



**Pontificia Universidad
Católica del Ecuador**
Seréis mis testigos

**FACULTAD DE
CIENCIAS EXACTAS Y NATURALES**

PROCESAMIENTO Y ANÁLISIS DE IMÁGENES

Andrés Merino

Septiembre 2024

Escuela de Ciencias Físicas y Matemática

Escuela de Verano de Aprendizaje Automático PUCE 2024

CONTENDIO

1. Conceptos básicos
2. ¿Qué es una imagen?
3. Convolución
4. Redes convolucionales
5. Arquitecturas de CNN





**Pontificia Universidad
Católica del Ecuador**
Seréis mis testigos

FACULTAD DE
CIENCIAS EXACTAS Y NATURALES

CONCEPTOS BÁSICOS



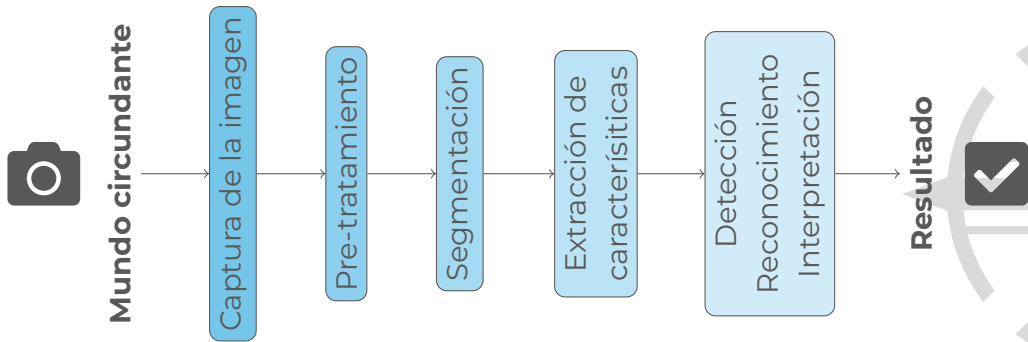
DEFINICIÓN DE PROCESAMIENTO Y ANÁLISIS DE IMÁGENES

Procesamiento de Imágenes:

- Ingresa una imagen y sale una imagen.
- Ejemplo: Segmentación.

Análisis de Imágenes:

- Ingresa una imagen y sale un resultado.
- Ejemplo: Clasificación.





**Pontificia Universidad
Católica del Ecuador**
Seréis mis testigos

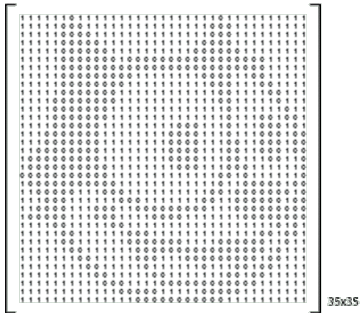
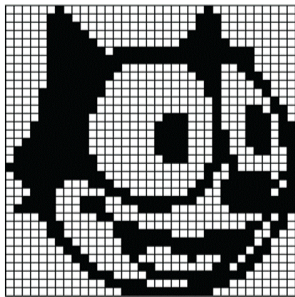
FACULTAD DE
CIENCIAS EXACTAS Y NATURALES

¿QUÉ ES UNA IMAGEN?



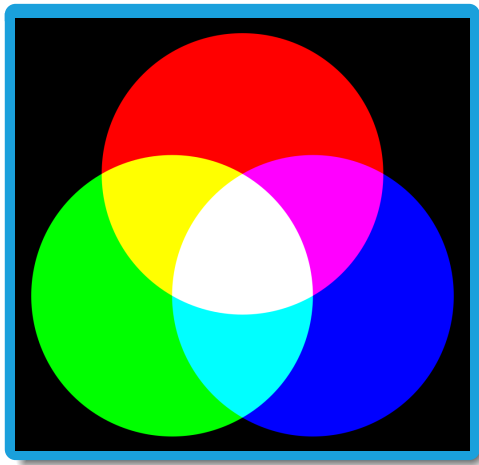
¿QUÉ ES UNA IMAGEN?

Una imagen digital se representa como una matriz de valores, donde cada elemento (píxel) de la matriz almacena la intensidad de luz en un punto específico de la imagen.



EL FORMATO RGB

- Modelo de color basado en la síntesis aditiva
- Es posible representar un color mediante la mezcla por adición de los tres colores de luz primarios: **rojo**, **verde** o **azul**.
- Tiene rangos de 0 a 255 (enteros) o de 0 a 1 (reales).

