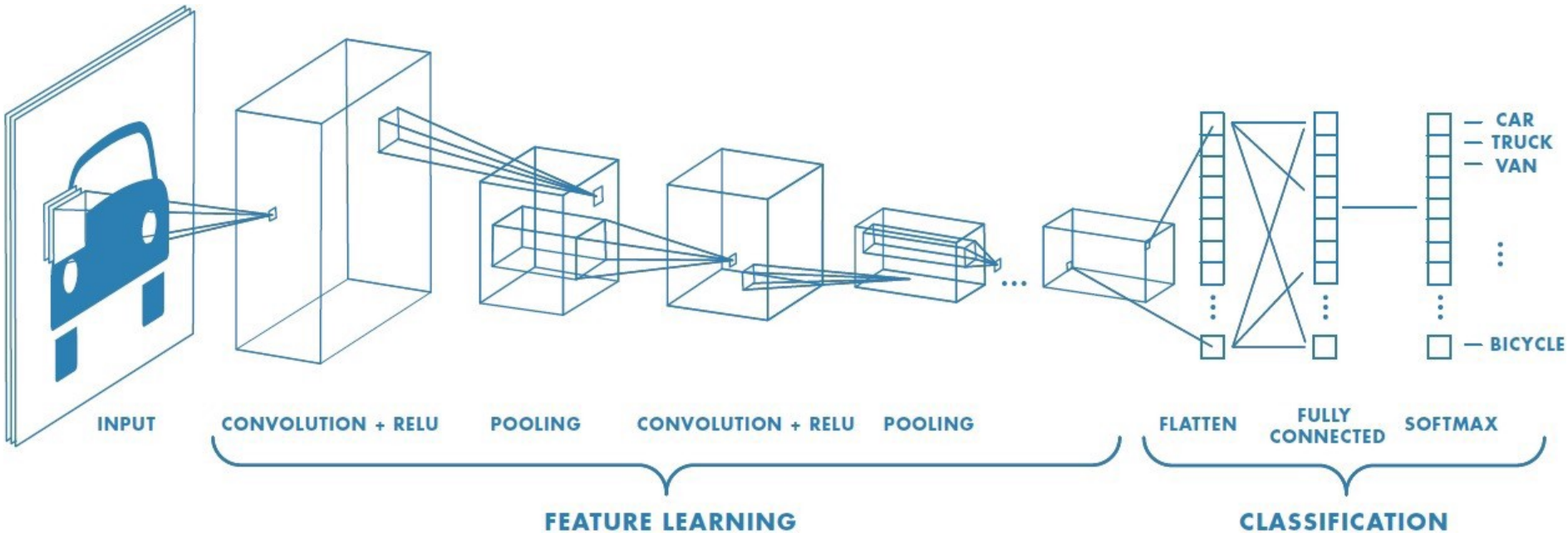
Convolutional Neural Networks

Dr. Mauricio Toledo-Acosta

Diplomado Ciencia de Datos con Python

Table of Contents

CNN


CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

Source

Utilidad de las CNNs

- Una imagen no es más que una matriz de valores de píxeles. **¿Por qué no basta con aplanar la imagen y alimentar la MLP para tareas de clasificación?**

Utilidad de las CNNs

- Una imagen no es más que una matriz de valores de píxeles. **¿Por qué no basta con aplanar la imagen y alimentar la MLP para tareas de clasificación?**
- Para imágenes muy básicas, este enfoque puede exhibir un desempeño razonable al realizar tareas de clasificación, pero **tendría poca precisión cuando se trata de imágenes complejas que tienen dependencias entre píxeles**.

Utilidad de las CNNs

- Una imagen no es más que una matriz de valores de píxeles. **¿Por qué no basta con aplanar la imagen y alimentar la MLP para tareas de clasificación?**
- Para imágenes muy básicas, este enfoque puede exhibir un desempeño razonable al realizar tareas de clasificación, pero **tendría poca precisión cuando se trata de imágenes complejas que tienen dependencias entre píxeles**.
- **Una CNN es capaz de capturar las dependencias espaciales de una imagen mediante la aplicación de filtros**. La arquitectura se ajusta mejor al conjunto de datos de la imagen gracias a la reducción del número de parámetros implicados.

CNN

Las Convolutional Neural Networks son muy similares a las redes neuronales ordinarias:

- Estan hechas de neuronas que aprenden pesos y sesgos.

CNN

Las Convolutional Neural Networks son muy similares a las redes neuronales ordinarias:

- Estan hechas de neuronas que aprenden pesos y sesgos.
- Cada neurona recibe una entrada, realiza un producto punto y tiene una activación no lineal.

CNN

Las Convolutional Neural Networks son muy similares a las redes neuronales ordinarias:

- Estan hechas de neuronas que aprenden pesos y sesgos.
- Cada neurona recibe una entrada, realiza un producto punto y tiene una activación no lineal.
- La red recibe las imágenes como conjuntos de píxels en un lado y produce scores de clases en el otro (en tareas de clasificación).

CNN

Las Convolutional Neural Networks son muy similares a las redes neuronales ordinarias:

- Estan hechas de neuronas que aprenden pesos y sesgos.
- Cada neurona recibe una entrada, realiza un producto punto y tiene una activación no lineal.
- La red recibe las imágenes como conjuntos de píxels en un lado y produce scores de clases en el otro (en tareas de clasificación).
- La red tiene una función de pérdida.

CNN

Las Convolutional Neural Networks son muy similares a las redes neuronales ordinarias:

- Estan hechas de neuronas que aprenden pesos y sesgos.
- Cada neurona recibe una entrada, realiza un producto punto y tiene una activación no lineal.
- La red recibe las imágenes como conjuntos de píxels en un lado y produce scores de clases en el otro (en tareas de clasificación).
- La red tiene una función de pérdida.
- Los pesos se actualizan con descenso de gradiente u otros optimizadores.

Capas de una CNN

Una red CNN se compone principalmente de tres tipos de capas:

- Capas convolucionales (convolutional).

