UT3

Computer Vision

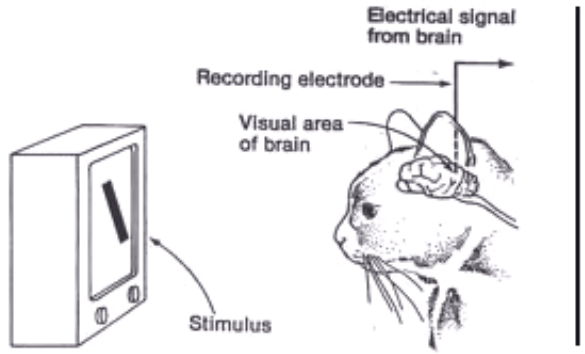
Fundamentos del Aprendizaje Automático

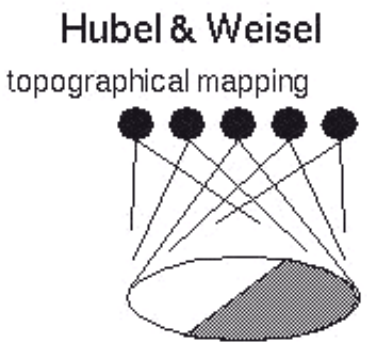
Profesor: Ing. Juan Francisco Kurucz

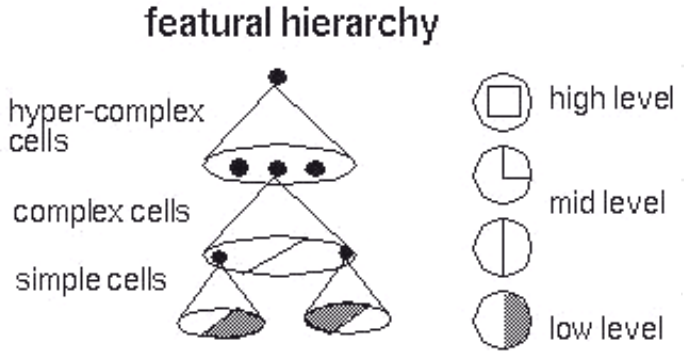
juan.kuruczsosa@ucu.edu.uy

Corteza visual

- Hubel & Wiesel (1959 y 1962)
- Organización jerárquica
  - Células simples
  - Células complejas
  - Células hipercomplejas

