

UACM

Universidad Autónoma
de la Ciudad de México

NADA HUMANO ME ES AJENO

Redes Neuronales

Sabino Miranda

Redes Neuronales Convolucionales

Redes Neuronales Convolucionales (1)

- Las Redes Neuronales Convolucionales (Convolutional Neural Networks, CNN) son un tipo de red neuronal diseñada para procesar datos con una estructura de tipo espacial o temporal, como imágenes, videos o secuencias.
- Estas redes están especialmente indicadas para manejar datos que poseen relaciones locales significativas, aprovechando su capacidad para identificar patrones jerárquicos en los datos de entrada.
- Su principal característica es el uso de la operación de convolución (correlación cruzada), que les permite detectar patrones locales (como bordes, texturas y formas) para luego combinarlos y reconocer estructuras más complejas.

Redes Neuronales Convolucionales (2)

- Las redes convolucionales son redes neuronales que utilizan la convolución en lugar de la multiplicación de matrices general en al menos una de sus capas.

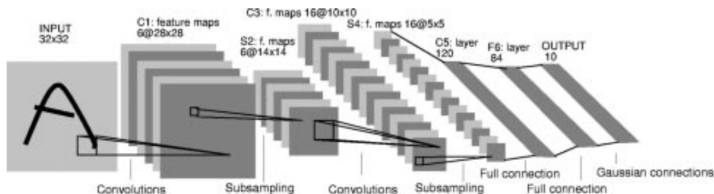


Figura: Ejemplo de CNN: Arquitectura de Lenet-5 (1)

(1) LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>

Redes Neuronales Convolucionales (revisión) (1)

- Los éxitos de algunas arquitecturas de redes neuronales convolucionales en concursos de clasificación de imágenes después de 2011 llevaron a una mayor atención al campo del aprendizaje profundo.
- Los puntos de referencia como ImageNet con una tasa de error de clasificación de más del 25 % (en los 5 mejores sistemas) se redujeron a menos del 4 % en los años entre 2011 y 2015.
- Las redes neuronales convolucionales son muy adecuadas para el proceso de ingeniería de características jerárquicas con profundidad.

Redes Neuronales Convolucionales (revisión) (2)

- La motivación inicial para las redes neuronales convolucionales se derivó de los experimentos de David Hubel y Torsten Wiesel en la corteza visual de un gato.
- La corteza visual tiene pequeñas regiones de células que son sensibles a regiones específicas en el campo visual. Es decir, si se excitan áreas específicas del campo visual, entonces esas células en la corteza visual también se activarán.
- Las células excitadas también dependen de la forma y la orientación de los objetos en el campo visual. Por ejemplo, los bordes verticales hacen que algunas células neuronales se exciten, mientras que los bordes horizontales hacen que otras células neuronales se exciten.

Redes Neuronales Convolucionales (revisión) (3)

- Las células están conectadas mediante una arquitectura en capas, y este descubrimiento llevó a la conjetura de que los mamíferos utilizan estas diferentes capas para construir porciones de imágenes en diferentes niveles de abstracción.
- Desde el punto de vista del aprendizaje automático, este principio es similar al de la extracción de características jerárquicas.
- Basándose en estas inspiraciones biológicas, el primer modelo neuronal fue el neocognitrón.
- Basándose en el neocognitrón, se desarrolló una de las primeras arquitecturas convolucionales, conocida como LeNet-5. Esta red fue utilizada por los bancos para identificar números escritos a mano en cheques.

Redes Neuronales Convolucionales (revisión) (4)

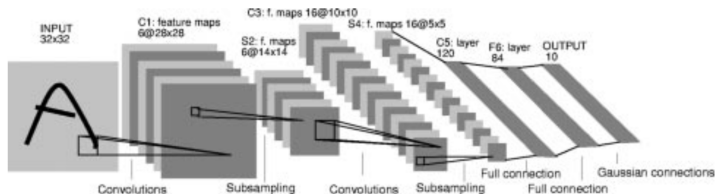


Figura: Arquitectura de Lenet-5, un red neuronal convolucional usada para reconocimiento de dígitos (1). Cada plano es un mapa de características.

(1) LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>

Redes Neuronales Convolucionales (revisión) (5)

- Las redes neuronales convolucionales han sido ganadoras constantes del concurso ImageNet desde 2012, cuando se propuso AlexNet.

Redes Neuronales Convolucionales (revisión) (6)

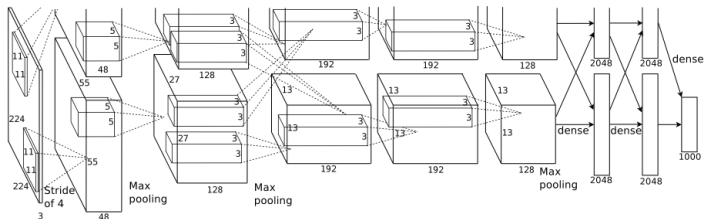


Figura: Una ilustración de la arquitectura de AlexNet, CNN (2). La entrada de la red tiene 150,528 dimensiones y la cantidad de neuronas en las capas restantes de la red está dada por 253,440 – 186,624 – 64,896 – 64,896 – 43,264 – 4096 – 4096 – 1000.