Capas de una CNN

Una red CNN se compone principalmente de tres tipos de capas:

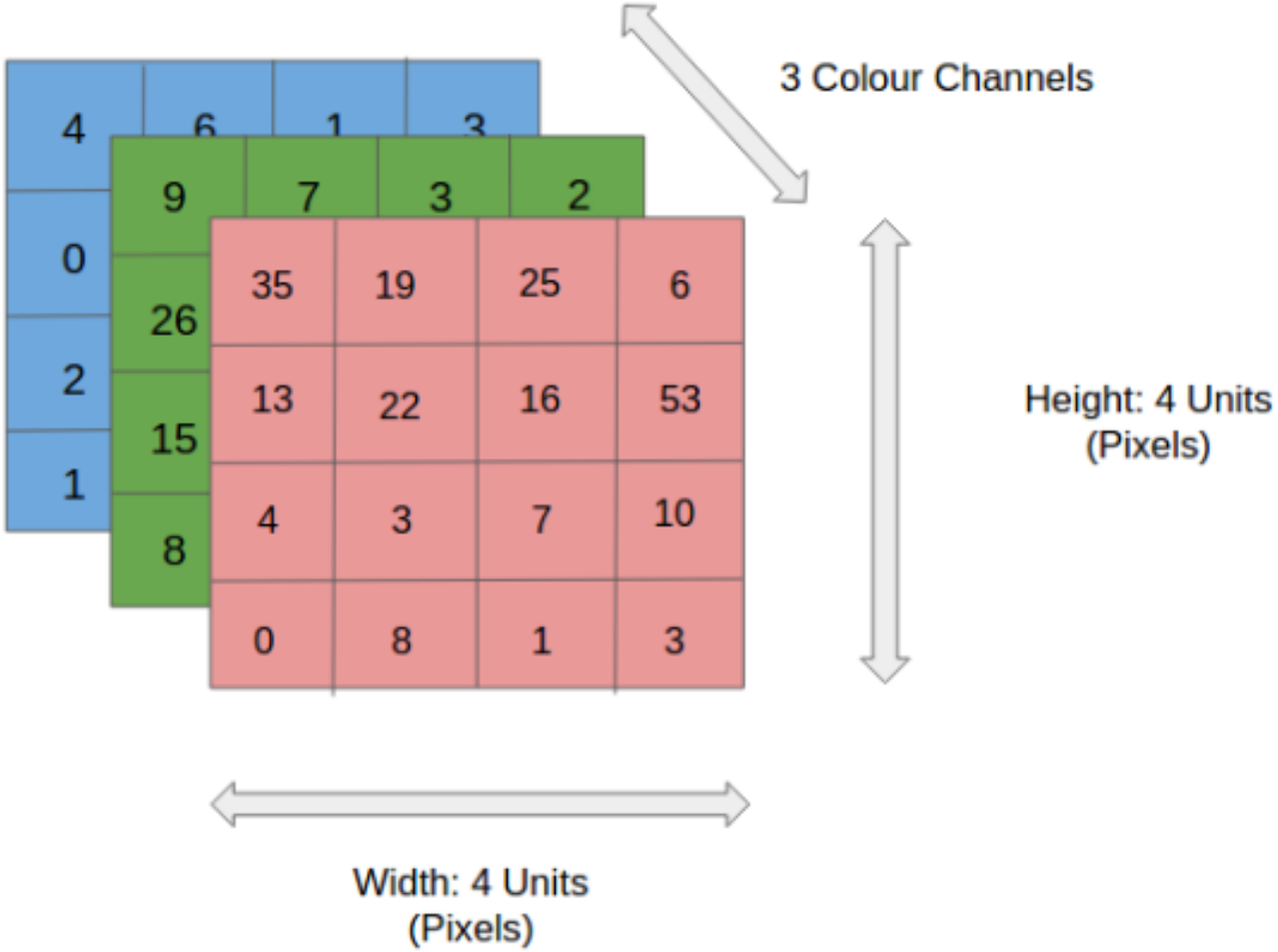
- Capas convolucionales (convolutional).
- Capas de pooling (pooling)

Capas de una CNN

Una red CNN se compone principalmente de tres tipos de capas:

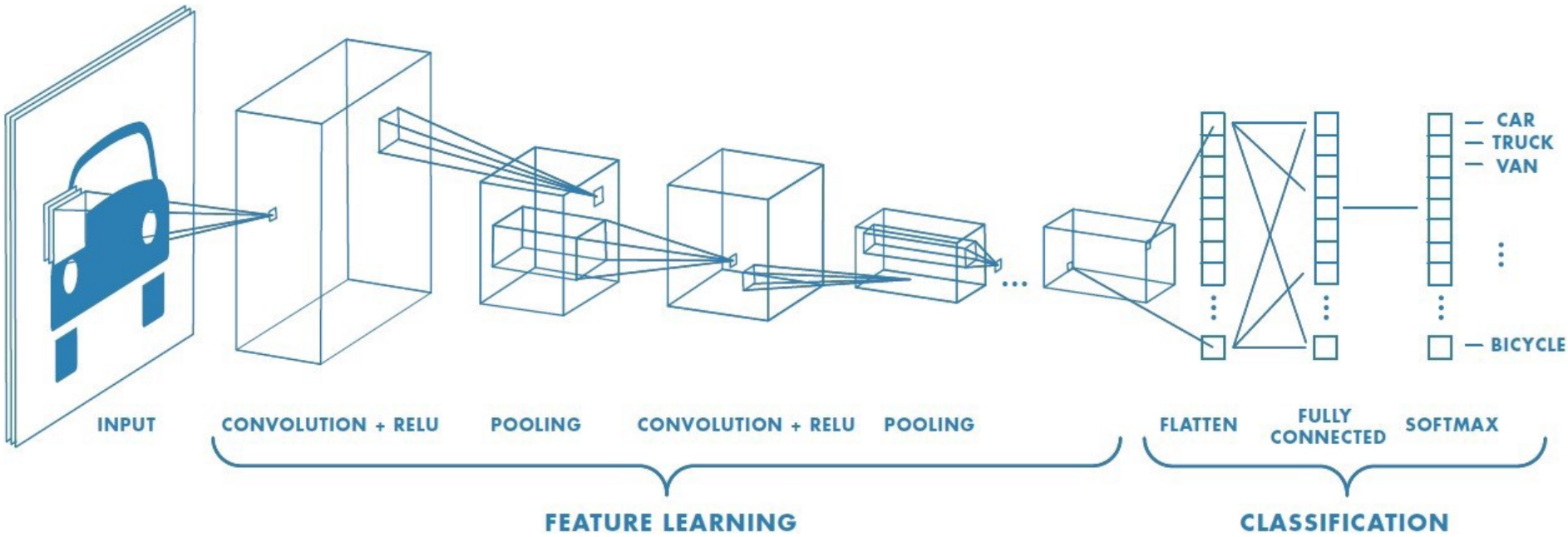
- Capas convolucionales (convolutional).
- Capas de pooling (pooling)
- Capas totalmente conectadas (fully connected).