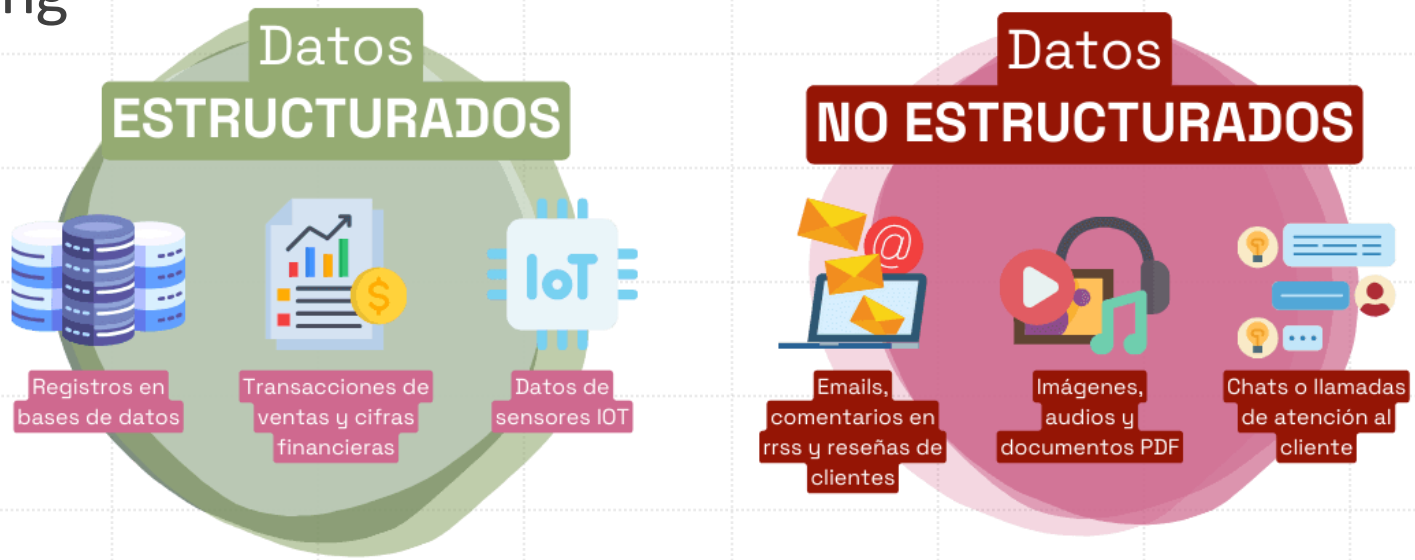






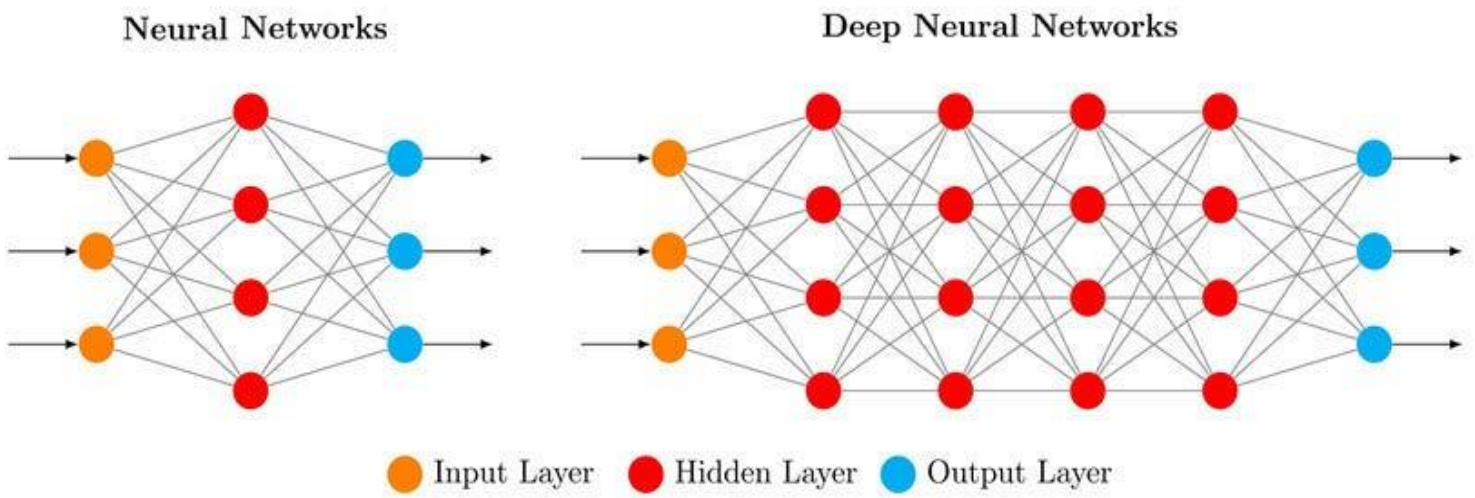
Deep Learning

- 2 clases de datos:
  - Datos estructurados (tablas, CSV): Feature extraction
  - Datos no estructurados (imágenes, sonidos, texto): Representation learning



Deep Learning - DNN

- Múltiples capas
- Cada capa aprende una forma distinta de representación
- Características jerarcas

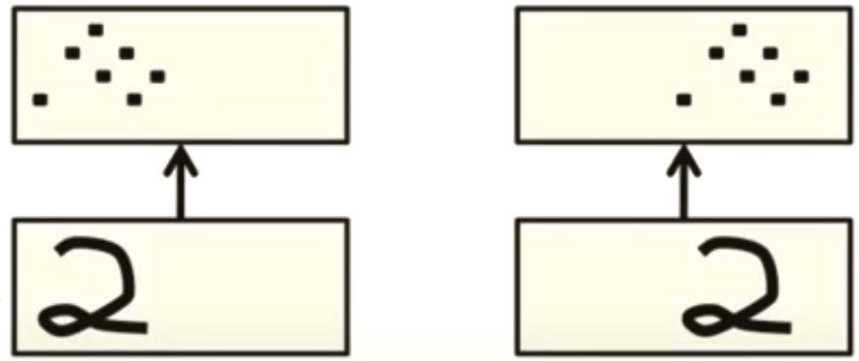


Problemas Redes Neuronales comunes

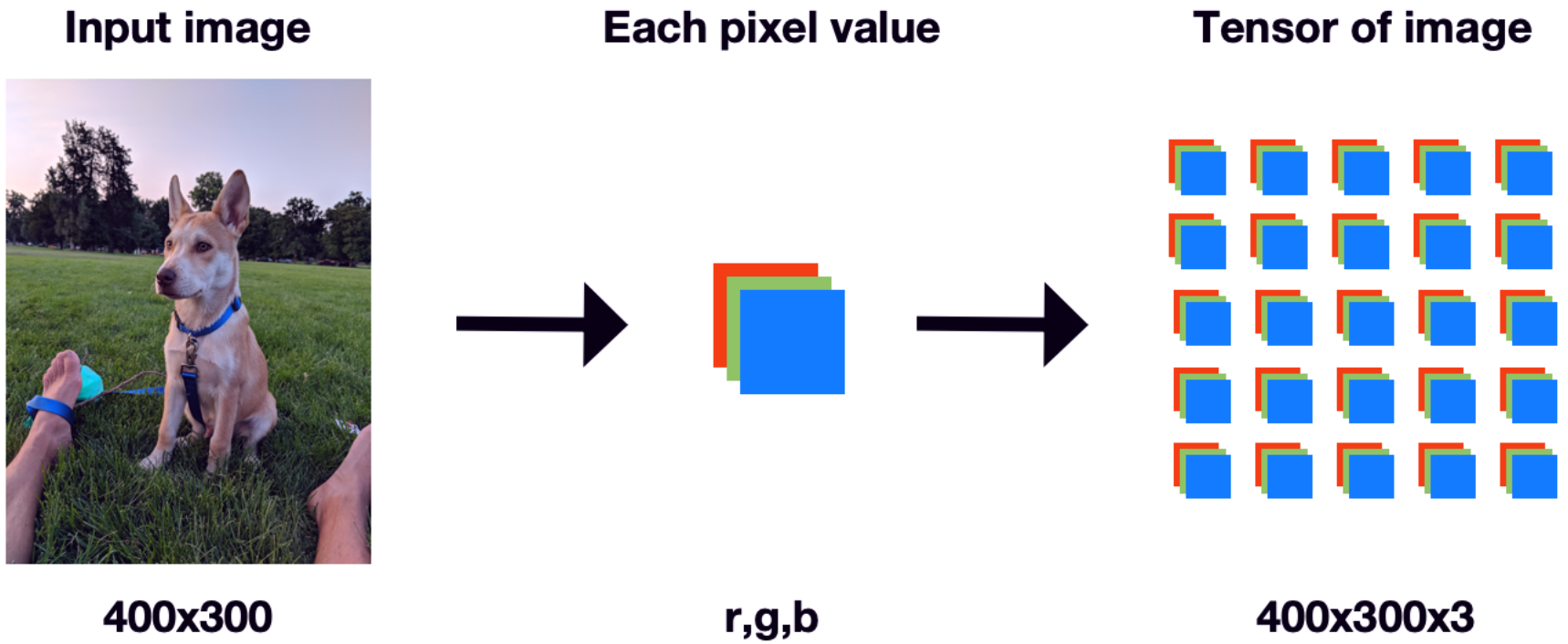
- Cada neurona pondera todas las entradas
- Cantidad de parámetros
  - Entrada: imagen 32 x 32 x 3
  - 3072 conexiones por neurona!!
- Información local

