

Trabajo: Técnicas de Reconocimiento de Patrones

Jorge Bengoa Pinedo Bogurad Barański Barańska Sergio Izquierdo Morona

22 de enero de 2026

Índice

1. Introducción a la clasificación de imágenes térmicas	2
1.1. Redes Convolucionales...	2
1.2. Metodología Yolo	2
2. Fundamentos de funcionamiento de la red Yolo	2
2.1. Neurona de la red Yolo	2

1. Introducción a la clasificación de imágenes térmicas

Explicar el estado del arte un poco de que hay y tal.... redes tal, este otro tipo de clasificación

1.1. Redes Convolucionales...

1.2. Metodología Yolo

2. Fundamentos de funcionamiento de la red Yolo

A continuación se desarrolla el funcionamiento de la red YOLO mediante la teoría de Marr para lo cuál es necesario describir el nivel implementacional y algorítmico. En este caso las redes neuronales artificiales se componen de 3 componentes que se describirán a continuación:

- Caracterización de la neurona.
- Definición de una topología de interconexión
- Definición de unas reglas de ajuste de pesos

2.1. Neurona de la red Yolo

La unidad fundamental de procesamiento en la arquitectura YOLO es la capa convolucional (típica de las CNN). En esta sección se describe el funcionamiento de dicha operación, tomando como caso de estudio la capa de entrada, analizando tanto el entrenamiento específico realizado para imágenes térmicas como la implementación por defecto, tal como se ilustra en la Figura 1.

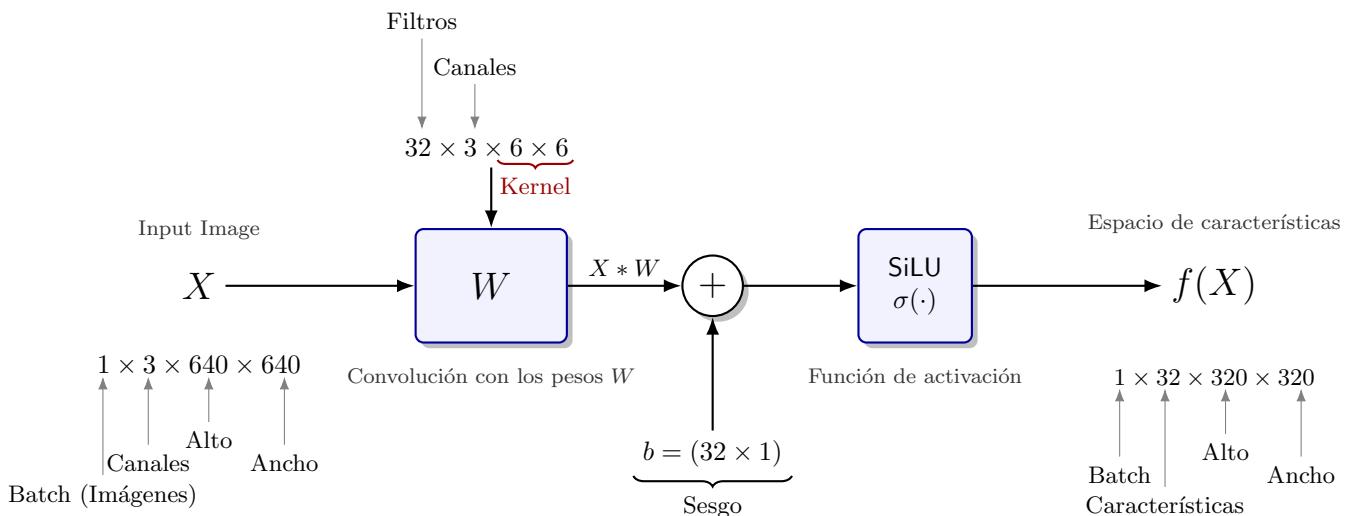


Figura 1: Estructura de la neurona a la entrada de la red YOLO.

La operación de esta neurona convolucional está definida por los siguientes hiperparámetros:

- **Número de filtros (Características C_{out}):** Determinado por la primera dimensión del tensor de pesos. Define cuántos mapas de características se generarán.
- **Canales de entrada (C_{in}):** Determinado por la segunda dimensión del tensor de pesos (e.g., 3 para RGB, 1 para imágenes térmicas).

