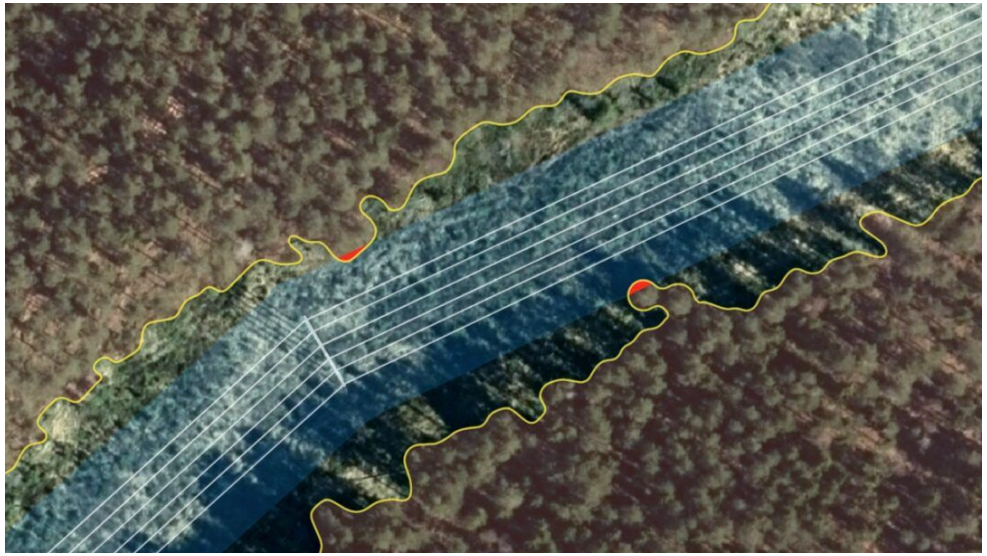
https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