Características replicadas

- En una imagen los objetos pueden aparecer en diferentes partes
- Reutilizar detectores de características en mas de una posición
  - Reduce los parámetros para aprender
- Conocimiento invariante

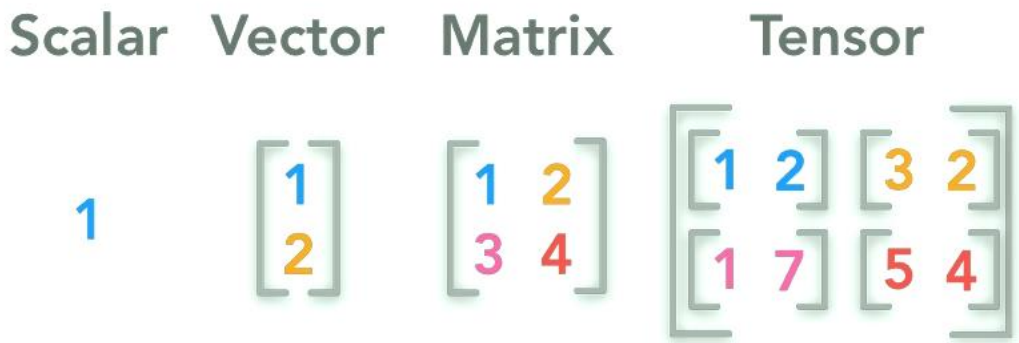


Representación de imágenes



Tensor

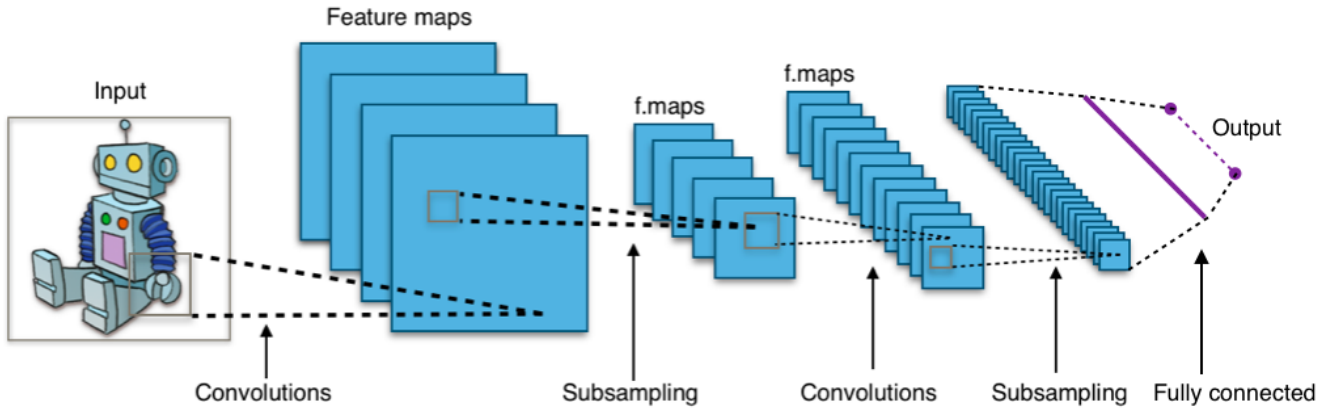
- Estructura de datos muy usada en Machine Learning y Deep Learning
- Arreglo multidimensional
  - Orden 0: Escalar
  - Orden 1: Vector
  - Orden 2: Matriz
  - Orden n: Tensor



Desafios de imágenes

Redes Neuronales Convolucionales (CNN)

- Especializadas para el reconocimiento de patrones complejos
- Exitosas para problemas de Computer Vision
  - Reconocimiento de objetos, caras, etc
- Procesamiento de todo tipo de señales
  - Imágenes
  - Audio
  - Video
  - Medicina



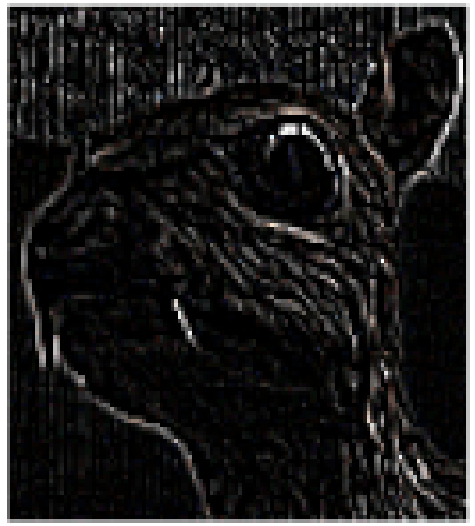
Convolución

Input image

Convolution Kernel

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Feature map

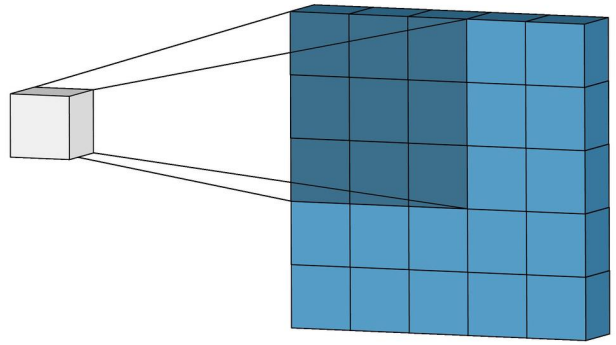


https://setosa.io/ev/image-kernels/

Filtro o Kernel

- Extraer características de la entrada
- Detección de patrones
  - Bordes
  - Esquinas
  - Figuras
  - Objetos
  - Etc
- Cuanto mas profunda es la red, mas sofisticados son los filtros.