Imagen como tensor



Source

CNN


CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

Kernel de Convolución

- También llamada matriz de convolución o máscara.

Kernel de Convolución

- También llamada matriz de convolución o máscara.
- Esta matriz es utilizada para transformar los valores de la imágen por medio de los valores del kernel.

Kernel de Convolución

- También llamada matriz de convolución o máscara.
- Esta matriz es utilizada para transformar los valores de la imágen por medio de los valores del kernel.
  - Es cuadrada y pequeña (3×3, 5×5).

Kernel de Convolución

- También llamada matriz de convolución o máscara.
- Esta matriz es utilizada para transformar los valores de la imágen por medio de los valores del kernel.
  - Es cuadrada y pequeña (3×3, 5×5).
  - Cuanto más grande es la matriz, más información local se pierde.

Kernel de Convolución

- También llamada matriz de convolución o máscara.
- Esta matriz es utilizada para transformar los valores de la imágen por medio de los valores del kernel.
  - Es cuadrada y pequeña (3×3, 5×5).
  - Cuanto más grande es la matriz, más información local se pierde.
- Permite efectos de *área* como desenfoque, nitidez y detección de bordes.

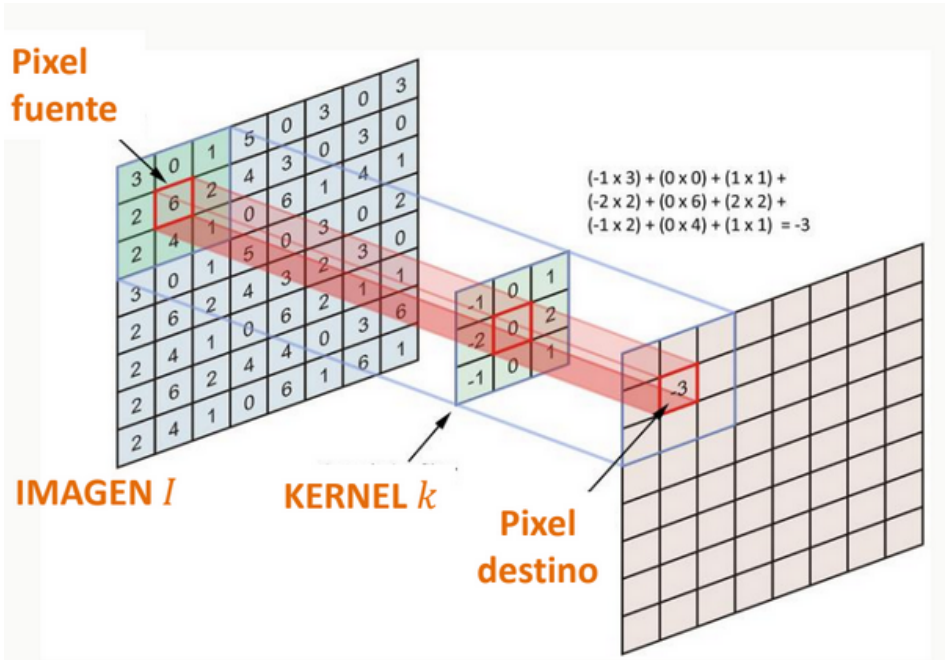
Kernel de Convolución

- También llamada matriz de convolución o máscara.
- Esta matriz es utilizada para transformar los valores de la imágen por medio de los valores del kernel.
  - Es cuadrada y pequeña (3×3, 5×5).
  - Cuanto más grande es la matriz, más información local se pierde.
- Permite efectos de *área* como desenfoque, nitidez y detección de bordes.
- No es una multiplicación de matrices.