- **Tamaño del Kernel (k):** Define las dimensiones espaciales de la ventana de convolución (últimas dos dimensiones de los pesos). En la arquitectura YOLO analizada, se emplean principalmente los tamaños (6,6) para la entrada, y (3,3) o (1,1) para capas profundas.
- **Padding (P):** Parámetro de relleno utilizado para controlar las dimensiones espaciales de la salida e indicar los bordes al llenarlos de ceros. En YOLO se utilizan típicamente valores de (2,2), (1,1), o (0,0).
- **Stride (S):** Controla el paso o desplazamiento del filtro sobre la imagen. Un $S = 1$ mueve el filtro píxel a píxel (preservando resolución), mientras que $S = 2$ lo desplaza cada dos píxeles, reduciendo la dimensión espacial a la mitad.
- **Dilatación (d):** Controla la separación entre los puntos que analiza el kernel (también conocida como *atrous convolution* cuando $d > 1$). En esta implementación de YOLO, el parámetro es (1,1), lo que indica una convolución estándar compacta sin huecos entre los elementos del kernel.
- **Agrupación (g):** Controla la conectividad entre los canales de entrada y salida. Dado que en YOLO se utiliza $g = 1$, se realiza una convolución estándar donde todos los canales de entrada contribuyen al cálculo de cada filtro de salida.

A partir de estos hiperparámetros, es posible determinar las dimensiones espaciales del mapa de características de salida. Dado que se contempla el factor de dilatación, se debe utilizar la relación generalizada descrita por [1].

$$\text{Salida} = \left\lceil \frac{\text{Entrada} + 2 \cdot P - [k + (k-1)(d-1)]}{S} \right\rceil + 1 \quad (1)$$

En cuanto a la activación de cada neurona, el proceso consiste en aplicar la función SiLU (*Sigmoid Linear Unit*) al resultado de la operación lineal (convolución más sesgo). Matemáticamente, la activación $f(X)$ se expresa como:

$$f(X) = \underbrace{[(W * X) + b]}_z \cdot \sigma(\underbrace{[(W * X) + b]}_z) \quad (2)$$

donde $\sigma(z) = \frac{1}{1+e^{-z}}$ es la función sigmoide logística. Se utiliza esta función de activación porque permite el cálculo de gradientes y solo permite valores negativos pequeños para mejorar el flujo de gradientes [2, 3], se ilustra en la Figura 2.

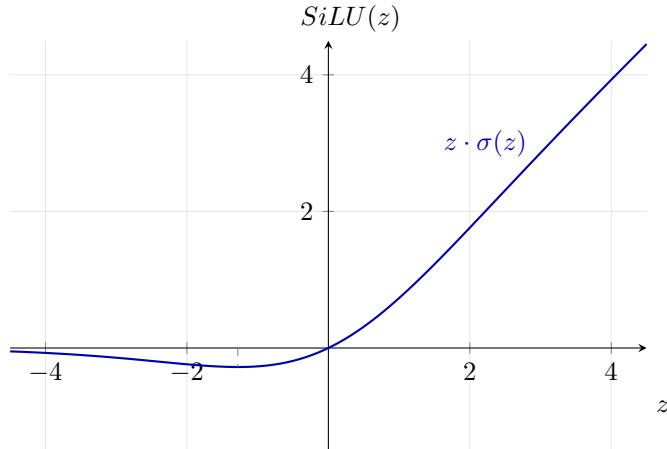


Figura 2: Representación gráfica de la función de activación SiLU.

Por otro lado, la operación de convolución discreta 2D ($W * X$) en el contexto de aprendizaje profundo se implementa técnicamente como una correlación cruzada (sin voltear el kernel) pues solo se busca conocer los pesos y no aplicar un filtro con una topología concreta. Para un filtro de salida específico (c_{out}) en la posición espacial (h, w) , la operación se define formalmente como [1]:

$$(W * X)(c_{out}, h, w) = \sum_{l=0}^{\frac{C_{in}}{g}-1} \sum_{i=0}^{K_H-1} \sum_{j=0}^{K_W-1} W(c_{out}, l, i, j) \cdot X(l, h \cdot S + i \cdot d - P, w \cdot S + j \cdot d - P) \quad (3)$$

Referencias

- [1] V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning,” 2018.
- [2] S. Elfwing, E. Uchibe, and K. Doya, “Sigmoid-weighted linear units for neural network function approximation in reinforcement learning,” 2017.
- [3] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for activation functions,” 2017.