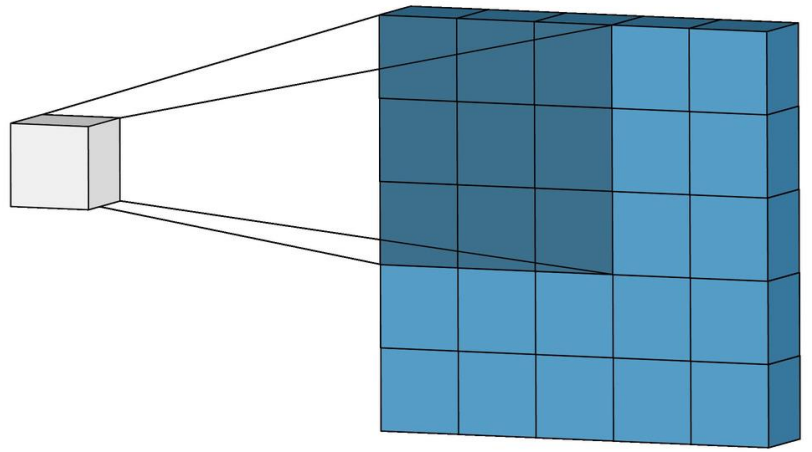
Filtros – funcionamiento



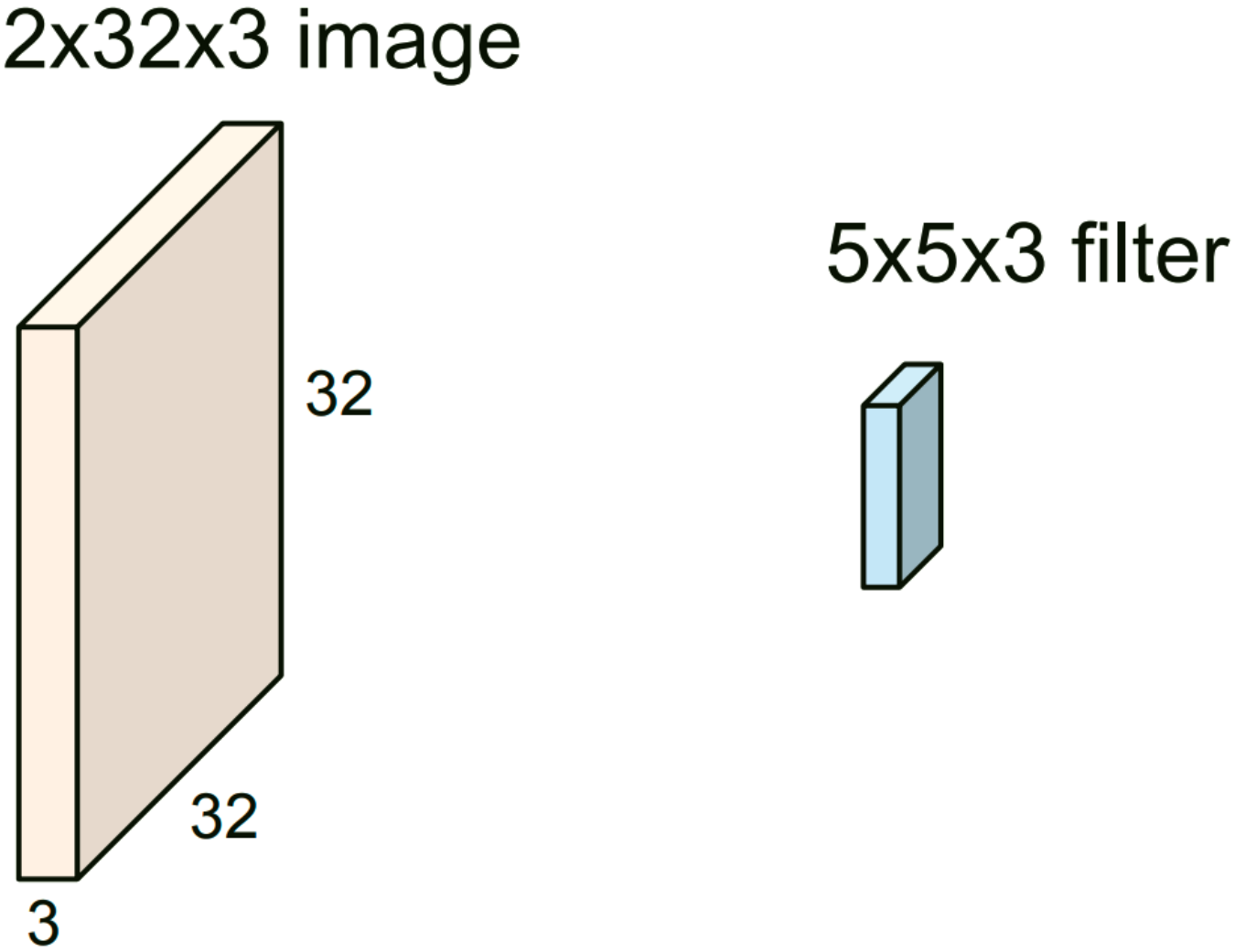
- Ventana pequeña
- Matriz de tamaño definido
- Se va “desplazando” por los valores de entrada
- Operación de convolución