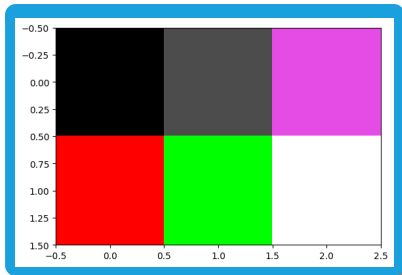


EJEMPLO

$$R = \begin{pmatrix} 0,0 & 0,3 & 0,9 \\ 1,0 & 0,0 & 1,0 \end{pmatrix}; \quad G = \begin{pmatrix} 0,0 & 0,3 & 0,3 \\ 0,0 & 1,0 & 1,0 \end{pmatrix}; \quad B = \begin{pmatrix} 0,0 & 0,3 & 0,9 \\ 0,0 & 0,0 & 1,0 \end{pmatrix}.$$

EJEMPLO

$$R = \begin{pmatrix} 0,0 & 0,3 & 0,9 \\ 1,0 & 0,0 & 1,0 \end{pmatrix}; \quad G = \begin{pmatrix} 0,0 & 0,3 & 0,3 \\ 0,0 & 1,0 & 1,0 \end{pmatrix}; \quad B = \begin{pmatrix} 0,0 & 0,3 & 0,9 \\ 0,0 & 0,0 & 1,0 \end{pmatrix}.$$



REDUCCIÓN Y AMPLIACIÓN DE IMÁGENES

Reducción de Imágenes:

- Proceso para disminuir el tamaño de una imagen.
- Puede provocar la pérdida de información al disminuir la cantidad de píxeles.

Ampliación de Imágenes:

- Proceso para aumentar el tamaño de una imagen.
- Generan píxeles adicionales que no estaban en la imagen original.

REDUCCIÓN DE IMÁGENES

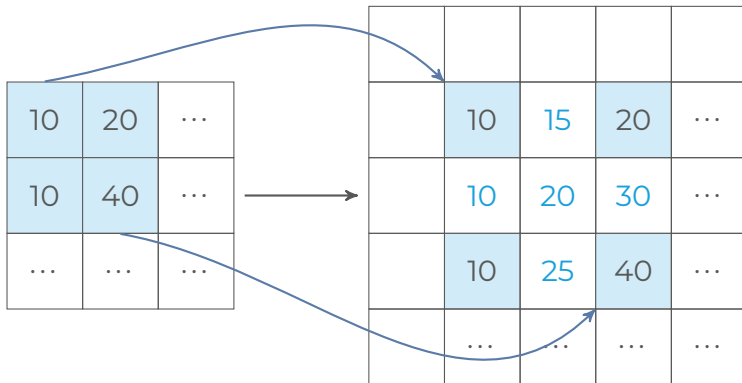
10	13	20	...
10	25	35	...
10	25	40	...
...

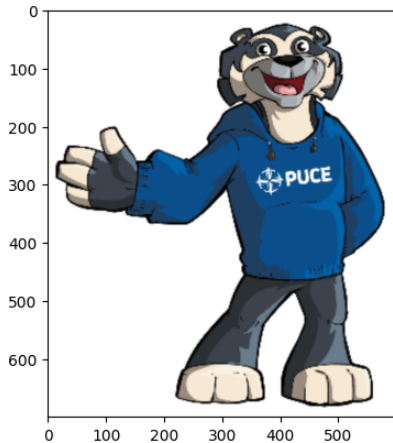
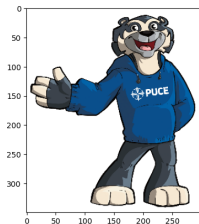
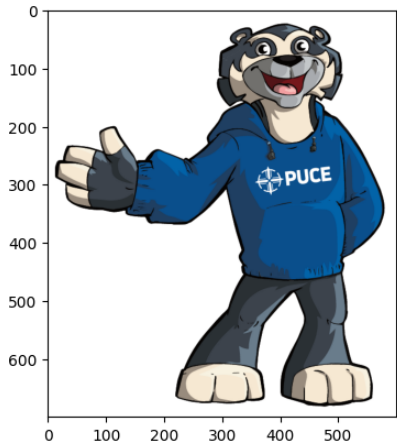


10	20	...
10	40	...
...

