

Análisis de Datos

Redes neuronales artificiales

Red neuronal convolucional

Dr. Wilfrido Gómez Flores

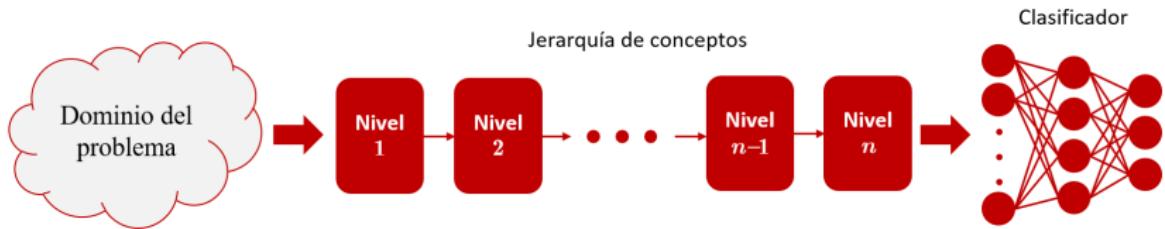


Introducción

Tradicionalmente, la extracción de características se hace mediante ingeniería humana: un diseñador propone un conjunto de descriptores *hand-crafted* para un problema específico. En el paradigma de aprendizaje profundo, el conjunto de descriptores es aprendido automáticamente mediante técnicas de inteligencia artificial. 

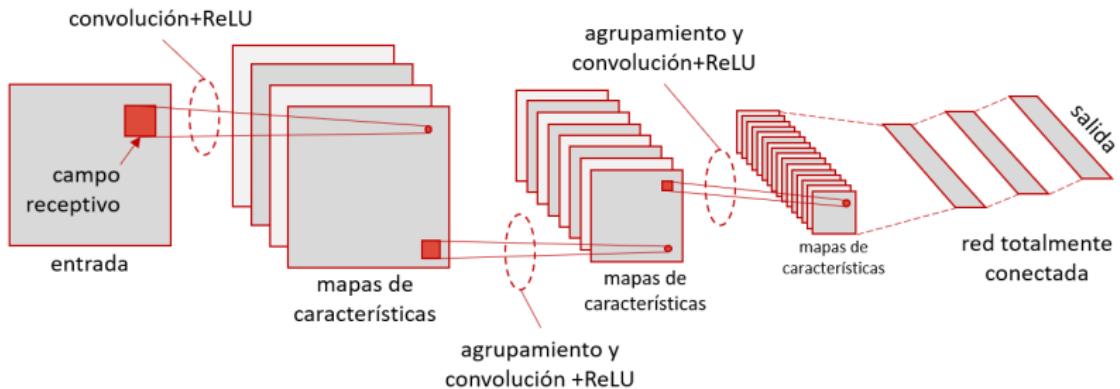
Introducción

El aprendizaje profundo es un subcampo de la inteligencia artificial en el que una máquina aprende mediante jerarquías de conceptos con diferentes niveles de abstracción, evitando la necesidad de expertos que especifiquen todo el conocimiento.



Red neuronal convolucional

Una red neuronal convolucional (CNN) es un tipo de red neuronal profunda para procesar datos que tienen una topología en forma de cuadrícula (e.g., imágenes).



El nombre “red neuronal convolucional” indica que la red aplica la operación matemática de convolución en al menos una de sus capas. En la imagen se muestra la red LeNet-5 propuesta para reconocimiento de dígitos manuscritos.

Red neuronal convolucional

- **Convolución:** operador matemático sobre dos funciones f y h que produce una tercera función $f * h$ que expresa cómo la forma de una es modificada por la otra: 

$$(h * f)(x, y) = \sum_{i=-a}^a \sum_{j=-b}^b h(i, j) f(x - i, y - j), \quad (1)$$

donde f es una matriz de tamaño $M \times N$ y h es un filtro (máscara) de tamaño $m \times n$ con $a = (m - 1)/2$ y $b = (n - 1)/2$.

- Al resultado de la convolución se le aplica una función de activación no lineal; típicamente la función ReLU: $f(x) = \max(x, 0)$.

Red neuronal convolucional

Agrupamiento: proceso de remuestreo para reducir el tamaño de un mapa de características, lo cual provee una representación abstracta.