(2) Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional

Neural Networks. Communications of the ACM, 60(6), 84–90. <https://doi.org/10.1145/3065386>

Componentes de una CNN (1)

- ❶ **Capa de Convolución:** Realiza la operación de convolución, donde un filtro (*kernel*) se desliza sobre la entrada para extraer características locales.



Figura: (3) (padding=0, stride=1) Convolución de un kernel 3x3 (azul intenso, subíndices indican los pesos del kernel) sobre una entrada de 5x5 (azul claro), salida (mapa de características, *feature map*) (verde) (i=5, k=3, s=1, p=0)

Componentes de una CNN (2)

*Imágenes obtenidas de: (3) Dumoulin, V., Visin, F. (2018). A guide to convolution arithmetic for deep learning. arXiv. <https://arxiv.org/abs/1603.07285>

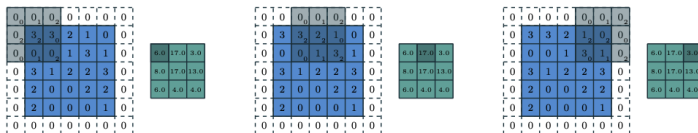


Figura: (padding=1, stride=2) Convolución de un kernel 3x3 (azul intenso, subíndices indican los pesos del kernel) sobre una entrada de 5x5 (azul claro), salida (mapa de características, *feature map*) (verde) (i=5, k=3, s=2, p=1)

Componentes de una CNN (3)

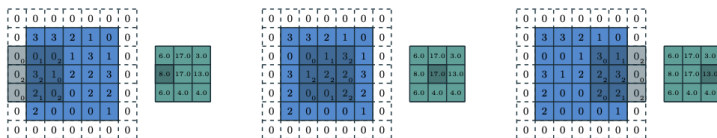


Figura: (padding=1, stride=2) Convolución de un kernel 3x3 (azul intenso, subíndices indican los pesos del kernel) sobre una entrada de 5x5 (azul claro), salida (mapa de características, *feature map*) (verde) (i=5, k=3, s=2, p=1)

- ② **Capa de Activación:** Introduce no linealidad mediante funciones como ReLU ($f(x) = \max(0, x)$), permitiendo que la red aprenda relaciones complejas.

Componentes de una CNN (4)

③ **Capa de Pooling:** Reduce la dimensión espacial de los mapas de características para disminuir el número de parámetros y la complejidad computacional.

- **Average pooling:** Calcula el promedio en una región.
- **Max pooling:** Toma el valor máximo en una región.

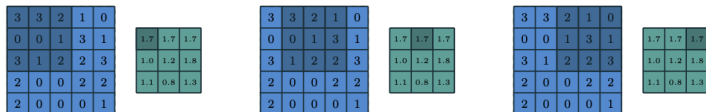


Figura: (padding=0, stride=1) Cálculo de los valores de salida de 3x3 (verde) con una operación *average pooling* sobre una entrada de 5x5 (azul claro) con una ventana de 3x3 (azul intenso) usando *strides* de 1x1

Componentes de una CNN (5)



Figura: (padding=0, stride=1) Cálculo de los valores de salida de 3x3 (verde) con una operación *max pooling* sobre una entrada de 5x5 (azul claro) con una ventana de 3x3 (azul intenso) usando *strides* de 1x1

- ④ **Capas Completamente Conectadas (*Fully Connected*):** Al final de la red, las características extraídas se transforman en un vector unidimensional que pasa por una o más capas densas para hacer predicciones.

Ejemplo de red convolucional en PyTorch

Clases:

- ❶ **Conv2d**: Aplica una convolución 2D sobre una señal de entrada compuesta por varios planos de entrada.
- ❷ **MaxPool2d**: Aplica una agrupación máxima 2D sobre una señal de entrada compuesta por varios planos de entrada.

Ejemplo en PyTorch: Creación de red convolucional para clasificación de texto

Consultar el Notebook: 18_CNN_Cov2d_WE.ipynb

Referencias (1)

- ❶ Deep Learning. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. MIT Press, 2016.
<http://www.deeplearningbook.org>
- ❷ Dive into Deep Learning. Aston Zhang, Zachary C. Lipton, Mu li, and Alexander J. Smola. Cambridge University Press, 2023. <https://d2l.ai>
- ❸ Neural Networks and Deep Learning A Textbook (2nd Edition). Charu C. Aggarwal. Springer, 2023.
<https://doi.org/10.1007/978-3-031-29642-0>
- ❹ Deep Learning: Foundations and Concepts. Christopher M. Bishop and Hugh Bishop. Springer, 2024.
<https://doi.org/10.1007/978-3-031-45468-4>

Referencias (2)

- ⑤ PyTorch documentation.
<https://pytorch.org>
- ⑥ Numpy documentation.
<https://numpy.org>
- ⑦ Python documentation.
<https://www.python.org>