



Redes Neuronales Convolucionales (CNN)

Diego Armando Pérez Rosero ·, Santiago Pineda Quintero

Departamento de Ingeniería Eléctrica, Electrónica y Computación
Universidad Nacional de Colombia
Manizales, Colombia
Octubre 2023

Contenido



- 1 ¿Qué es una CNN?
- 2 Hiperparámetros de la Capa de Convolución
- 3 Capa de Pooling
- 4 Arquitectura Clásica de una CNN
- 5 Observaciones

¿Qué es una CNN?



Una red neuronal convolucional (CNN) es un tipo de red neuronal artificial utilizada en el campo de la visión artificial para **tratar** de emular la visión humana; pues este tipo de redes se caracterizan por aprovechar muy bien el espacio geométrico de una imagen.

¿Cómo funciona una CNN?



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Este tipo de redes neuronales son llamadas convolucionales; ya que lo que hacen es aplicar operaciones de convolución sobre una imagen, es decir, escanean la imagen píxel por píxel desde la esquina superior izquierda hasta la esquina inferior derecha, con el fin de ir extrayendo características de la imagen: Bordes, formas, texturas, colores, etc.

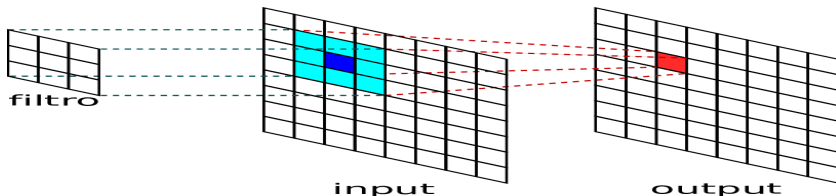


Figure: Operación de Convolución

Operación de Convolución



UNIVERSIDAD
NACIONAL
DE COLOMBIA

La operación de convolución consiste en escanear una imagen a través de un filtro o kernel e ir realizando operaciones de producto punto, para finalmente obtener un mapa de características de la imagen.

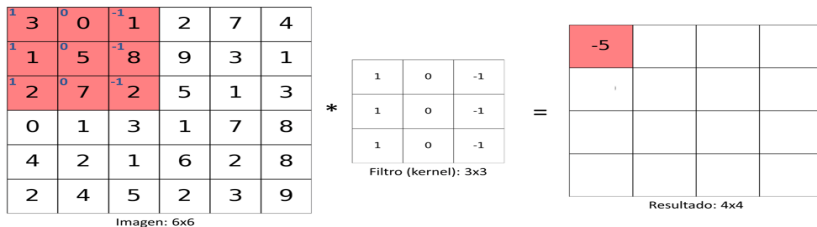


Figure: Capa Convolutiva

Kernel o Filtro



- Es una matriz cuadrada que escanea la imagen desde la esquina superior izquierda hasta la esquina inferior derecha haciendo operaciones de convolución, con el fin de encontrar los patrones relevantes de la imagen.
- Los valores del kernel se elijen de acuerdo a las características que se quieren extraer de la imagen de acuerdo a la tarea a realizar (Ya hay kernels predefinidos para detectar ciertos patrones en las imágenes).

Kernels para Detectar Bordes Horizontales y Verticales



UNIVERSIDAD
NACIONAL
DE COLOMBIA

1	1	1
0	0	0
-1	-1	-1

a)

1	0	-1
1	0	-1
1	0	-1

b)

Figure: Tipos de Kernels

Padding



- aplicar la operación de convolución, el kernel escanea una sola vez los píxeles de las esquinas de la imagen, mientras que los demás píxeles los escanea varias veces; lo cual lleva a que se ignore información posiblemente importante de las esquinas de la imagen.
- El padding consiste en rellenar los bordes de las imágenes con ceros (Para no afectar la información de la imagen), para lograr que los píxeles de las esquinas de la imagen sean escaneados más veces al efectuar la convolución.

Padding



UNIVERSIDAD
NACIONAL
DE COLOMBIA

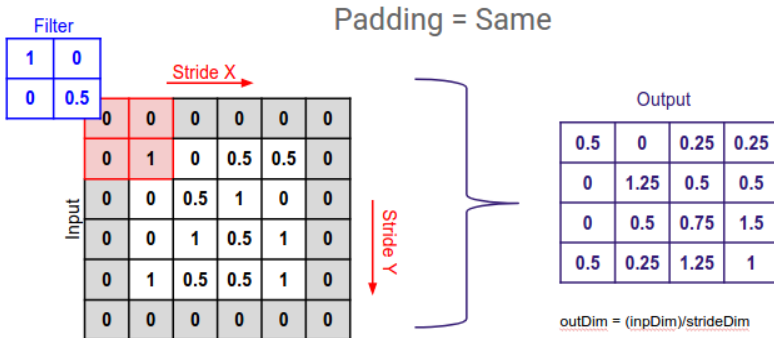
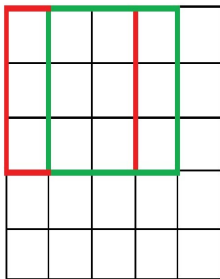


Figure: Padding

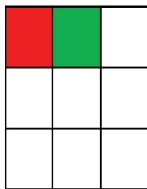
Stride

El stride es la cantidad de pasos que da el kernel cuando está escanando la imagen.

