

Apuntes de Clase: Redes Neuronales Convolucionales

Tomás Granados Preciado
Escuela de Ingeniería en Computación
Instituto Tecnológico de Costa Rica
Cartago, Costa Rica

Abstract—Este documento compila los apuntes de la clase de Inteligencia Artificial centrados en Redes Neuronales Convolucionales (CNN), integrando conceptos teóricos, arquitectura de capas convolucionales y de pooling, capas fully connected, pesos compartidos y un repaso de arquitecturas clásicas como LeNet, AlexNet, ZFNet, GoogleNet, VGG16, ResNet y DenseNet.

Index Terms—Redes Neuronales Convolucionales, CNN, Pooling, Arquitectura, LeNet, AlexNet

I. DUDAS DE LA TAREA 2

El grupo no presentó dudas significativas sobre los cambios necesarios para adaptar la red a problemas de regresión.

II. RECOMENDACIONES DEL PROYECTO 2

- Generar las imágenes del dataset antes del entrenamiento para evitar procesos repetitivos y costosos en tiempo de ejecución.
- Probar la red con un subconjunto reducido de datos para evaluar resultados y optimizar recursos computacionales.

III. REDES NEURONALES CONVOLUCIONALES

A. Limitaciones de las Fully Connected Networks

En un esquema fully connected, cada neurona de una capa está conectada a todas las neuronas de la capa anterior, lo cual provoca un número elevado de parámetros y hace el proceso poco escalable cuando las entradas (por ejemplo, imágenes con múltiples canales) son de alta dimensión.

B. Arquitectura de las CNN

1) *Capa Convolutiva*: Cada neurona procesa una región local (receptive field) de la capa anterior mediante filtros (kernels) de tamaño $k \times k$. Los parámetros clave son:

- **Stride (s)**: número de pasos con que se mueve el kernel.
- **Padding (p)**: píxeles añadidos alrededor de la imagen, calculado como $p = \frac{k-1}{2}$.

- **Dimensión de salida**: dado un mapa de entrada de tamaño m , la salida es:

$$\left\lfloor \frac{m - k + 2p}{s} \right\rfloor + 1.$$

2) *Pesos Compartidos*: Para reducir drásticamente el número de parámetros, se usa el mismo conjunto de pesos (filtros) en todas las posiciones espaciales de la imagen.

3) Tareas Principales:

- Clasificación de imágenes
- Segmentación de objetos e instancias
- Procesamiento y detección de características

4) *Capa de Pooling*: Realiza un downsampling espacial para disminuir el tamaño de los mapas de características:

- **Kernel de pooling** de tamaño k y stride s .
- **Dimensión de salida**: $W' = \left\lfloor \frac{W-k}{s} \right\rfloor + 1$, análogo para H .
- **Técnicas**: Max pooling, Average pooling (histórico), L2-Norm pooling.

5) *Capa Fully Connected*: Al terminar las etapas convolucionales y de pooling, las activaciones se aplanan y se conectan a una o varias capas densas que calculan las probabilidades de pertenencia a cada clase.

IV. ARQUITECTURAS CONVOLUCIONALES

- **LeNet (1998)**: arquitectura pionera de 5 capas (2 conv, 2 pooling, 1 fully connected) diseñada por Yann LeCun.
- **AlexNet (2012)**: introdujo filtros grandes (11×11, 5×5, 3×3) y 5 capas convolucionales; popularizó el uso de ReLU y dropout [2].
- **ZFNet (2013)**: redujo los tamaños de kernel de AlexNet a 7×7 y luego 3×3 para mejorar la localización de características.

- **GoogleNet/Inception (2014)**: utiliza módulos Inception para combinar múltiples tipos de convolución en paralelo, reduciendo parámetros de 60M a 4M [3].
- **VGG16 (2014)**: apila múltiples convoluciones pequeñas (3×3) para profundizar la red, con 16 capas de profundidad [4].
- **ResNet (2016)**: introduce conexiones residuales para entrenar redes muy profundas sin degradación del gradiente [5].
- **DenseNet (2017)**: conecta cada capa con todas las anteriores, facilitando la reutilización de características y reduciendo parámetros.

V. CONCLUSIONES

Las CNN aprovechan la localidad espacial y el concepto de pesos compartidos para procesar imágenes de forma eficiente. La combinación de capas convolucionales, pooling y fully connected permite resolver tareas de clasificación y segmentación con alto rendimiento. Las arquitecturas evolutivas (LeNet a DenseNet) demuestran la mejora continua mediante ajustes en la profundidad, el tamaño de filtros y las conexiones internas.

REFERENCIAS

REFERENCES

- [1] Y. LeCun *et al.*, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [2] A. Krizhevsky, I. Sutskever y G. Hinton, "Imagenet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, pp. 1097–1105, 2012.
- [3] C. Szegedy *et al.*, "Going deeper with convolutions," *Proc. CVPR*, pp. 1–9, 2015.
- [4] K. Simonyan y A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, 2014.
- [5] K. He *et al.*, "Deep residual learning for image recognition," *Proc. CVPR*, pp. 770–778, 2016.