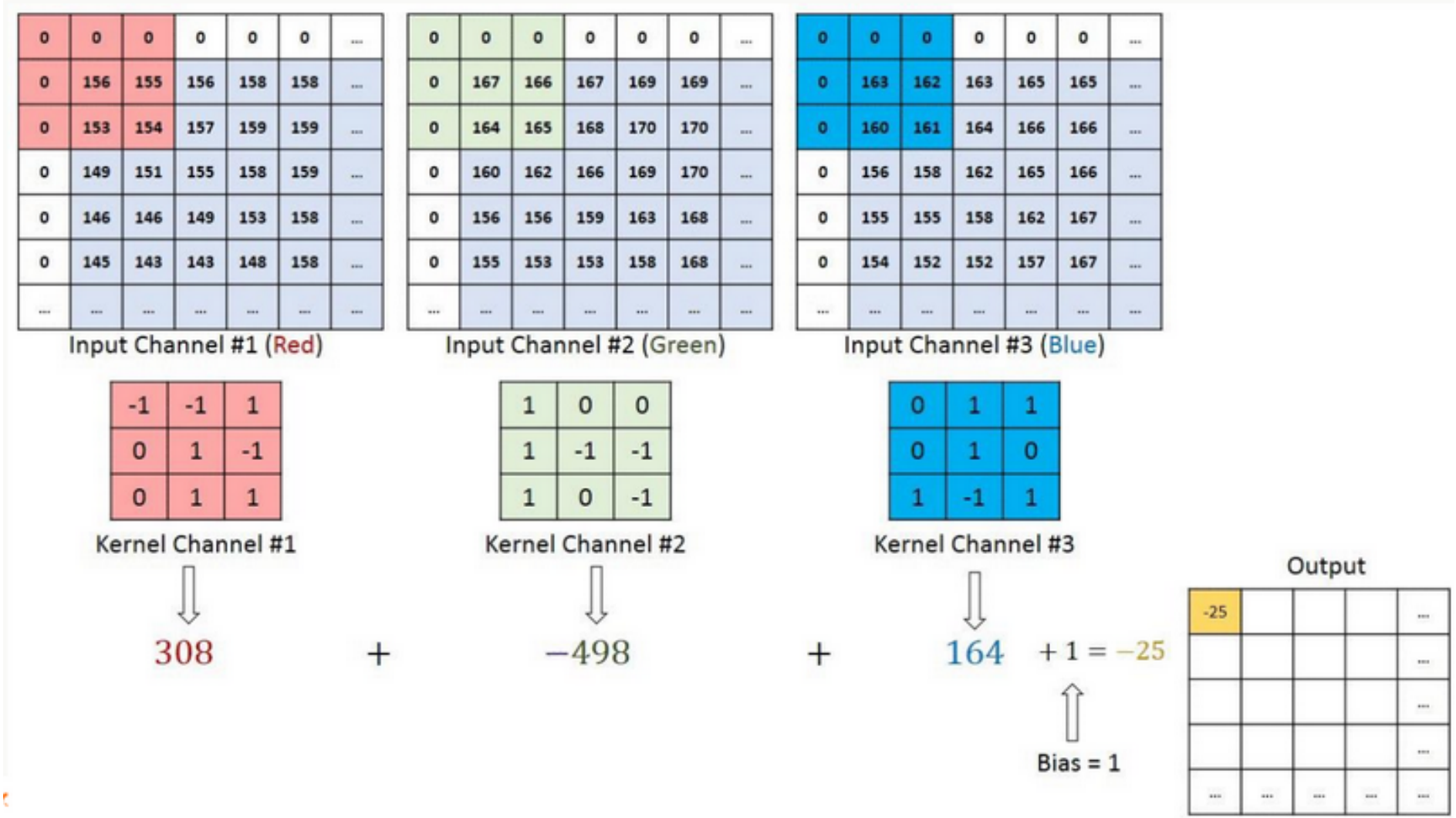
Kernel de Convolución

Al aplicar el kernel de convolución k a una entrada (i, j) de la imágen I, esta entrada se transforma en