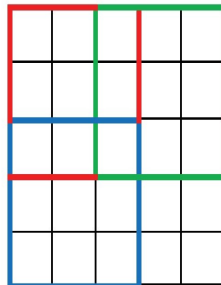
Convolution
with Stride=1



Output



Convolution
with Stride=2



Output

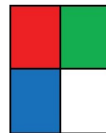


Figure: Stride

Stride



- Notese que un stride muy grande reduce considerablemente la resolución del mapa de características.
- Un stride muy grande implica disminuir costo computacional (pues es menor la cantidad de operaciones de producto punto que se realizan), pero así mismo implica que el mapa de características obtenido no contiene de forma muy específica las características de la imagen, y viceversa. Es por esto que es importante de acuerdo a la aplicación a que se le debe dar prioridad.

Capa de Pooling



- Tener múltiples filtros en cada capa convolucional, hace que la complejidad de la red neuronal convolucional aumente, es decir, que se tenga una red con muchos parámetros, lo cual hace que estos modelos sean muy costosos computacionalmente.
- El objetivo de las capas de pooling es disminuir la resolución de los mapas de características, para de este modo reducir el número de parámetros del modelo y así mismo el costo computacional demandado por el modelo.

Capa de Pooling



UNIVERSIDAD
NACIONAL
DE COLOMBIA

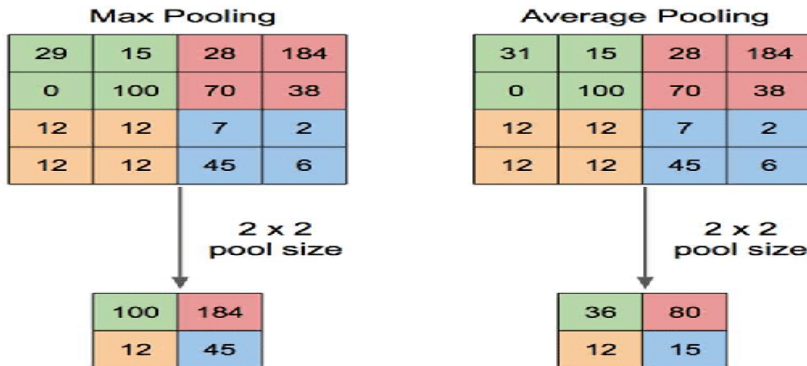


Figure: Average Pooling y Max Pooling

¿Con las Capas de Pooling se Pierde Información de la Imagen?



Por el contrario, con una capa de pooling se está disminuyendo el ruido presente en la imagen; pues se está consolidando en un solo píxel la información más relevante de toda una región de la imagen.

Arquitectura Clásica de una CNN



UNIVERSIDAD
NACIONAL
DE COLOMBIA

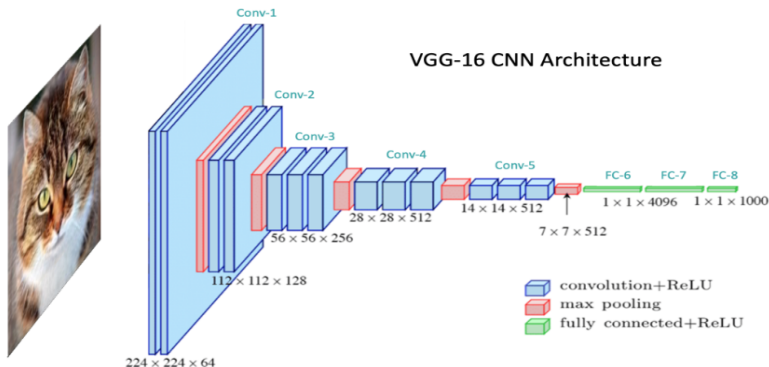


Figure: Arquitectura clásica de una CNN

Arquitectura Clásica de una CNN



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Las CNN tienen la siguiente arquitectura: Varias capas de convolución (para extraer las características de la imagen), luego, va una capa de pooling (para reducir la complejidad del modelo), repitiendo esto múltiples veces. Posteriormente, se pasa el mapa de características por una capa de flatten y por último, el vector aplanado obtenido se pasa por una capa densa encargada de realizar la predicción según la tarea a realizar.

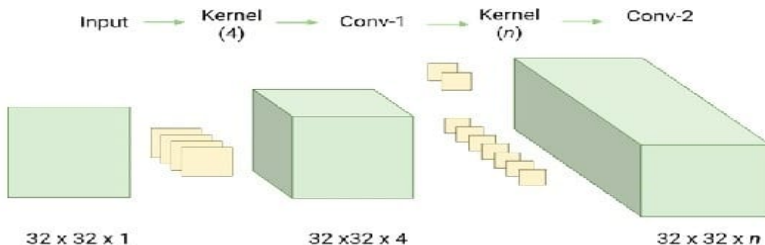
Arquitectura Clásica de una CNN



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Cada capa de convolución aumenta la profundidad del mapa de características; por tanto, entre más capas convolucionales y más cantidad de filtros en estas capas, mayor será la profundidad del mapa de características obtenido.

