

Redes Neuronales Convolucionales (CNN)

Jose Pablo Granados Siles
Instituto Tecnológico de Costa Rica
Apuntes del 06 de mayo del 2025

Abstract—Las Redes Neuronales Convolucionales (CNN) son un tipo especial de red neuronal diseñada específicamente para el procesamiento de datos con estructura de cuadrícula, como imágenes. Este artículo presenta sus principales características, compara su funcionamiento con redes totalmente conectadas, describe sus componentes clave, expone conceptos fundamentales y resume las arquitecturas históricas más relevantes, así como sus aplicaciones.

Index Terms—CNN, redes neuronales, visión por computadora, arquitectura, clasificación de imágenes

I. INTRODUCCIÓN

Las Redes Neuronales Convolucionales (CNN) están optimizadas para trabajar con datos espaciales y son especialmente efectivas en tareas de visión por computadora. A diferencia de las redes tradicionales, aprovechan la estructura espacial local de los datos para extraer características relevantes.

A. Características Principales

- Transformación durante el forward pass
- Predicción de valores de salida basados en características espaciales
- Manejo eficiente de datos multidimensionales (imágenes a color)
- Ejemplo: Una imagen de $32 \times 32 \times 3$ tiene 3072 píxeles de entrada

II. COMPARACIÓN: FULLY CONNECTED VS CONVNET

A. Redes Fully Connected

- Requieren gran cantidad de parámetros
- Entrenamiento lento
- Rendimiento limitado con datos espaciales
- No aprovechan la estructura local de las imágenes

B. Redes Convolucionales

- Neuronas organizadas en 3 dimensiones (alto, ancho, profundidad)
- Conexiones locales entre capas
- Funcionan como extractores de características
- Reducción eficiente de dimensionalidad
- Capa de salida típica: $1 \times 1 \times N$ (donde N es el número de clases)

III. COMPONENTES PRINCIPALES

A. Capa Convolutiva

- Función: Extrae características locales mediante filtros
- Filtros con tamaño y *stride* definidos
- Parámetros aprendibles
- Función de activación (típicamente ReLU)

B. Capa de Pooling

- Función: Reducción de dimensionalidad espacial
- Tipos: Max Pooling, Average Pooling, L2 Norm Pooling
- No tiene parámetros aprendibles
- Mantiene la profundidad de los canales

C. Capa Fully Connected (FC)

- Función: Clasificación final
- Conexión completa con la capa anterior
- Parámetros aprendibles
- Activación: softmax para clasificación

IV. CONCEPTOS FUNDAMENTALES

$$\text{INPUT} \rightarrow [\text{conv} \rightarrow \text{relu}]^n \rightarrow [\text{pool}]^m \rightarrow [\text{fc} \rightarrow \text{relu}]^k \rightarrow \text{fc}$$
$$3 \geq n \geq 0, m \geq 0, k \geq 0$$

A. Campo Receptivo

- Área de la imagen que afecta a una neurona en la capa de salida
- Aumenta progresivamente a través de las capas
- Cálculo: Tamaño del filtro \times Número de capas convolucionales

B. Padding

- Propósito: Control del tamaño de salida
- Tipos: Valid (sin padding), Same (mantener dimensiones)
- Beneficios: Preserva información en bordes, mejora rendimiento

C. Reglas Prácticas

- Usar campos receptivos pequeños: 3×3 o 5×5 con *stride* 1
- Padding de ceros para conservar dimensiones:
 - $F = 3, P = 1$
 - $F = 5, P = 2$

V. ARQUITECTURAS HISTÓRICAS

A. *LeNet-5*

- Primera CNN exitosa
- 5 capas convolucionales, 2 de pooling
- 2 capas fully connected, 1 de salida

B. *AlexNet*

- Avance significativo en 2012
- Clasificación de 1000 clases
- Arquitectura paralela para GPUs
- Uso de ReLU y Dropout

C. *ZFNet*

- Mejora de AlexNet
- Ajuste de hiperparámetros
- Mayor comprensión de características aprendidas

D. *GoogleNet (Inception)*

- Reducción drástica de parámetros (4M vs 60M de AlexNet)
- Módulos Inception
- Reemplazo de capas FC por *average pooling*

E. *VGG16*

- Arquitectura profunda, bloques 3×3
- Estructura uniforme y simple
- Excelente rendimiento en *transfer learning*

F. *ResNet*

- Conexiones residuales (*skip connections*)
- Solución al desvanecimiento del gradiente
- Entrenamiento de redes muy profundas

VI. APLICACIONES

- Clasificación de imágenes
- Segmentación semántica
- Segmentación de instancias
- Detección de objetos
- Procesamiento de imágenes
- Reconocimiento facial
- Análisis médico de imágenes