Generalmente se utiliza agrupamiento máximo sin traslape, aunque también se ha propuesto utilizar el agrupamiento promedio. ➔

Red neuronal convolucional

El aprendizaje de una CNN ajusta los coeficientes W de las máscaras de convolución mediante la minimización de una función de pérdida $J(W)$ usando *backpropagation*:

$$\begin{aligned} W(k = 0) & \quad \text{arbitrario} \\ W(k + 1) &= W(k) + \eta \frac{\partial J(W)}{\partial W} \quad k \geq 1 \end{aligned} \tag{2}$$

Ajuste de seis máscaras de convolución de tamaño 5×5 . 

Red neuronal convolucional

Cada capa convolucional extrae mapas de características a partir de las salidas de la capa anterior, los cuales describen rasgos más sutiles a medida que aumenta la profundidad. A la salida, una red completamente conectada (FC) genera un vector, cuyo tamaño es igual al número de clases, que es clasificado por la función softmax.



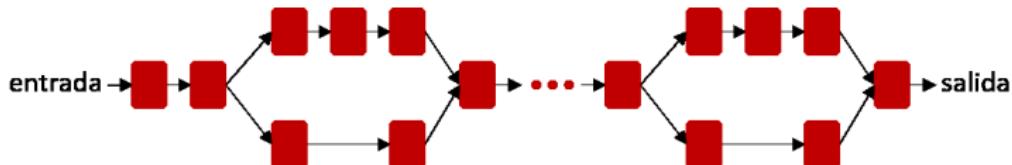
Arquitecturas CNN

Tipos de arquitecturas CNN:

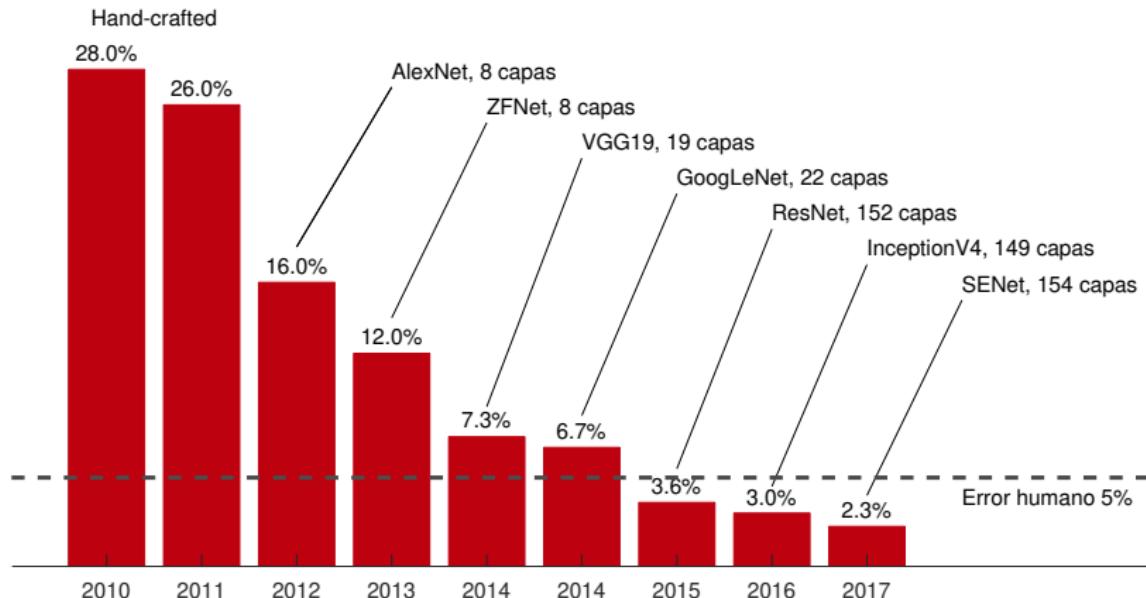
- **Red en serie:** las capas se disponen una tras otra, con una capa de entrada única y una capa de salida única (poca profundidad).



- **Red de grafo acíclico dirigido (DAG):** las capas tienen entradas de varias capas y salidas a varias capas (mucha profundidad).