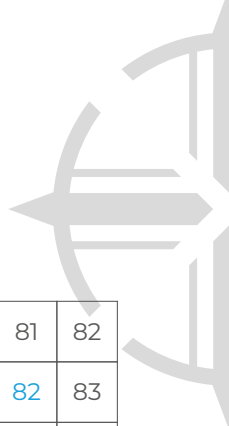
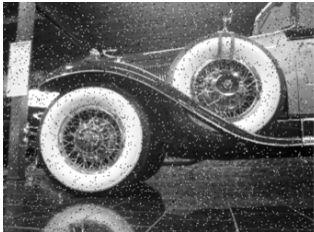


AMPLIACIÓN DE IMÁGENES





MEJORAR IMÁGENES



82	81	82
81	200	83
80	83	84



Lista ordenada

80	81	81	82	82	83	83	84	200
----	----	----	----	----	----	----	----	-----



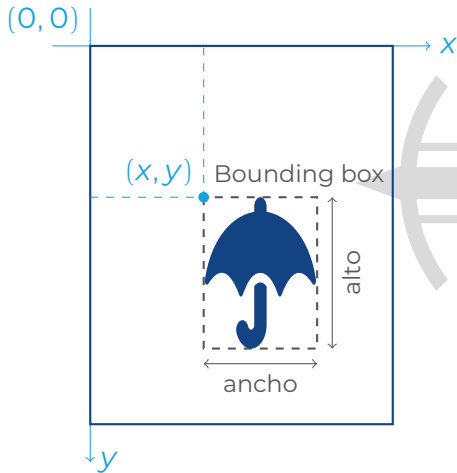
valor seleccionado

82	81	82
81	82	83
80	83	84

UN PAR MÁS DE CONCEPTOS

Bounding Box

- Delimita un objeto dentro de una imagen utilizando un rectángulo.
- Proporciona coordenadas (x, y) y dimensiones (ancho, alto) del rectángulo.



UN PAR MÁS DE CONCEPTOS

Segmentación Semántica

- Asigna una etiqueta a cada píxel de la imagen, clasificándolos según el tipo de objeto al que pertenecen.
- Proporciona una delimitación precisa de los contornos de los objetos.





**Pontificia Universidad
Católica del Ecuador**
Seréis mis testigos

FACULTAD DE
CIENCIAS EXACTAS Y NATURALES

CONVOLUCIÓN



Convolución

La convolución entre dos funciones f y g se define como:

$$s(t) = (f * g)(t) = \int_{-\infty}^{+\infty} f(\tau)g(t - \tau) d\tau$$

- f es la función de entrada.
- g es el filtro o kernel.

Convolución Discreta

Para datos discretos, la convolución se define como:

$$s(t) = (f * g)(t) = \sum_{\tau=-\infty}^{\infty} f(\tau)g(t - \tau)$$

Convolución en Dos Dimensiones (Imágenes)

La convolución entre una imagen I y un kernel K es:

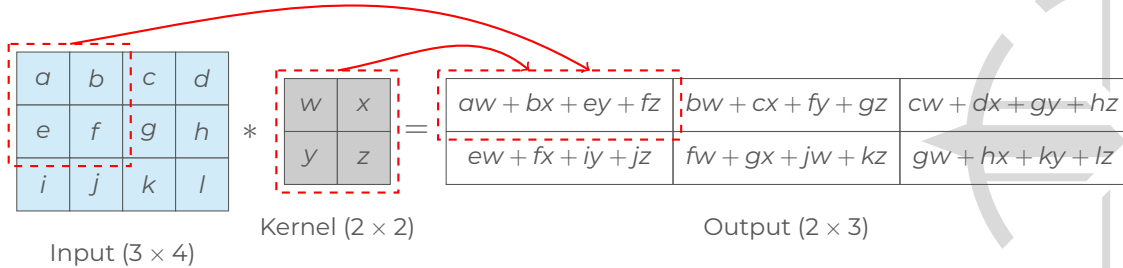
$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n)$$

Correlación Cruzada

La correlación cruzada entre una imagen I y un kernel K es:

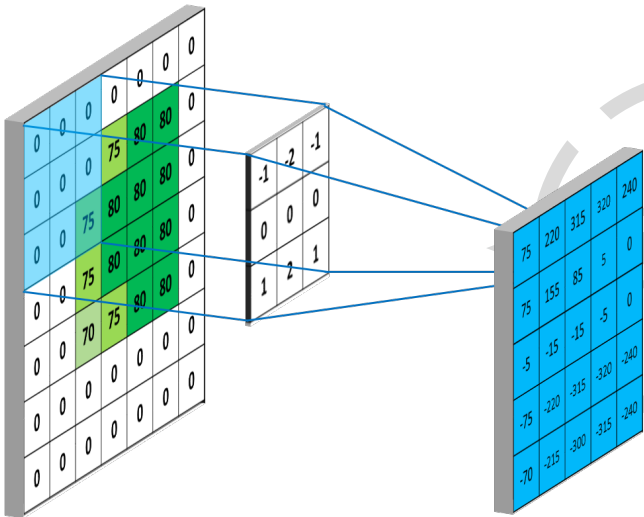
$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n)$$

- En la práctica, se usa correlación cruzada en lugar de convolución.
- La correlación cruzada evita invertir el kernel, lo que simplifica la implementación



FILTROS

Ver animación.



La dimensión de salida después de aplicar una convolución es:

$$(l_1 - K_1 + 1, l_2 - K_2 + 1),$$

dónde

- l_1, l_2 son las dimensiones de la imagen de entrada, y
- K_1, K_2 son las dimensiones del kernel



**Pontificia Universidad
Católica del Ecuador**
Seréis mis testigos

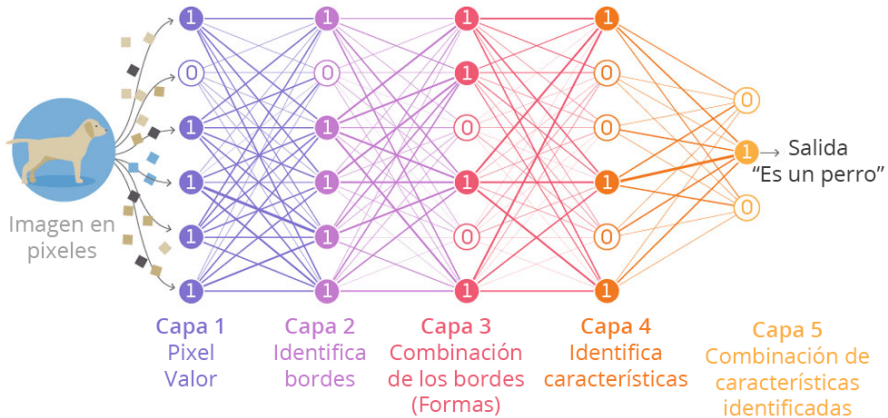
FACULTAD DE
CIENCIAS EXACTAS Y NATURALES

REDES CONVOLUCIONALES



REDES CONVOLUCIONALES

- Se inspiran en la corteza visual de los animales.
- Desarrollo desde Neocognitron (Fukushima, 1982) a LeNet-5 (LeCun, 1989).
- Avance impulsado por hardware mejorado y algoritmos optimizados a partir de 2012.
- CNNs aplican filtros convolucionales para extraer patrones como bordes, texturas, y formas a diferentes niveles de abstracción.



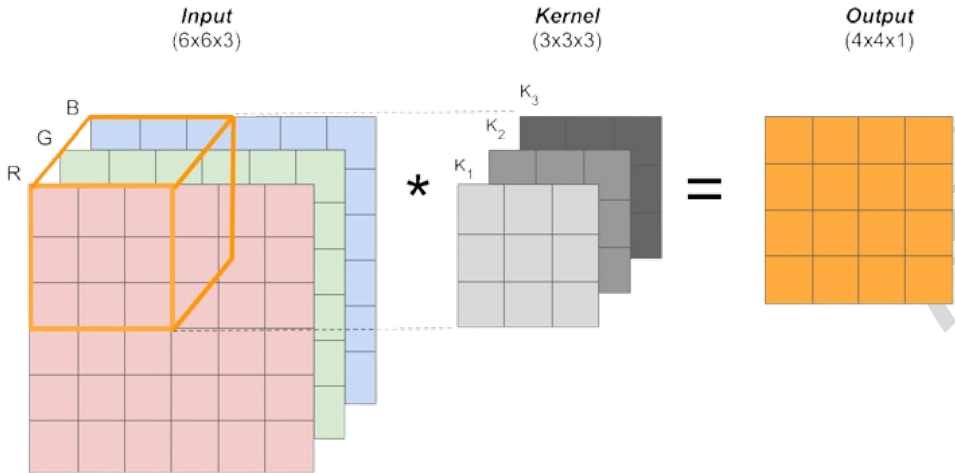
Fuente: <https://www.quantamagazine.org/>