$$I_{i,j} = \sum_{x,y=1}^{n} I_{x-i,y-j} k_{x,y}$$



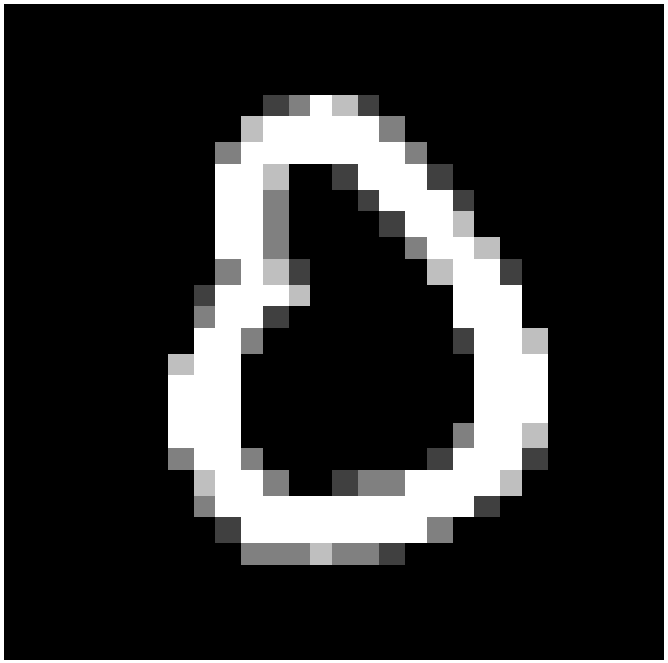
Convolución en varios canales



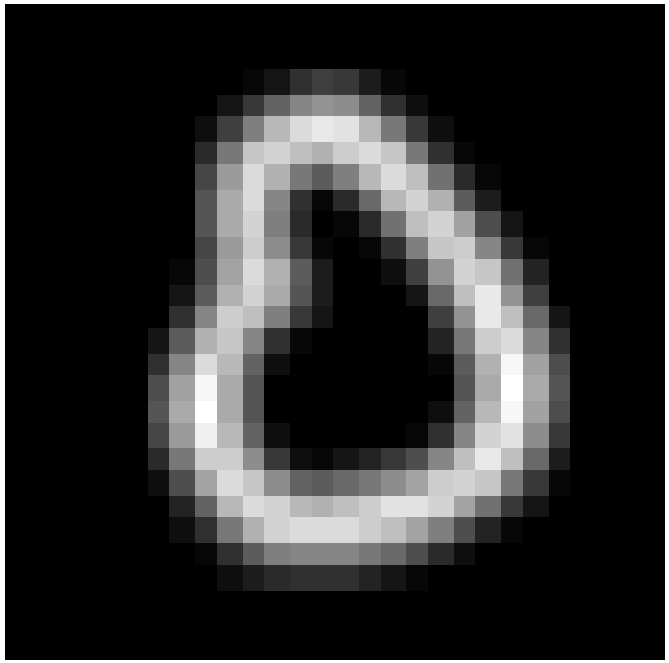
Convolución: Ejemplo

$$K = \frac{1}{9}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

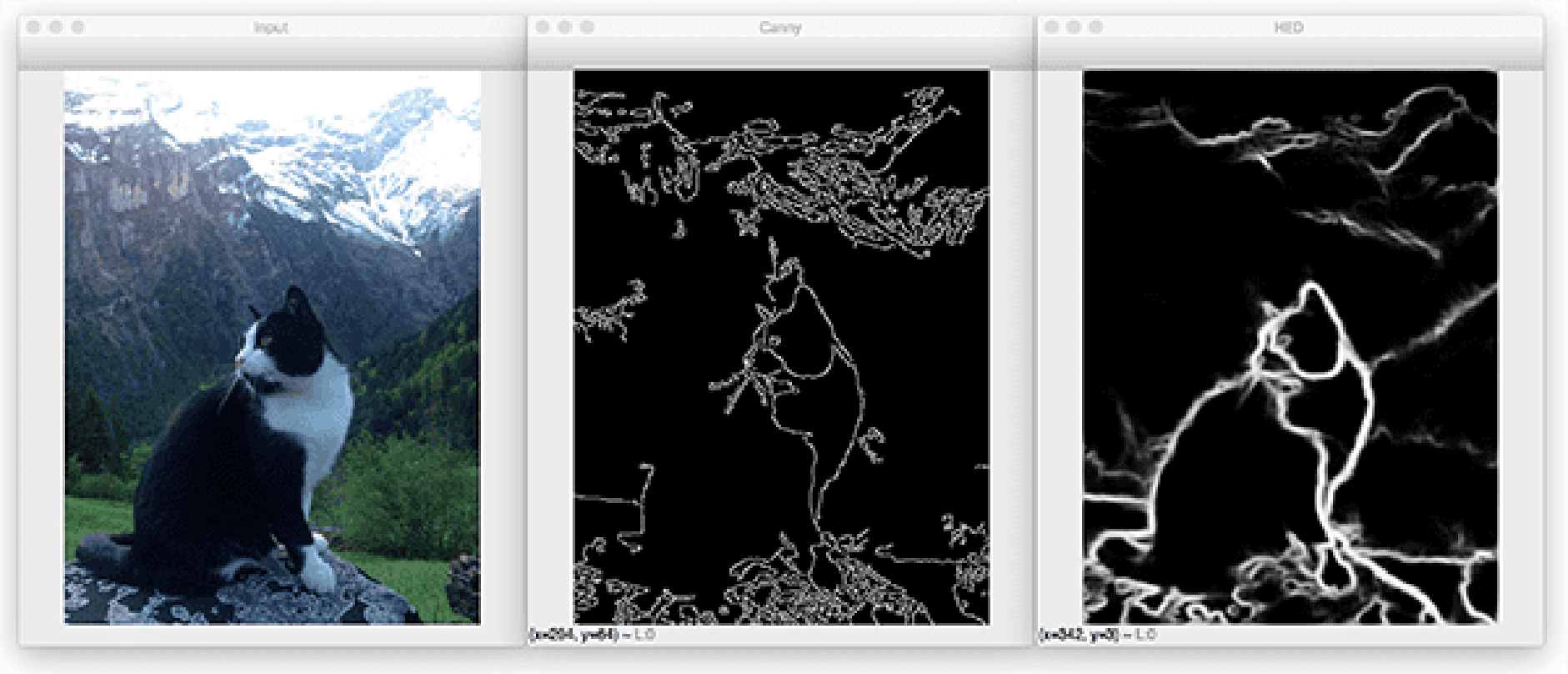
Before



After



State of the Art: Edge Detection



Holistically-Nested Edge Detection

Capas de Pooling

- Uno de los principales objetivos de las CNN es aprender los filtros (kernels).

Capas de Pooling

- Uno de los principales objetivos de las CNN es aprender los filtros (kernels).
- Los filtros sirven para detectar rasgos discriminativos en imágenes.

Capas de Pooling

- Uno de los principales objetivos de las CNN es aprender los filtros (kernels).
- Los filtros sirven para detectar rasgos discriminativos en imágenes.
- El problema con este enfoque es que el proceso es que es sensible a la ubicación donde se encuentren estos rasgos.

Capas de Pooling

- Uno de los principales objetivos de las CNN es aprender los filtros (kernels).
- Los filtros sirven para detectar rasgos discriminativos en imágenes.
- El problema con este enfoque es que el proceso es que es sensible a la ubicación donde se encuentren estos rasgos.
- Una solución a esto es subsamplear estas salidas para hacerlas más robustas al cambio de posición en la imágen.

Capas de Pooling

- Uno de los principales objetivos de las CNN es aprender los filtros (kernels).
- Los filtros sirven para detectar rasgos discriminativos en imágenes.
- El problema con este enfoque es que el proceso es que es sensible a la ubicación donde se encuentren estos rasgos.
- Una solución a esto es subsamplear estas salidas para hacerlas más robustas al cambio de posición en la imágen.
- Aquí es donde entran las capas de **pooling**. El pooling se aplica después de una capa de convolución.

Pooling

La operación de pooling consiste en subsamplear la imágen de entrada. Esta operación se especifica, en vez de aprenderse.

Pooling

La operación de pooling consiste en subsamplear la imágen de entrada. Esta operación se especifica, en vez de aprenderse. Las dos maneras típicas que se usan son:

- Average Pooling: Calcula el valor promedio para cada porción de la imágen.
- Maximum Pooling (Max Pooling): Calcula el valor máximo para cada porción de la imágen.