ferro@cimat.mx

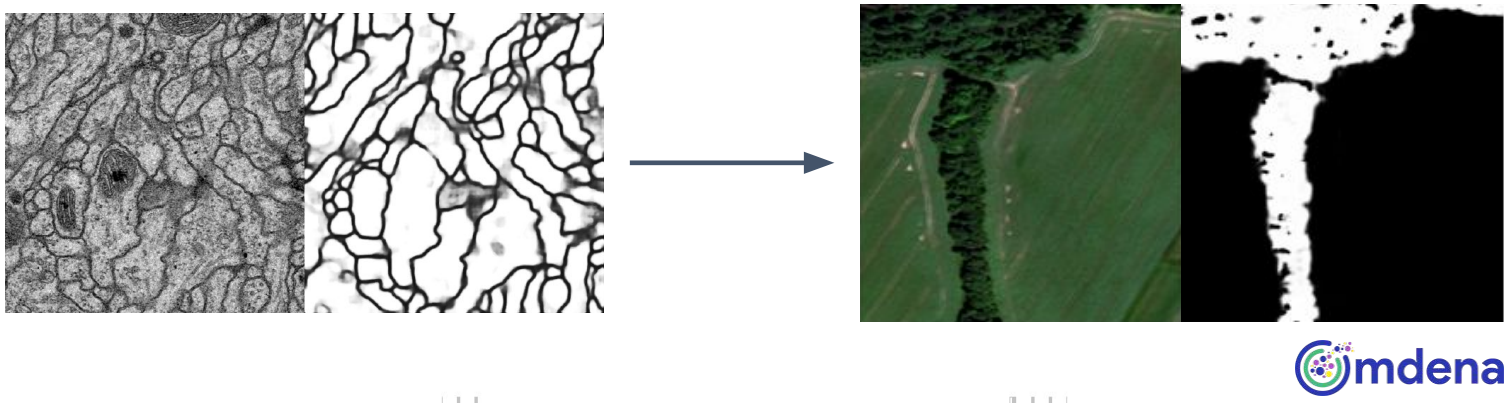
Caso de estudio

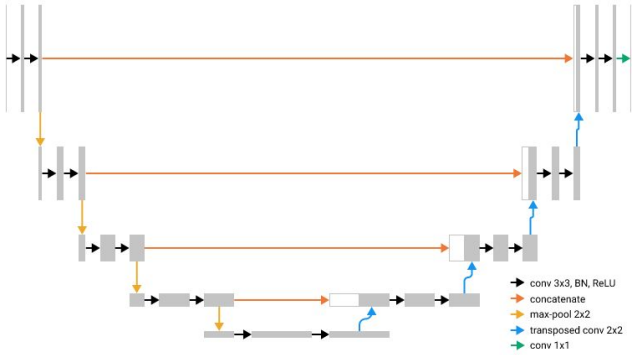
ferro@cimat.mx

Prevención de incendios



https://omdena.com/projects/ai-prevent-forest-fires/





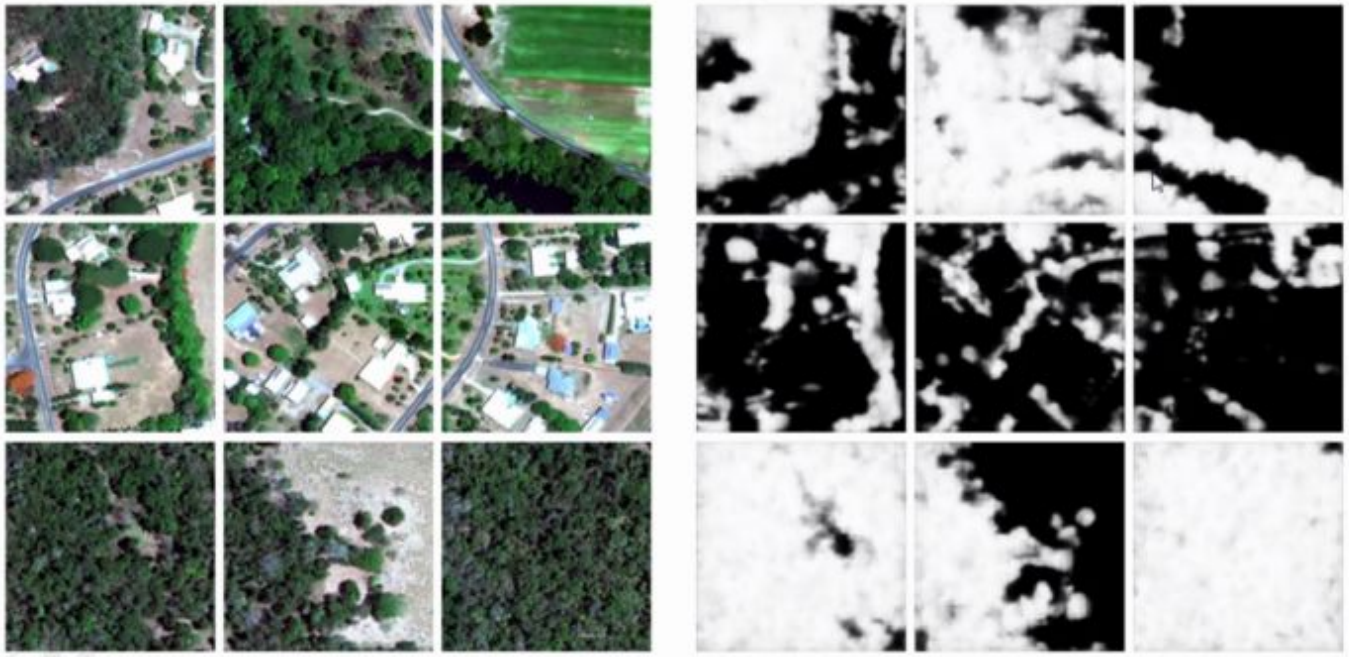
U-Net: Convolutional Networks for Biomedical Image Segmentation.

ferro@cimat.mx

Prevención de incendios

- **Omdena** + **Spacept**, una startup sueca
- 36 colaborador@s a nivel global (al menos 3 mexicanos)
- Dataset de 200 imágenes satelitales (Australia)
- Desarrollamos un modelo con 95% de precisión
- Desarrollamos un módulo que preprocesa imágenes
- **¡Resolvimos el challenge!**

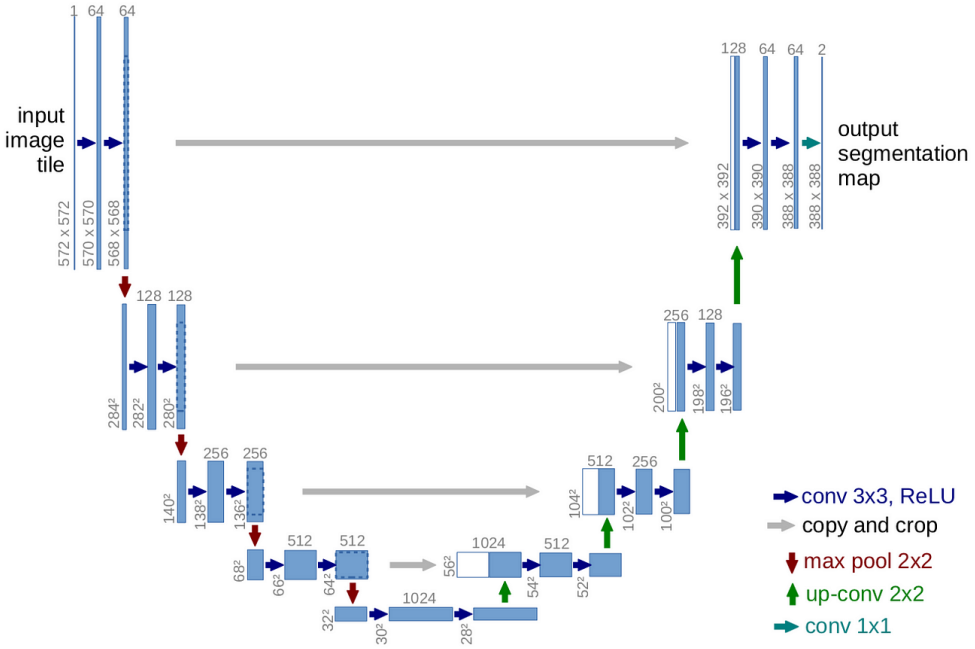
https://omdena.com/projects/ai-prevent-forest-fires/



https://omdena.com/projects/ai-prevent-forest-fires/

ferro@cimat.mx

Ejercicio: Entrenamiento de U-Net



ferro@cimat.mx