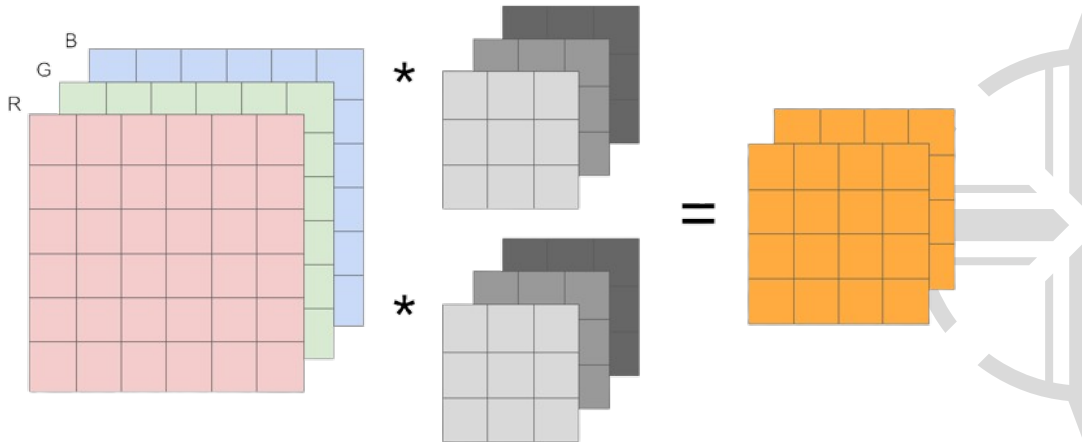
CAPA DE CONVOLUCIÓN

- Se hace mediante la aplicación de Filtros.
- El filtro es la matriz de pesos de una capa de convolución dentro de una CNN.
- Si se tiene 3 canales, se aplica un filtro por canal y se suman los resultados de la siguiente manera:

$$R * K_1 + G * K_2 + B * K_3.$$

- El kernel siempre tiene el mismo número de canales que la entrada.





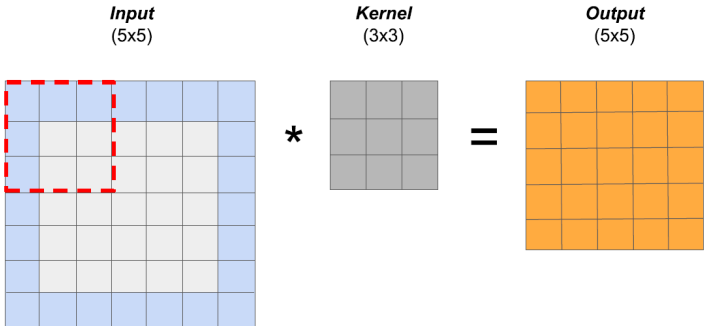
DIMENSIONES

- Entrada: (n_w, n_h, n_c) .
- Kernel: (n_{k_1}, n_{k_2}, n_c) .
- Número de kernels: n .
- Salida:

$$(n_w - n_{k_1} + 1, n_h - n_{k_2} + 1, n).$$

PADDING (RELLENO)

Evita la reducción del tamaño de la salida agregando ceros alrededor de la entrada.



DIMENSIONES

- Entrada: (n_w, n_h) .
- Kernel: (n_k, n_k) .
- Valor de relleno: p .
- Salida:

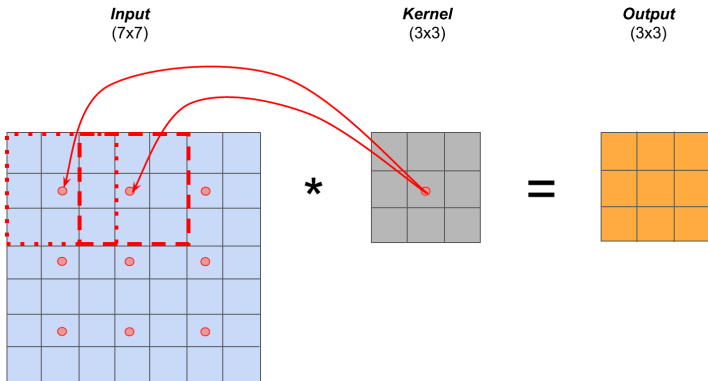
$$(n_w - n_k + 2p + 1, n_h - n_k + 2p + 1)$$

Para mantener el tamaño de la entrada, el padding p se calcula como:

$$p = \left\lfloor \frac{n_k - 1}{2} \right\rfloor$$

CONVOLUCIONES POR PASOS (STRIDE)

Aplican la operación de convolución saltando celdas en la entrada.