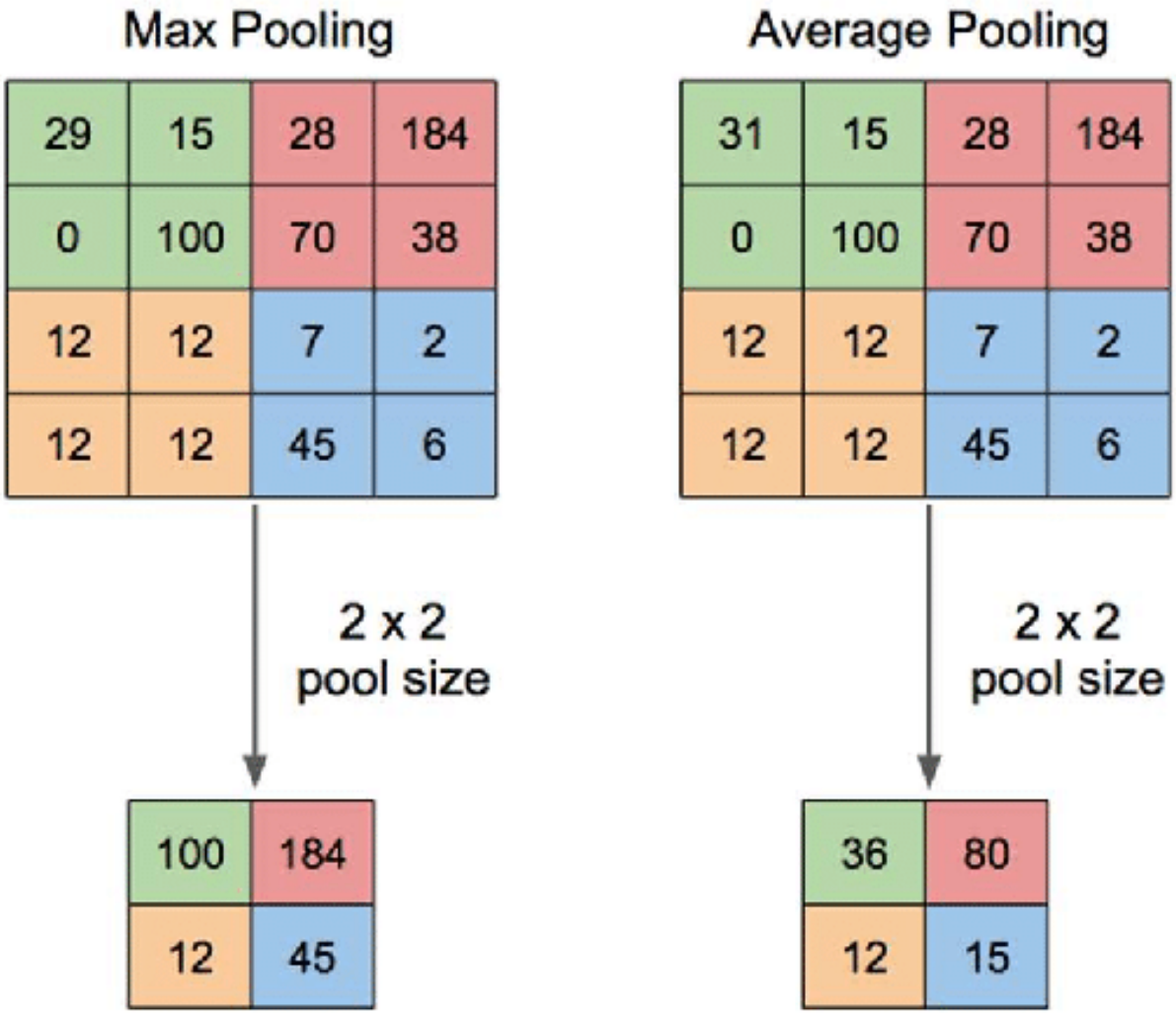
Pooling

La operación de pooling consiste en subsamplear la imágen de entrada. Esta operación se especifica, en vez de aprenderse. Las dos maneras típicas que se usan son:

- Average Pooling: Calcula el valor promedio para cada porción de la imágen.
- Maximum Pooling (Max Pooling): Calcula el valor máximo para cada porción de la imágen.

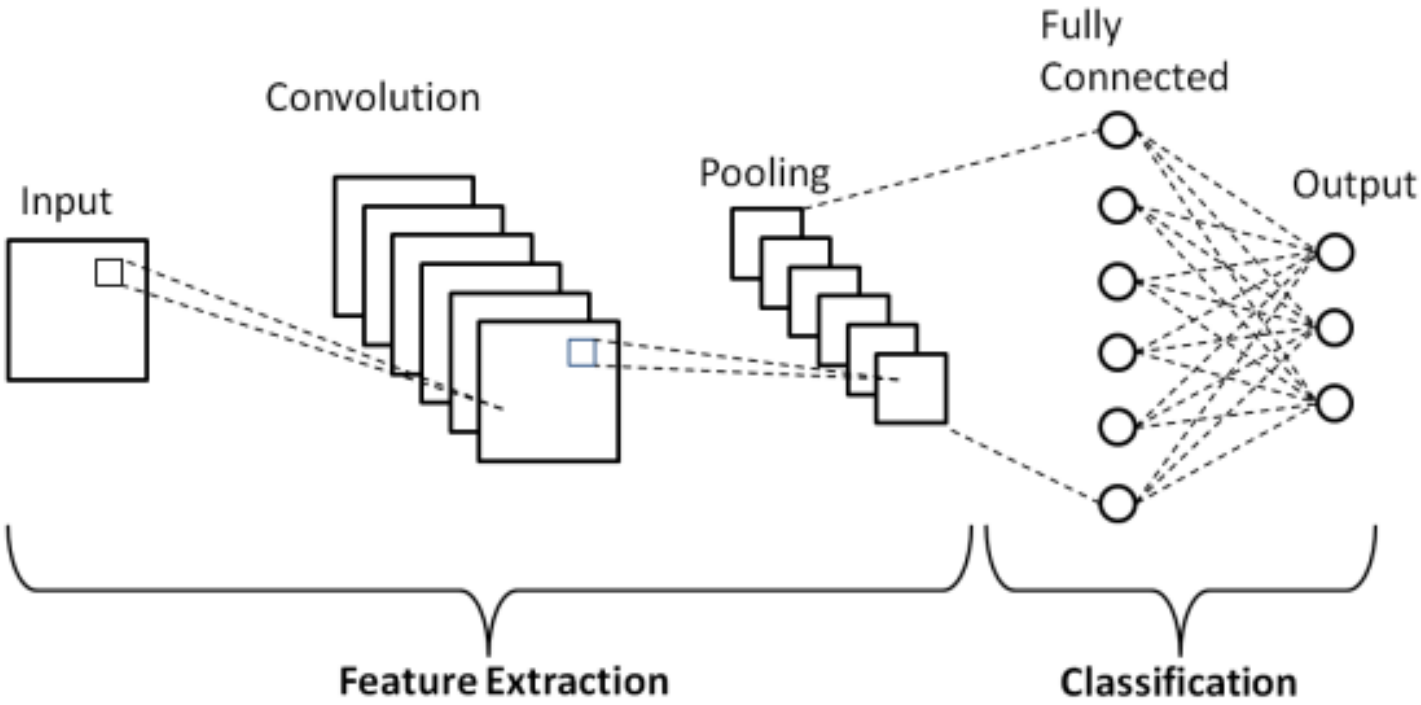
El tamaño de la operación de pooling es más pequeña que el tamaño de la imágen; casi siempre es de 2×2 pixeles con un paso de 2 pixels. En este caso, se reduce el tamaño a la mitad. Por ejemplo, una capa de pooling aplicada a una imágen de 6×6 resultará en una salida de 3×3.

