

Aprendizaje Automático II

Práctica 3: U-Net

La arquitectura de red tipo **U-Net** está diseñada para realizar segmentación semántica en imágenes. Para ello, la red consta de dos partes: la primera parte se denomina "contracción" o "encoder", mientras que la segunda parte se llama "expansión" o "decoder". La sección de contracción tiene como objetivo capturar el contexto y las características de alto nivel de la imagen, reduciendo progresivamente su dimensionalidad a través de capas convolutivas y de agrupamiento máximo (max pooling). Por otro lado, la sección de expansión reconstruye la resolución espacial mediante operaciones como convoluciones transpuestas o interpolaciones, combinadas con las características extraídas en la etapa anterior.

Una característica clave de la U-Net es el uso de conexiones de salto ("skip connections") que vinculan las características de la etapa de contracción con las de la etapa de expansión. Esto permite preservar información espacial relevante, mejorando la precisión en los bordes y detalles de las áreas segmentadas.

Tareas

Se presenta el código de una red U-Net y un sencillo conjunto de imágenes y etiquetas para su entrenamiento. Una vez entrenada, observaremos los resultados comprobando visualmente el desempeño del modelo con unas muestras de test.

Dado el buen funcionamiento del modelo, llevaremos a cabo una simplificación de la de la red reduciendo progresivamente el número de canales en cada etapa de las partes encoder y decoder.

Después de esto, realizaremos un estudio de ablación eliminando progresivamente las conexiones "skip" entre encoder y decoder y viendo si se produce degradación en los resultados.

Entregables

Se pide una memoria que detalle los resultados obtenidos en las tareas mencionadas así como el código usado.