DIMENSIONES

- Entrada: (n_w, n_h) .
- Kernel: (n_k, n_k) .
- Valor de relleno: p .
- Paso o stride: s
- Salida:

$$\left(\left\lfloor \frac{n_w - n_k + 2p}{s} + 1 \right\rfloor, \left\lfloor \frac{n_h - n_k + 2p}{s} + 1 \right\rfloor \right)$$

Reduce la dimensión de la salida al aplicar un mayor paso en la convolución.

CAPA DE AGRUPAMIENTO (POOLING)

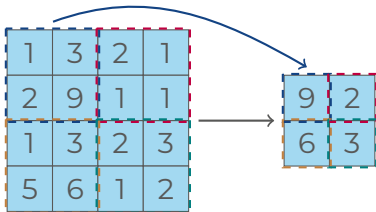
Reducen la dimensión de los datos de entrada.

Max-pooling

Valor máximo en una región de la entrada.

Average-pooling

Promedio de los valores en una región.



ESTRUCTURA DE UNA RED NEURONAL CONVOLUCIONAL (CNN)

- **Bloque convolucional:**

- Incluye capas de convolución, agrupamiento (pooling) y activación.
- Detecta patrones y características en los datos de entrada.

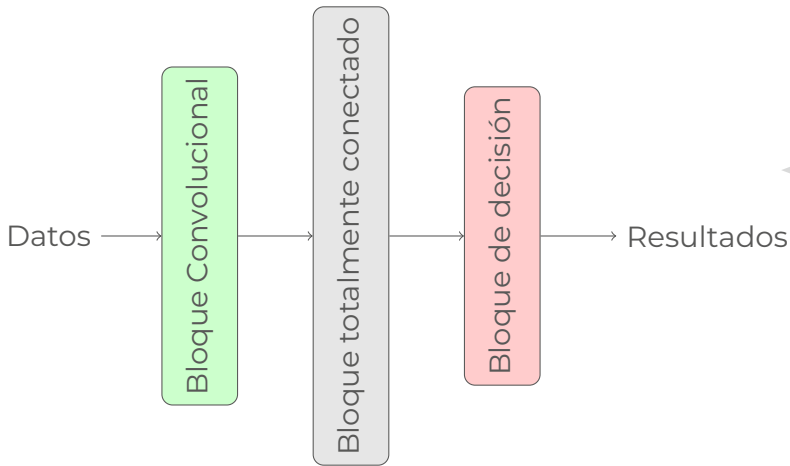
- **Bloque totalmente conectado:**

- Cada neurona se conecta con todas las neuronas de la capa anterior y siguiente.
- Correlacionan las características detectadas por las capas convolucionales.

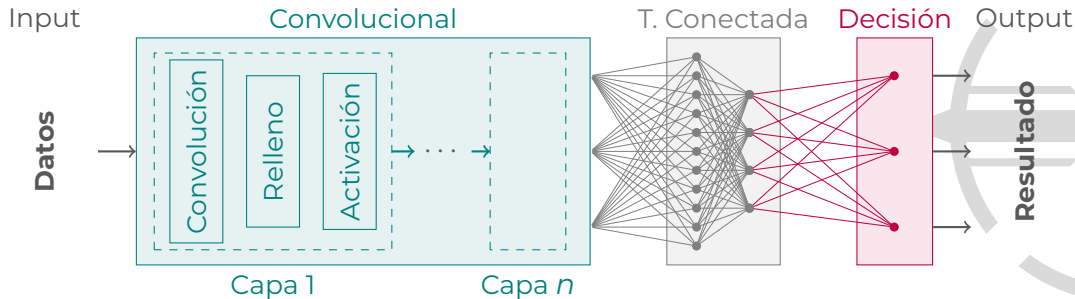
- **Bloque de decisión:**

- Genera un vector de probabilidades para cada clase.
- Usa funciones como softmax.