Capa convolutiva

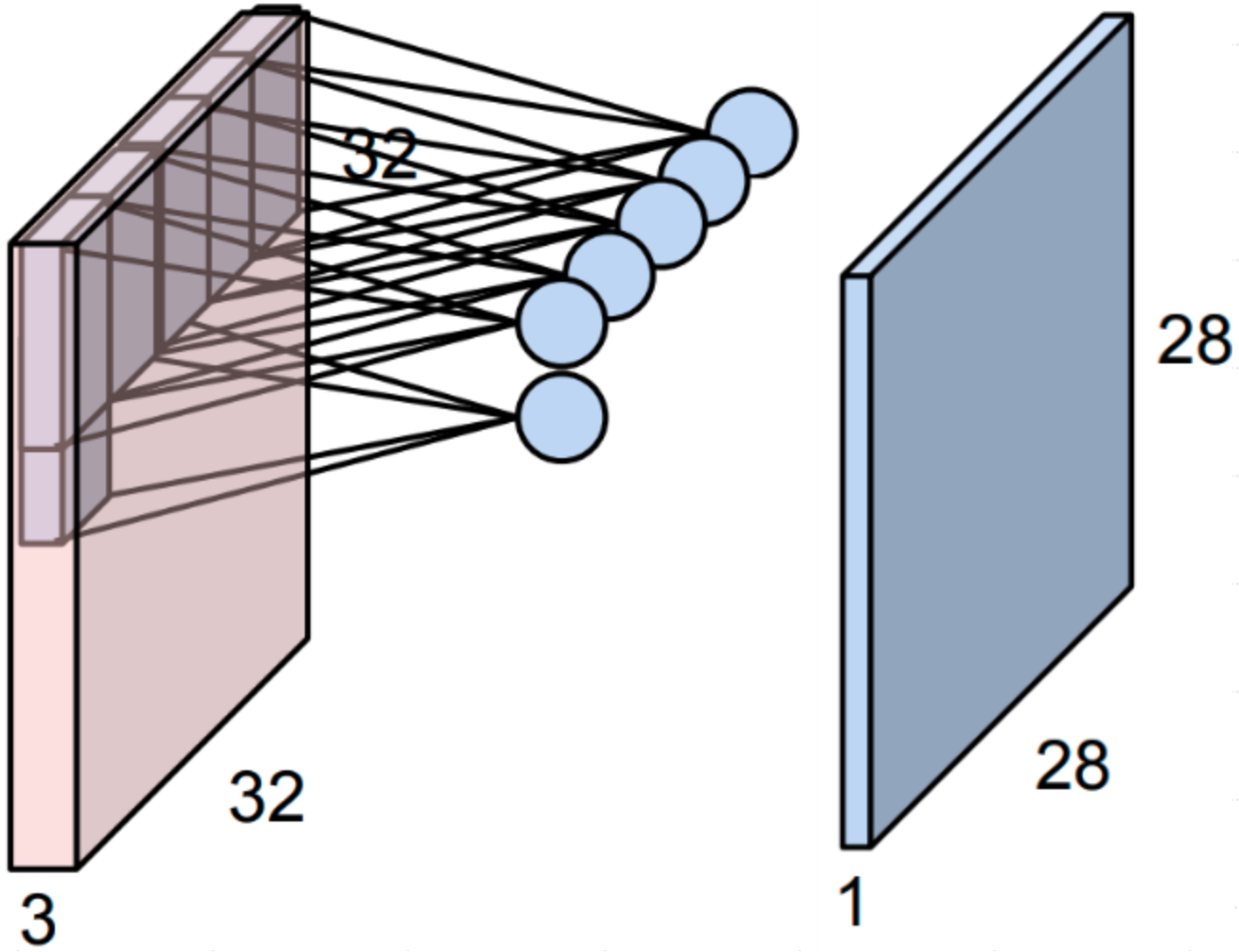
- Compuesta de unidades convolutivas (filtros)
- Los pesos de las neuronas son los valores del filtro
- Varios filtros por capa convolutiva



