AlexNet

AlexNet, descrita y utilizada en las referencias siguientes, ImageNet et al. (2020), Russakovsky et al. (2015), Krizhevsky et al. (2012), BVLC (2020), Oráculo (2021), es una de las primeras redes convolucionales capaz de reconocer una imagen y clasificarla por su contenido. En 2012, presentada por primera vez en el concurso ILSVRC - ImageNet Large Scale Visual Recognition Challenge - por Alex Krizhevsky junto a su director de tesis, logró reducir el error top-5-error del 25 % hasta el 17 %.

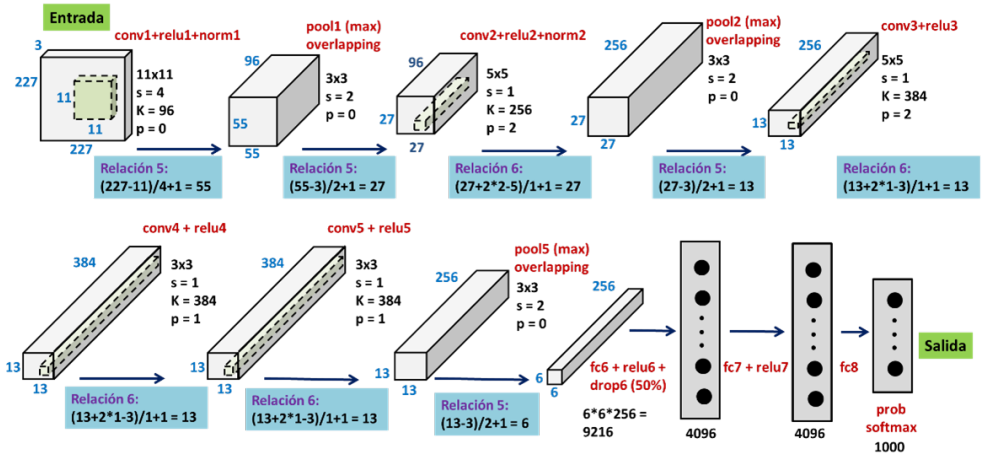

Figura 1: Estructura de capas de AlexNet

Su estructura, ilustrada en la figura 1, consta de cinco capas convolucionales y tres capas densas, con operaciones específicas como activaciones ReLU, normalización, agrupación (max-pooling), dropout y softmax. Su entrada es una imagen de 227×227 píxeles con tres canales de color (RGB). Los detalles de las operaciones y parámetros de cada capa se explican gráficamente y están resumidos en la tabla 1.

Las salidas de la primera, cuarta y quinta capas convolucionales pasan por una operación de max-pooling con un filtro 3×3 y un stride de 2. Este desplazamiento genera un cierto solapamiento entre las celdas agrupadas,

ayudando a capturar mejor los detalles espaciales de las características. En la quinta capa convolucional, se generan 256 mapas de características con dimensiones 13×13.

Las capas de pooling y convolución están seguidas por dos capas densas de 4096 neuronas cada una y una capa final con 1000 neuronas de salida. Estas últimas utilizan una función de activación softmax para realizar la clasificación.

Cada neurona de la capa de salida representa una categoría semántica específica. El valor de activación de cada neurona indica la probabilidad de que la imagen pertenezca a esa categoría.

La tabla 1 detalla los pesos W y bias b, el sesgo asociado a cada filtro, aprendidos en cada capa. Por ejemplo, la primera capa convolucional (conv1) aprende pesos $W = 11 \times 11 \times 3 \times 96$, correspondiendo 3 a los canales de entrada (RGB) y 96 al número de filtros, $b = 1 \times 1 \times 96$, produciendo una salida de $55 \times 55 \times 96$. Este esquema se repite para las demás capas convolucionales y densas.

Capas	**Dimensión de la salida**	**Pesos**
conv1	$55 \times 55 \times 96$	$W = 11 \times 11 \times 3 \times 96$, $b = 1 \times 1 \times 96$
conv2	$27 \times 27 \times 256$	$W = 5 \times 5 \times 48 \times 256$, $b = 1 \times 1 \times 256$
conv3	$13 \times 13 \times 384$	$W = 3 \times 3 \times 256 \times 384$, $b = 1 \times 1 \times 384$
conv4	$13 \times 13 \times 384$	$W = 3 \times 3 \times 192 \times 384$, $b = 1 \times 1 \times 384$
conv5	$13 \times 13 \times 256$	$W = 3 \times 3 \times 192 \times 256$, $b = 1 \times 1 \times 256$
fc6	$1 \times 1 \times 4096$	$W = 4096 \times 9216$, $b = 4096 \times 1$
fc7	$1 \times 1 \times 4096$	$W = 4096 \times 4096$, $b = 4096 \times 1$
fc8	$1 \times 1 \times 1000$	$W = 1000 \times 4096$, $b = 1000 \times 1$

Cuadro 1: Especificaciones de las capas de AlexNet

La red neuronal es simulada por dos GPUs, a cada una de las cuales les corresponde el cálculo de la mitad de los mapas de características de cada capa convolucional. Las dos GPUs pueden comunicarse directamente y acceder a la memoria de la otra sin necesidad de pasar por la CPU. Los mapas de características de la tercera capa convolucional se comunican con todos los mapas de la capa siguiente. Sin embargo, los mapas de la cuarta capa solo se comunican con los mapas de la capa tercera que residan en su GPU.

Durante el entrenamiento, AlexNet utiliza dropout en las capas densas, apagando aleatoriamente ciertas neuronas en cada lote (batch). Este proceso previene el sobreajuste (overfitting) y mejora la capacidad de generalización de la red.

Una característica distintiva de AlexNet es el uso de activaciones ReLU en lugar de funciones como la tangente hiperbólica, reduciendo significativamente el costo computacional. Además, los autores introdujeron el concepto de convolución inversa para analizar las características aprendidas, desmitificando la red como una "caja negra".

Referencias

BVLC (2020). Alexnet model. *https://github.com/bvlc/caffe/tree/master/models/bvlc_alexnet*. Disponible on-line. Accedido: diciembre, 2020.

ImageNet, 2020; Russakovsky, O., Deng, J., Su, W., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Maji, S., et al.; Krizhevsky, A., Sutskever, I., and Hinton, G. E. B. (2020). Alexnet and imagenet large scale visual recognition challenge. Disponible en línea: `http://www.image-net.org/challenges/LSVRC/2020/` y `https://github.com/BVLC/caffe/blob/master/models/bvlc_alexnet/readme.md`.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks.*

Oráculo, A. L. M. (2021). Alexnet: Red convolucional profunda para clasificación de imágenes. Disponible on-line. Accedido: abril 2025.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision.* Accedido: Febrero 2025.