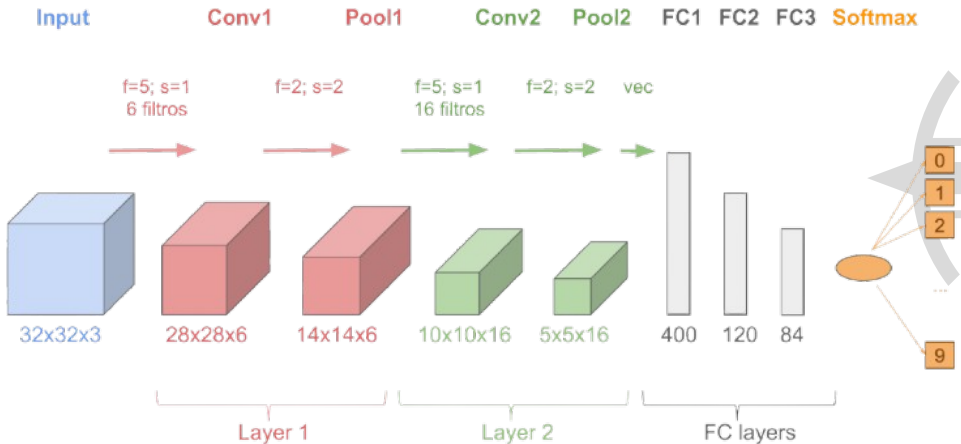
ESTRUCTURA DE UNA RED NEURONAL CONVOLUCIONAL (CNN)



ESTRUCTURA DE UNA RED NEURONAL CONVOLUCIONAL (CNN)



EJEMPLO DE UNA CNN: LENET-5



CAPA DE ENTRADA (INPUT LAYER)

- La capa de entrada recibe la imagen (o los datos de entrada) y la prepara para su procesamiento.
- No hay parámetros que aprender en esta capa.
- La imagen se representa como una matriz de píxeles, que puede ser en escala de grises o RGB.

CAPAS CONVOLUCIONALES (CONVOLUTIONAL LAYERS)

- Una capa convolucional tiene l canales de entrada, k canales de salida, y un tamaño de filtro f .
- El número total de pesos es:

$$f \times f \times l \times k$$

- Además, cada mapa de características tiene un valor de sesgo, por lo tanto, el número total de parámetros es:

$$f \times f \times (l + 1) \times k$$

CAPAS DE AGRUPAMIENTO (POOLING LAYERS)

- Las capas de agrupamiento (pooling) se encargan de reducir las dimensiones de las características, haciéndolas más manejables.
- No hay parámetros que aprender durante el proceso de entrenamiento en estas capas.

CAPAS TOTALMENTE CONECTADAS (FULLY-CONNECTED LAYERS)

- En las capas totalmente conectadas, cada unidad de entrada se conecta con todas las unidades de salida.
- El número de pesos es $n \times m$, donde n es el número de entradas y m el número de salidas.
- Además, cada neurona de salida tiene un valor de sesgo, por lo tanto, el número total de parámetros es:

$$(n + 1) \times m$$

CAPA DE SALIDA (OUTPUT LAYER)

- La capa de salida es generalmente una capa totalmente conectada.
- Genera las probabilidades o clasificaciones finales.
- El número total de parámetros en la capa de salida también es:

$$(n + 1) \times m$$

Capa	Dimensión	Tamaño	Parámetros
Input	$(32 \times 32 \times 3)$	3072	0
Conv1	$(28 \times 28 \times 6)$	4704	456
Pool1	$(14 \times 14 \times 6)$	1176	
Conv2	$(10 \times 10 \times 16)$	1600	
Pool2	$(5 \times 5 \times 16)$	400	
FC2	(120×1)	120	
FC3	(84×1)	84	
Softmax	(10×1)	10	
Total			

Capa	Dimensión	Tamaño	Parámetros
Input	$(32 \times 32 \times 3)$	3072	0
Conv1	$(28 \times 28 \times 6)$	4704	456
Pool1	$(14 \times 14 \times 6)$	1176	0
Conv2	$(10 \times 10 \times 16)$	1600	2416
Pool2	$(5 \times 5 \times 16)$	400	0
FC2	(120×1)	120	48120
FC3	(84×1)	84	10164
Softmax	(10×1)	10	850
Total			62006