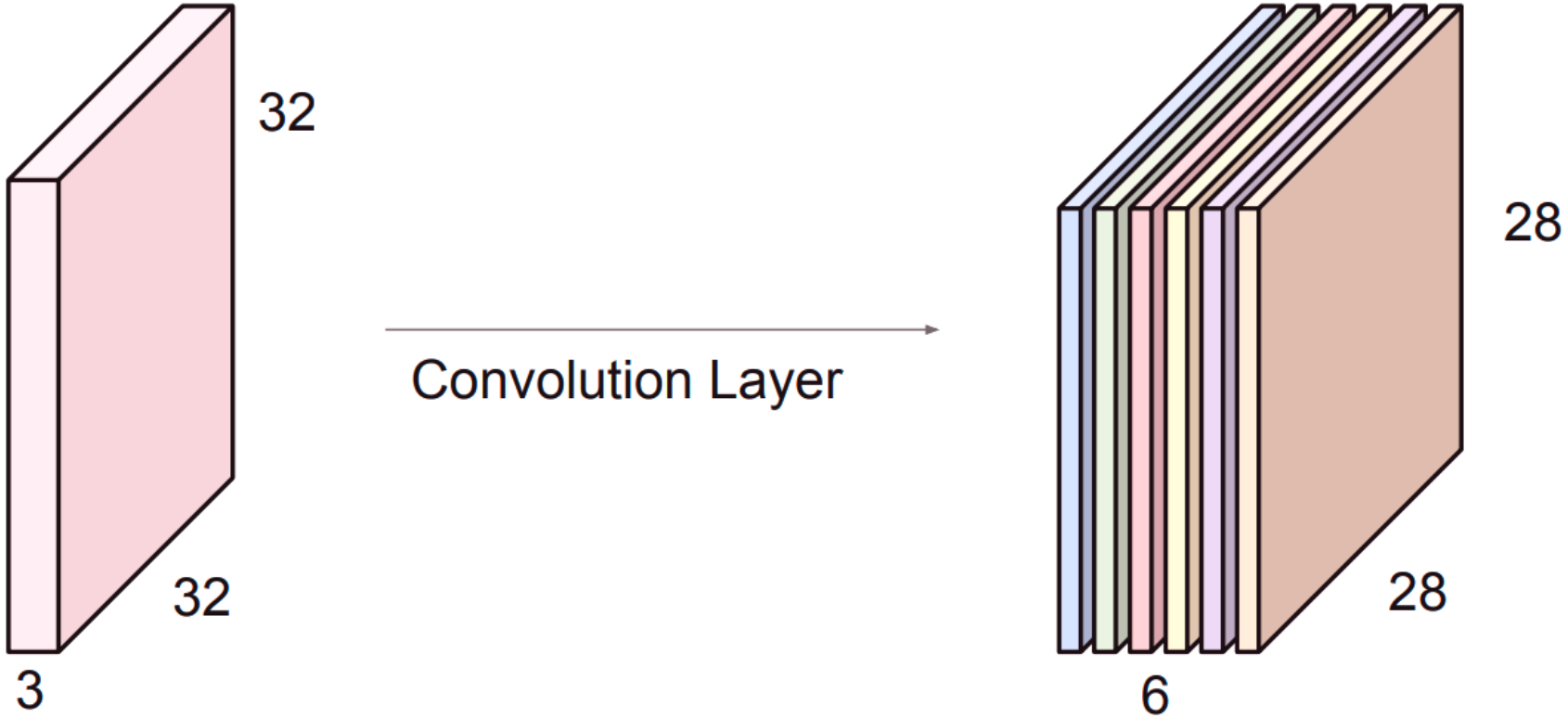
Capa convolutiva



Capa convolutiva



Capa convolutiva



Con 6 filtros de 5x5x3

¿Cuantos parámetros (w) hay en la capa?

NN vs CNN – Cantidad de Parámetros

Imagen 32x32x3

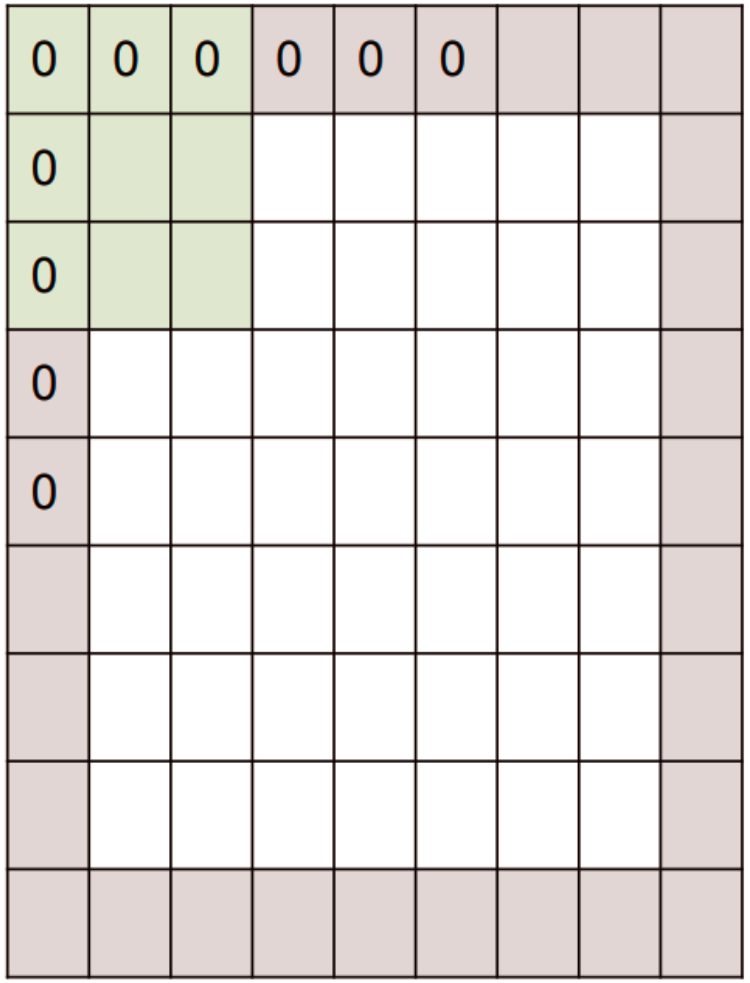
NN

- Cada neurona conecta con todas las anteriores
- 32 x 32 x 3 = 3.072 parámetros por cada neurona en la capa
- Con 100 neuronas son **307.200**

