

Task 2

Responda claramente y con una extensión adecuada las siguientes preguntas:

1. Defina en sus propias palabras qué es una caja de anclaje (anchor box). Asegúrese de mencionar cuál es su importancia en la detección de objetos hecha por YOLO y cuál es la diferencia y relación con las cajas delimitadoras (bounding boxes)

Es una caja predefinida que se utiliza para detectar objetos de formas específicas en imágenes. En la detección de objetos con el algoritmo de YOLO, las cajas de anclaje son cruciales porque ayudan a lidiar con el problema de múltiples objetos de diferentes formas y tamaños que se superponen en una imagen.

Además la importancia de las cajas de anclaje en YOLO es que permite que el modelo detecte múltiples objetos en una sola celda de la cuadrícula. Con las cajas de anclaje, cada celda de la cuadrícula puede predecir múltiples cajas delimitadoras, lo que aumenta la precisión y la capacidad de detección del modelo.

Mientras que las cajas de anclaje son cajas predefinidas que actúan como puntos de partida, las cajas delimitadoras (o "bounding boxes") son las cajas finales que se ajustan y se predicen para rodear un objeto detectado en la imagen. Esencialmente, las cajas de anclaje actúan como una especie de "plantilla" que se ajusta para obtener la caja delimitadora final que rodea un objeto.

2. Considere leer el paper original de YOLO (You Only Look Once: Unified, Real-Time Object Detection) y luego responda

a. ¿Cuál es la innovación clave que distingue a YOLO de los enfoques tradicionales de detección de objetos y cómo contribuye a la detección de objetos en tiempo real?

YOLO presenta un enfoque novedoso para la detección de objetos en comparación con los métodos tradicionales. En lugar de reutilizar clasificadores para realizar detecciones, YOLO enmarca la detección de objetos como un problema de regresión hacia cajas delimitadoras espacialmente separadas y las probabilidades de clase asociadas. Lo que es realmente innovador acerca de YOLO es que utiliza una sola red neuronal para predecir cajas delimitadoras y probabilidades de clase directamente desde imágenes completas en una sola evaluación. Esto significa que todo el pipeline de detección es una única red, y puede ser optimizado de extremo a extremo directamente en el rendimiento de la detección.

b. ¿Cómo maneja YOLO la detección de múltiples objetos dentro de una sola imagen y cuáles son las ventajas de este enfoque sobre los métodos tradicionales?

YOLO divide la imagen en una cuadrícula y para cada celda de la cuadrícula, predice cajas delimitadoras y las probabilidades de clase. Las cajas delimitadoras se ajustan a los objetos detectados y las probabilidades de clase indican la probabilidad de que la caja contenga un objeto específico. La combinación de estas cajas delimitadoras y probabilidades permite a YOLO detectar múltiples objetos en una sola imagen. Una ventaja clave de este enfoque es su velocidad; dado que todo el proceso se realiza en una sola pasada a través de la red, YOLO puede procesar imágenes en tiempo real. Además, al tratar la detección de objetos como un problema unificado en lugar de varios problemas separados, YOLO es menos propenso a cometer errores en ciertas clases de objetos que otros métodos.

c. ¿Cuáles son las limitaciones o desafíos asociados con YOLO y cuáles son las áreas potenciales de mejora o investigación adicional en el campo de la detección de objetos en tiempo real?

Una de las principales limitaciones de las primeras versiones de YOLO es que tiende a tener problemas con objetos pequeños que aparecen en grupos, ya que solo se permite una detección por celda de la cuadrícula. Otra limitación es que puede llegar a cometer errores de falsos positivos en comparación con otros enfoques que utilizan pasos múltiples. Las áreas potenciales de mejora podrían incluir la mejora en la detección de objetos pequeños y superpuestos, la reducción de falsos positivos, y la adaptación de la arquitectura para manejar diferentes tamaños y formas de objetos de manera más efectiva.