Pooling


Max Pooling
29 15 28 184
0 100 70 38
12 12 7 2
12 12 45 6
2 x 2
pool size
100 184
12 45
Average Pooling
31 15 28 184
0 100 70 38
12 12 7 2
12 12 45 6
2 x 2
pool size
36 80
12 15

Ejemplo de Arquitectura CNN

Consideremos el siguiente ejemplo de CNN



Análisis del ejemplo anterior

- La capa INPUT [32x32x3] tendrá los valores de los pixeles de la imágen, en este caso la imágen es de tamaño 32×32, con tres canales R,G,B.

Análisis del ejemplo anterior

- La capa INPUT [32x32x3] tendrá los valores de los pixeles de la imágen, en este caso la imágen es de tamaño 32×32, con tres canales R,G,B.
- La capa CONV calculará la convolución con cada filtro. Esto resultará en un volumen de imágenes [32x32x6] si es que usamos 12 filtros. A cada salida aplicamos la activación RELU, seguimos teniendo ([32x32x6]).

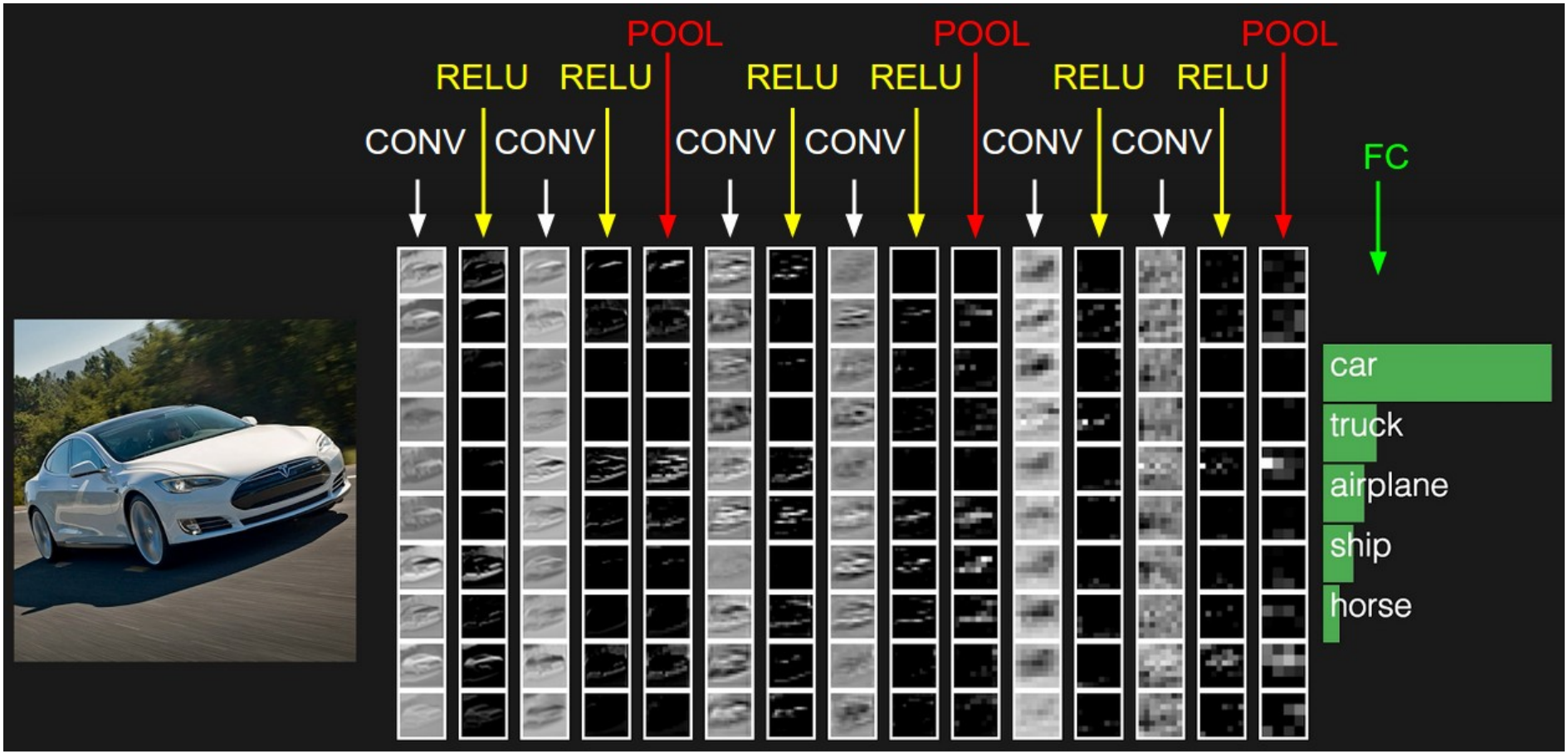
Análisis del ejemplo anterior

- La capa INPUT [32x32x3] tendrá los valores de los pixeles de la imágen, en este caso la imágen es de tamaño 32×32, con tres canales R,G,B.
- La capa CONV calculará la convolución con cada filtro. Esto resultará en un volumen de imágenes [32x32x6] si es que usamos 12 filtros. A cada salida aplicamos la activación RELU, seguimos teniendo ([32x32x6]).
- La capa POOL subsamplea a lo largo del ancho y largo de las imágenes, el resultado es de tamaño [16x16x6].

Análisis del ejemplo anterior

- La capa INPUT [32x32x3] tendrá los valores de los pixeles de la imágen, en este caso la imágen es de tamaño 32×32, con tres canales R,G,B.
- La capa CONV calculará la convolución con cada filtro. Esto resultará en un volumen de imágenes [32x32x6] si es que usamos 12 filtros. A cada salida aplicamos la activación RELU, seguimos teniendo ([32x32x6]).
- La capa POOL subsamplea a lo largo del ancho y largo de las imágenes, el resultado es de tamaño [16x16x6].
- El volumen de datos anterior se aplana y entra a la red FC, la cuál calculará los scores de clase resultando en un volumen de tamaño [`None`x3x1].