CNN - Convolución



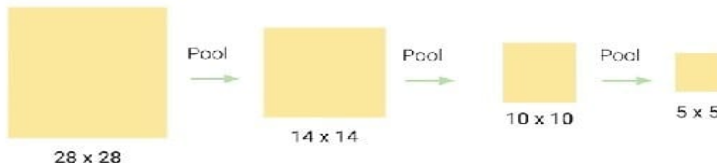
Arquitectura Clásica de una CNN



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Cada capa de pooling disminuye la resolución del mapa de características; por tanto, entre más capas de pooling, menor será la resolución del mapa de características.

CNN - Pooling

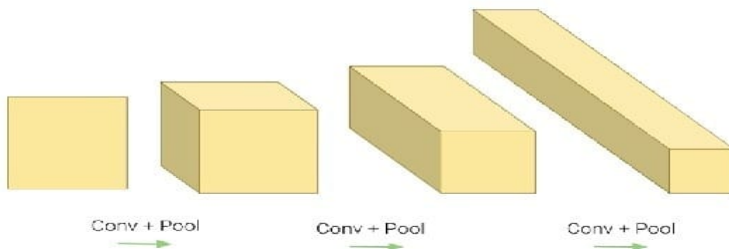


Arquitectura Clásica de una CNN



UNIVERSIDAD
NACIONAL
DE COLOMBIA

CNN - Todo junto

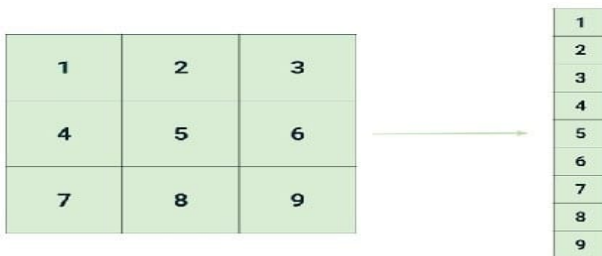


Capa de Flatten



Luego, cuando se tenga un mapa de características muy profundo y con muy baja resolución, este se pasa por una capa de flatten; para por último pasar dicho vector aplanado por una capa densa.

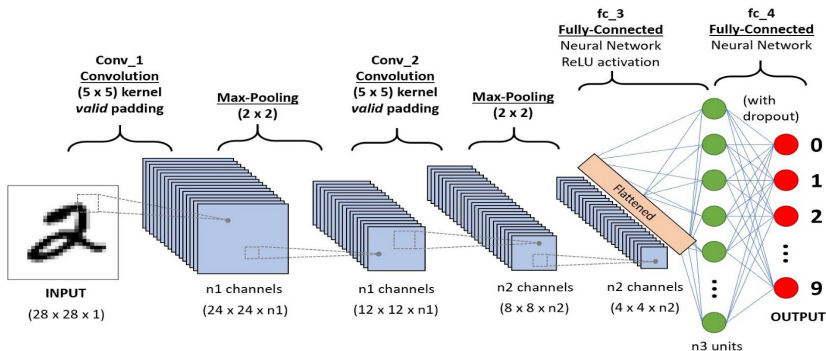
Flatten



Conclusión



UNIVERSIDAD
NACIONAL
DE COLOMBIA

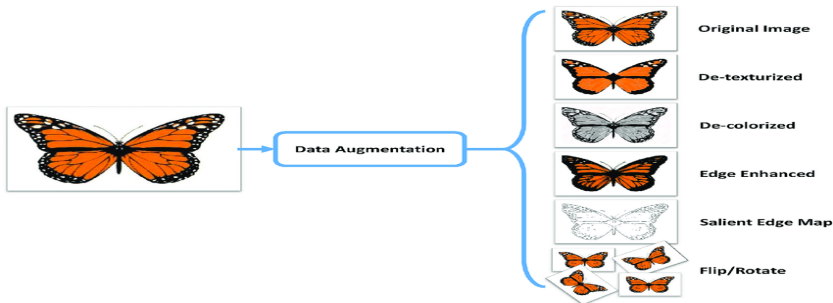


Observaciones



UNIVERSIDAD
NACIONAL
DE COLOMBIA

A menudo, el proceso de etiquetar imágenes puede ser tan engorroso, costoso y demorado, que se debe recurrir a técnicas como **Data Augmentation**, más sin embargo a la hora de utilizar este tipo de técnicas, tener cuidado en no caer en sobreentrenamiento.



Observaciones



- Siempre que el patrón color no sea relevante a la hora de llevar a cabo una tarea de visión por computador, trabajar con imágenes en escala de grises; pues esto reducirá considerablemente el costo computacional.
- A la hora de definir la arquitectura de una CNN, generalmente se comienza con arquitecturas sencillas, y en caso de caer en underfitting, se aumenta la complejidad de la arquitectura, mientras que si se cae en overfitting, se suele disminuir la complejidad de la arquitectura o utilizar otras técnicas para combatir el overfitting como: Capas de dropout y regularizadores.



Thanks!