CNN

- 6 filtros de 5x5
- 5 x 5 x 3 x 6 = **450** parámetros para toda la capa

Padding

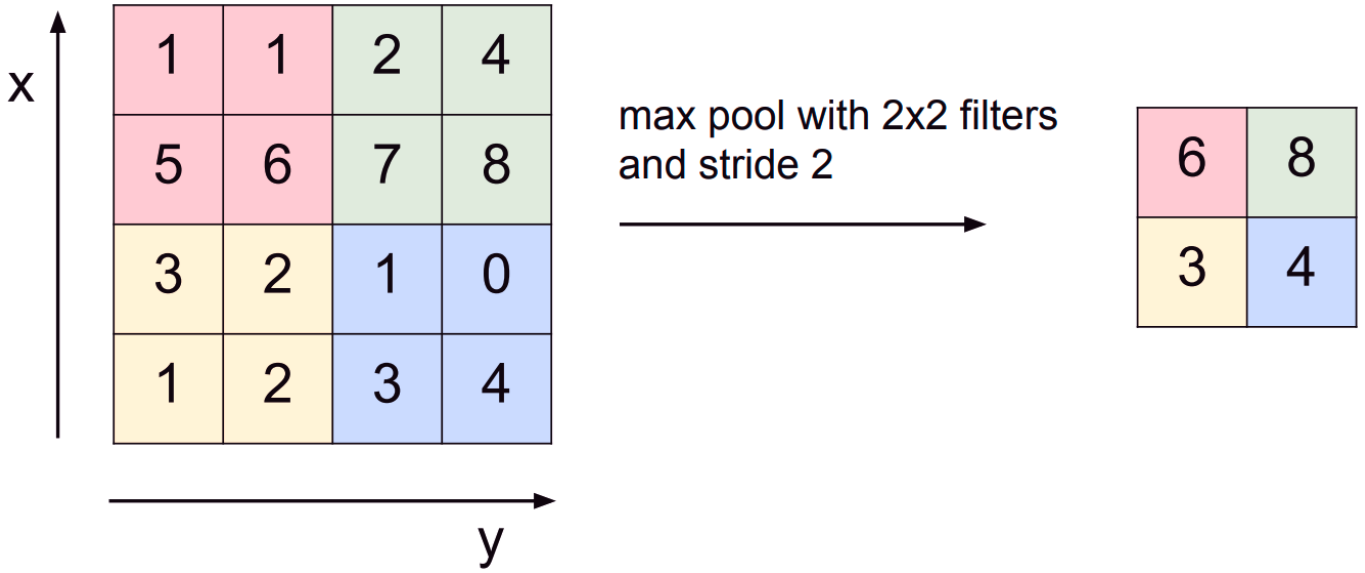


Hiperparametros CNN

- Cantidad de filtros o kernels
- Tamaño del filtro o kernel
- Desplazamiento del filtro o Stride
- Padding

Capa de pooling

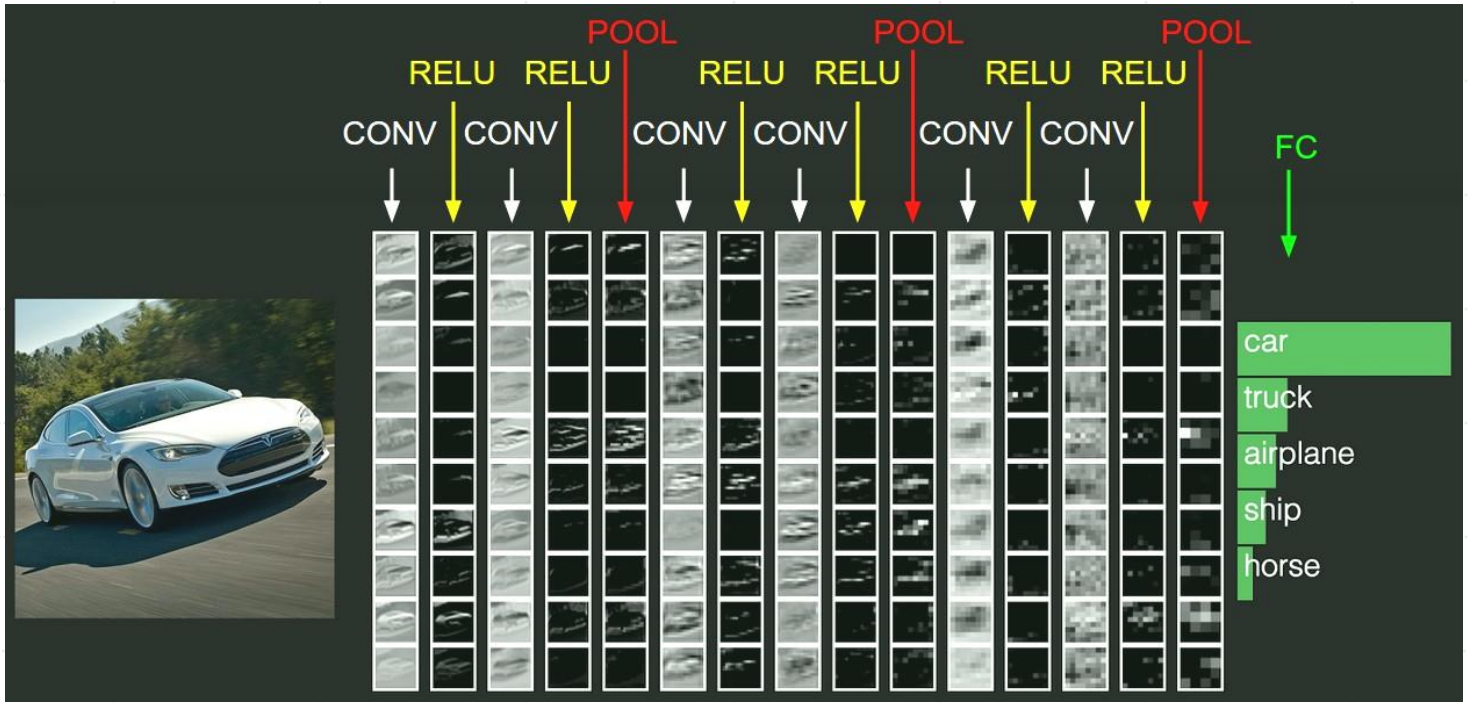
- Muestreo
- Sirve para condensar la información
- Tipos
  - Max pooling
  - Average pooling
  - Etc



- Previene variaciones en la posición de las características
- Reduce la memoria

Estructura CNN

- Combinación de capas convolucionales y pooling
- Detección de clases al final (red FC)



https://cs231n.stanford.edu/

Ejemplo: LeNet (Yann LeCun 1998)