Clasificación



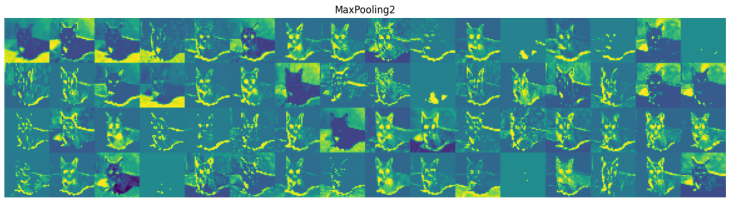
Ejemplo: Cats vs Dogs

Un gato

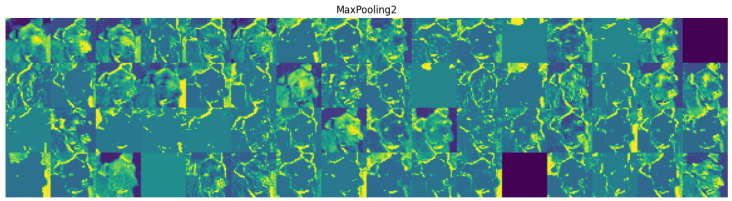
Un perro

Ejemplo: Cats vs Dogs

Un gato

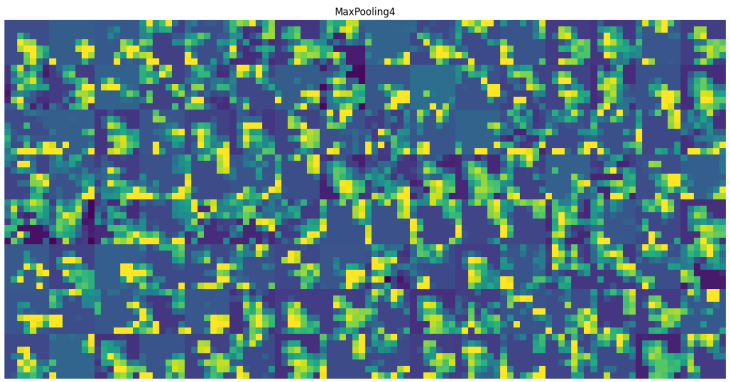


Un perro



Ejemplo: Cats vs Dogs

Un gato



Un perro

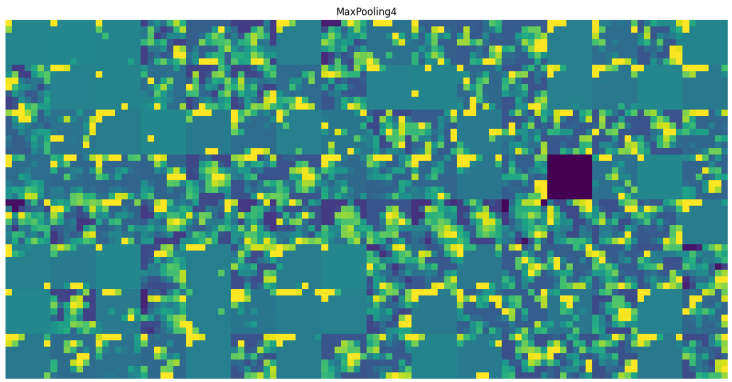
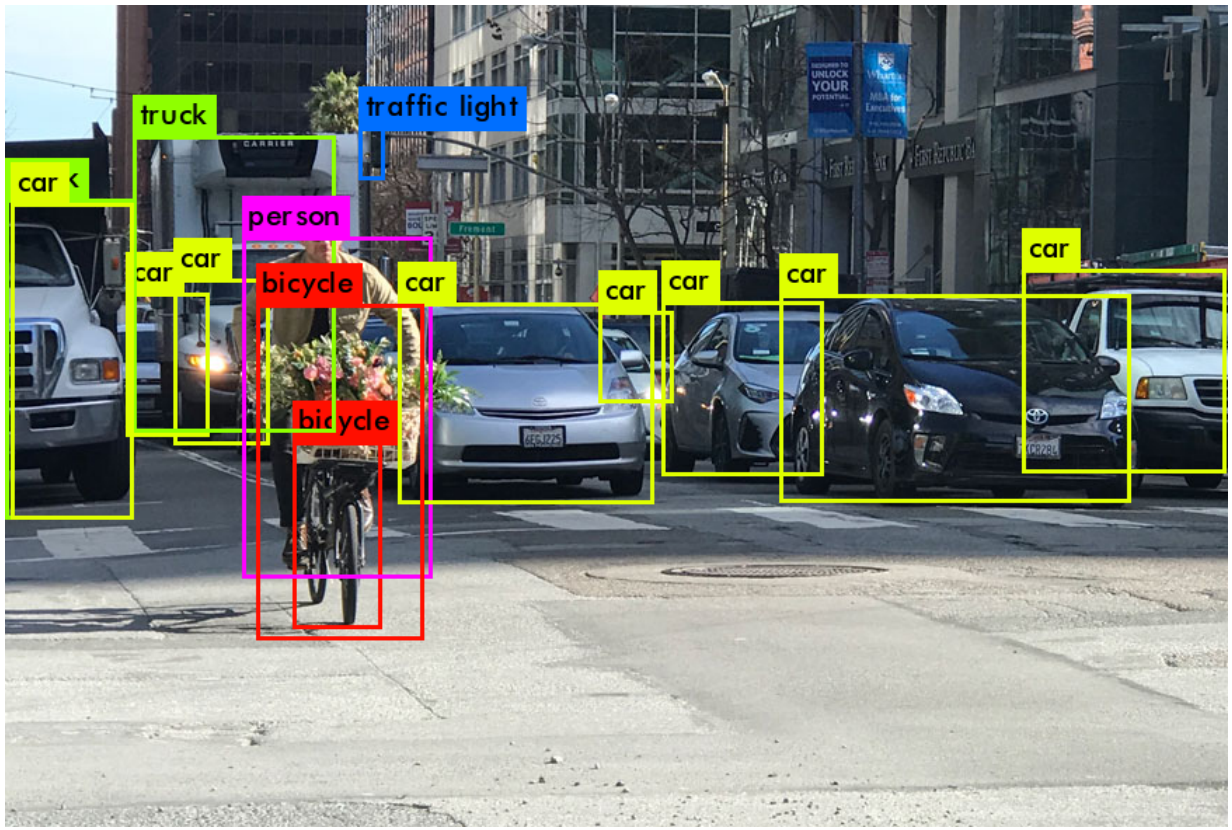


Table of Contents

Real-Time Object Detection

YOLO: You only look once



Video demo
Source