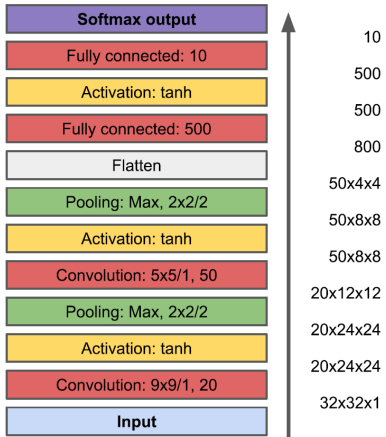
**Pontificia Universidad
Católica del Ecuador**
Seréis mis testigos

FACULTAD DE
CIENCIAS EXACTAS Y NATURALES

ARQUITECTURAS DE CNN

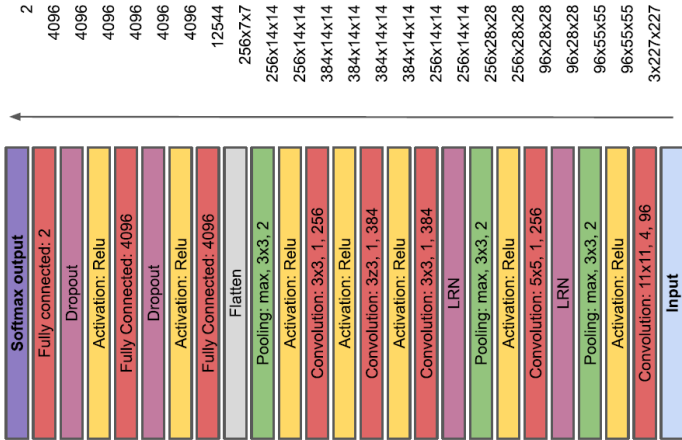


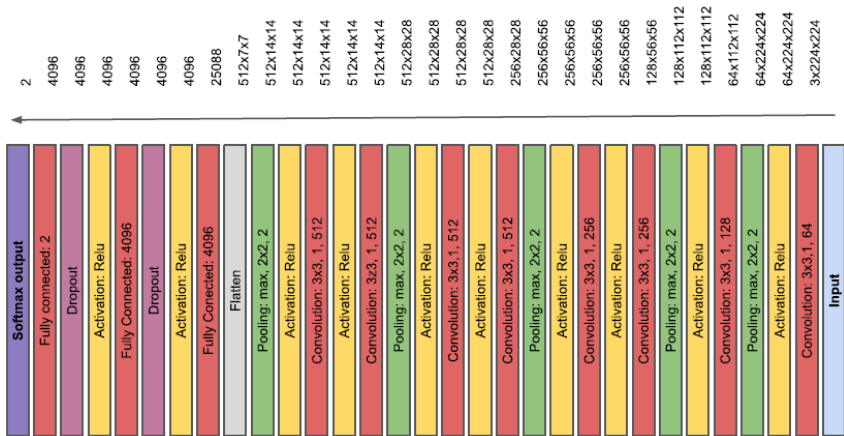
LENET-5



- Desarrollada por Bengio et al. en 1998.
- Objetivo: reconocimiento de dígitos escritos a mano.
- Pequeña y eficiente.
- Ideal para aprender y realizar pruebas iniciales.
- Puede ejecutarse en una CPU sin necesidad de GPU.

ALEXNET







TRANSFERENCIA DEL APRENDIZAJE

- Permite reutilizar un modelo entrenado en una tarea para aplicarlo en otra tarea similar.
- Mejora el rendimiento del modelo y acelera el progreso.
- Es popular en ****Deep Learning**** debido a los altos costos computacionales y de datos necesarios para entrenar modelos desde cero.
- La técnica más común es usar un **modelo pre-entrenado** y adaptarlo:
 - Seleccionar el modelo fuente.
 - Reutilizar todo o parte del modelo.
 - Ajustar el modelo a la nueva tarea (tunning).

TÉCNICAS DE AJUSTE (TUNNING)

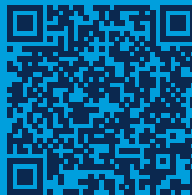
- **Extracción de características:** usar el modelo como extractor de características.
- **Congelación de capas:** mantener los pesos de algunas capas y reentrenar otras.



**Pontificia Universidad
Católica del Ecuador**

Seréis mis testigos

Gracias



Presentación

Contacto: aemerinot@puce.edu.ec