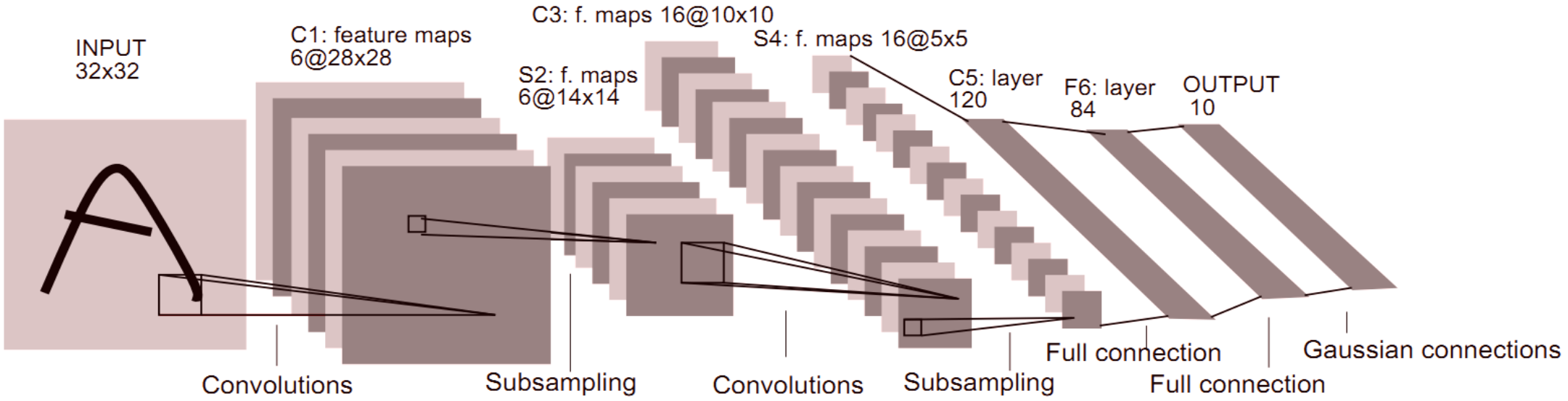
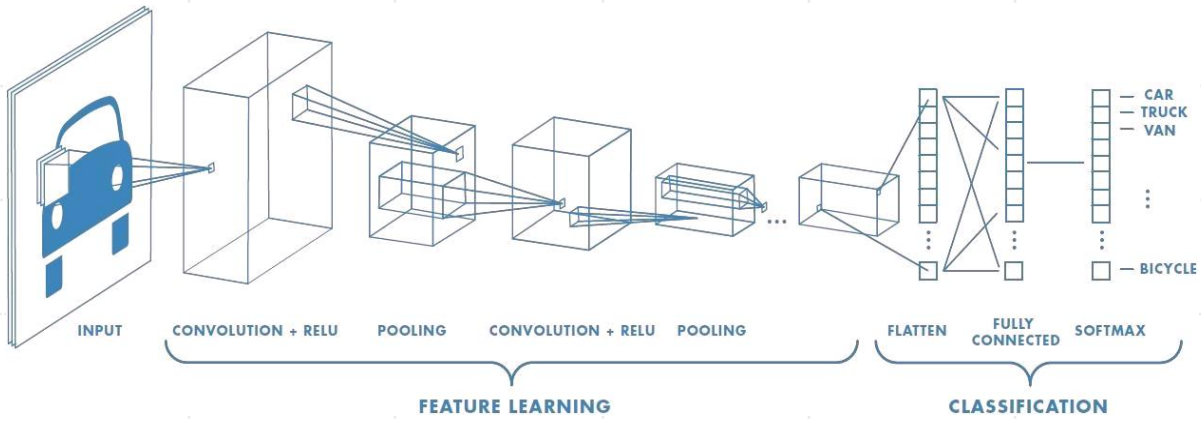


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Estructura CNN

- Primeras capas
  - características primitivas (bordes, contrastes, etc)
- Ultimas capas
  - características mas complejas, abstractas y especificas (formas, objetos, texto, etc)



Transfer Learning

El entrenamiento de una red neuronal requiere:

- Muchos datos disponibles y relevantes
- Gran poder de computo
- Tiempo

Solución:

- usar un modelo ya entrenado y adaptarlo a otro problema

Transfer Learning

Es necesario:

- remplazar capas del final con nuevas
- usar la red ya entrenada para la extracción de características
- entrenar solamente los parámetros de las capas nuevas

Transfer Learning

- Feature extraction:
  - Usar un modelo pre entrenado y adaptarlo agregando una capa de clasificación.
  - Se re usan los pesos del modelo base
  - Solo se entrena el clasificador
- Fine-Tuning:
  - Volver a entrenar las capaz superiores del modelo base junto con las capaz nuevas

Transfer Learning – “Model Zoo”

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Inception-ResNet-V2	215 MB	0.804	0.953	55,873,736	572
Xception	88 MB	0.79	0.945	22,910,480	126
Inception-v3	92 MB	0.788	0.944	23,851,784	159
DenseNet-201	80 MB	0.77	0.933	20,242,984	201
ResNet-50	99 MB	0.759	0.929	25,636,712	168
DenseNet-169	57 MB	0.759	0.928	14,307,880	169
DenseNet-121	33 MB	0.745	0.918	8,062,504	121
VGG-19	549 MB	0.727	0.91	143,667,240	26
VGG-16	528 MB	0.715	0.901	138,357,544	23
MobileNet	17 MB	0.665	0.871	4,253,864	88

Transfer Learning - Limitaciones

- No siempre se puede transferir conocimiento de un modelo a otro
  - Tipos de datos totalmente distintos: imágenes, audio, etc.
  - Arquitectura totalmente diferente

Data Augmentation

- Técnica para mejorar los resultados y prevenir overfitting
- Expandir el dataset creando versiones modificadas de las imágenes
- Transformaciones sobre las imágenes:
  - shift, flip, rotation, brightness, zoom, etc