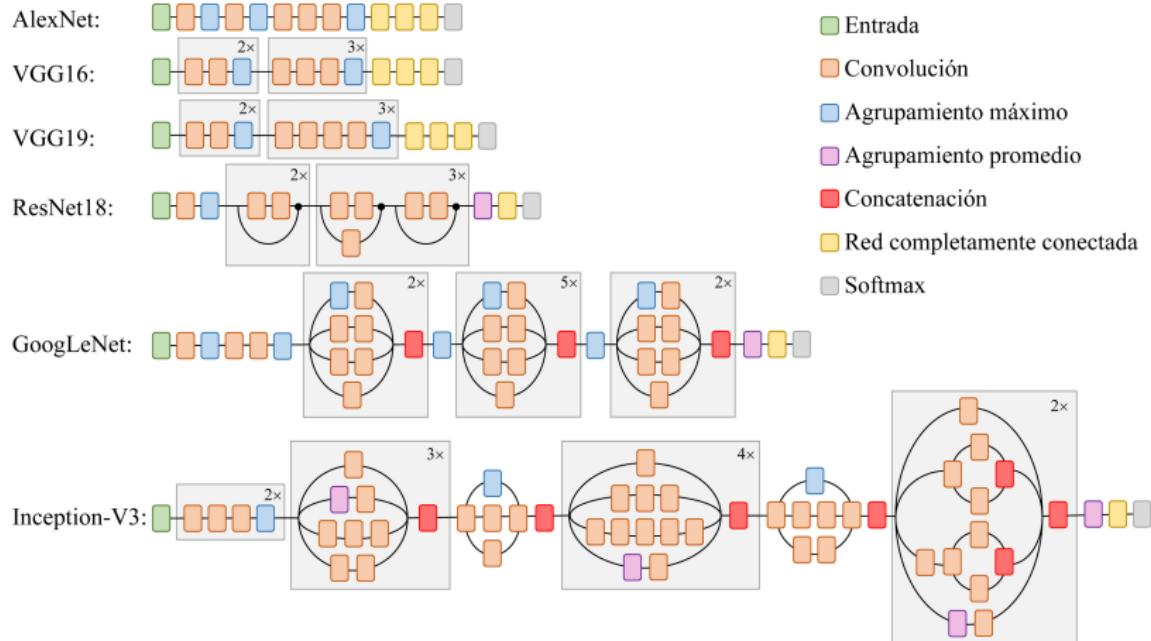


Arquitecturas CNN



Los resultados del desafío *The ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) muestran que las arquitecturas CNNs están superando los niveles humanos de precisión al aumentar la profundidad de la red. Las CNNs se entrenaron sobre la base de datos ImageNet que considera 1 millón de imágenes etiquetadas en 1000 clases de objetos.

Arquitecturas CNN



Ejemplos de algunas arquitecturas CNNs entrenadas sobre ImageNet.

Arquitecturas CNN

Características de los modelos CNN mostrados en la diapositiva anterior. Los símbolos † y ‡ indican que la red es seriada y DAG, respectivamente. El número de parámetros entrenables está dado en millones.

Modelo CNN	Entrada	Capas	Parámetros
AlexNet [†]	227×227	8	60 M
VGG16 [†]	224×224	16	138 M
VGG19 [†]	224×224	19	144 M
ResNet18 [‡]	224×224	18	11 M
GoogLeNet [‡]	224×224	22	7 M
Inception-V3 [‡]	299×299	48	24 M

Transferencia de aprendizaje

La transferencia de aprendizaje es un tipo de aprendizaje automático en el que un modelo obtenido al resolver un problema se aplica para resolver otro problema distinto pero relacionado.

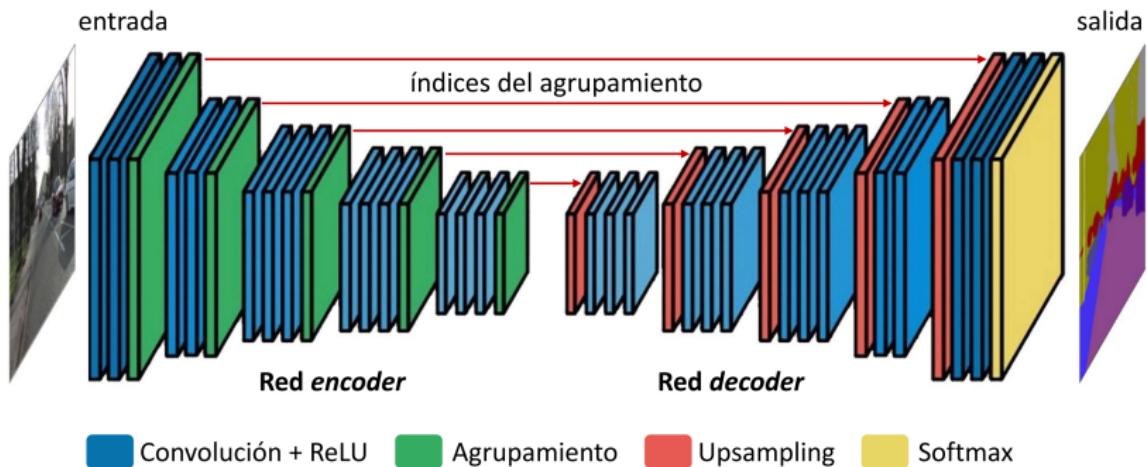
Transferencia de aprendizaje en CNNs:

- Transferir los parámetros de una CNN desde el dominio de origen al dominio de destino.
- Mejorar el punto de partida en el dominio de destino y únicamente realizar un ajuste fino de parámetros.
- Útil para reducir el sobre-ajuste de un modelo entrenado sobre conjuntos de datos pequeños.

Transferencia de aprendizaje

Un modelo CNN preentrenado en la base de datos ImageNet es reutilizado para distinguir entre perros y gatos. Las capas convolucionales originales se mantienen y se reemplazan las capas finales completamente conectadas y la capa softmax para clasificar entre dos clases de animales. Después se hace un ajuste fino de los parámetros de la nueva CNN mediante *backpropagation*. Se pueden “congelar” algunas capas convolucionales para no ajustar sus parámetros. ➤

Arquitectura *encoder-decoder*

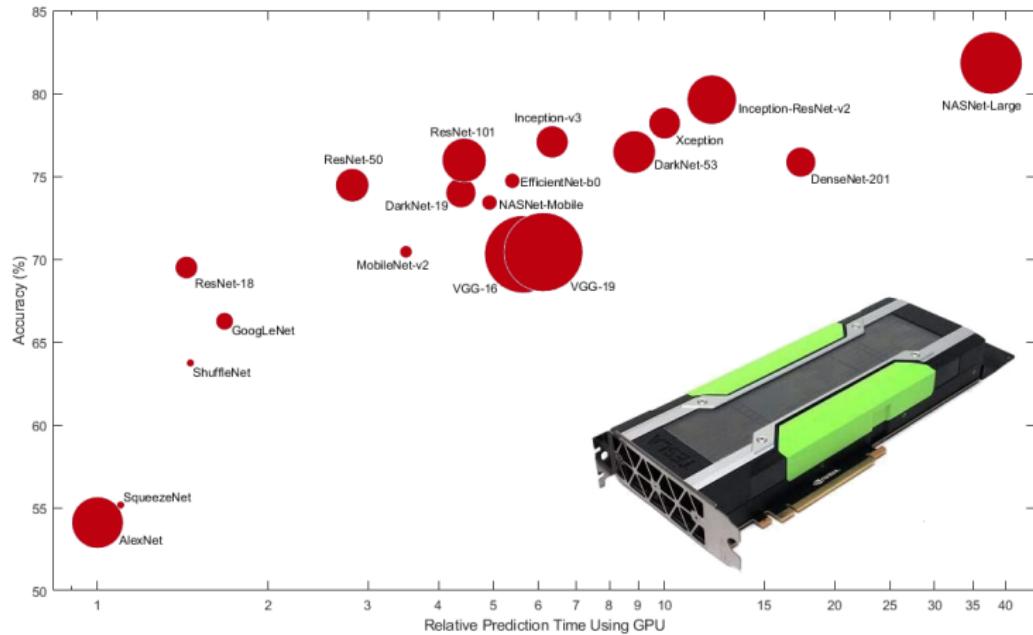


Una **red encoder** toma una señal de entrada $X \in \mathbb{R}^{d_x}$ y la mapea hacia un espacio de características $Z \in \mathbb{R}^{d_z}$, mientras que una **red decoder** toma este espacio de características como entrada, y produce una señal de salida $Y \in \mathbb{R}^{d_y}$. Este tipo de arquitectura generalmente se usa para la **segmentación semántica** de imágenes para realizar clasificación a nivel de píxel. En la imagen se muestra el modelo SegNet.

Arquitectura *encoder-decoder*

La segmentación semántica es muy utilizada en la conducción autónoma. 

CNNs y GPUs



La fase de entrenamiento de un modelo CNN es la tarea que más consume recursos computacionales. Las GPU están optimizadas para entrenar modelos de inteligencia artificial y aprendizaje profundo, ya que pueden procesar múltiples cálculos simultáneamente.