23- S2-Q3

Q: (a) R-CNN YoLOV7 ?

Solution

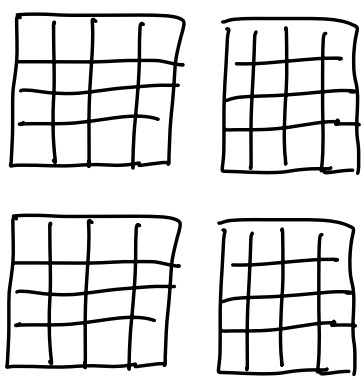
R-CNN: Two-stage detector

YoLoV7: One-stage detector

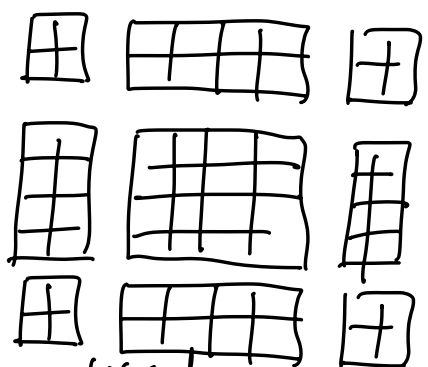
YoLoV7 is a more suitable choice as it is a more recent one-stage object detection algorithm.

One-stage detector is generally faster than two-stage detector such as RCNN

(b) diagram



window-base self-attention



shifted window-base self-attention

objectives

① improve Computational efficiency

window-base self-attention reduces the quadratic complexity of standard self attention, such as VIT, to linear complexity relative to image size

② Hierarchical Representation

The use of windows allow the Swin Transform to build hierarchical feature representation

③ Enhanced Contextual understanding

shifted windows enable the model to capture relationships between distant patches improving on vision tasks requiring global context

cc) key steps in tracking-by-detection multiple-object tracking:

① detection: Use a detector to initialize tracks at frame t.

② Motion prediction: predict the next position of objects using motion model

③ Data association : Match predicted position with detections at frame t+1

(d) ① Encode motion information without using optical flow

② Temporal modeling without 3D convolution

③ No Addition Parameters : TSM use temporal modeling

④ minimal computational overhead : TSM use shift operation which are memory-efficient