

23- S2-Q3

Q: (a) R-CNN YOLOV7 ?

Solution

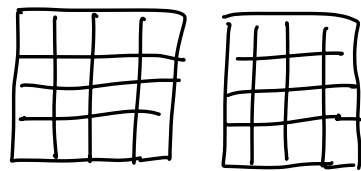
R-CNN : Two-stage detector

YOLOV7 : One-stage detector

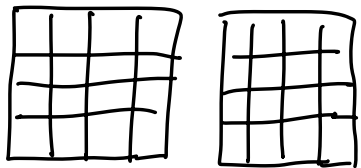
YOLOV7 is a more suitable choice as it is a more recent one-stage object detection algorithm.

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

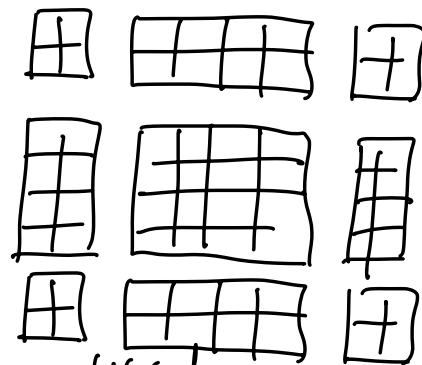
(b) diagram



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

c) 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 $t+1$

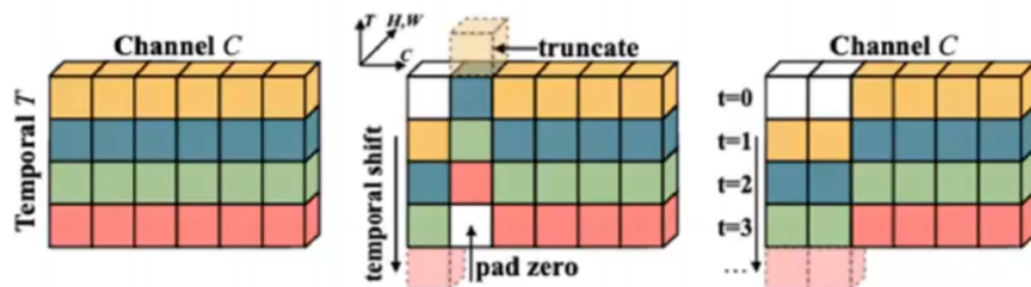
(d) ① Encode motion information without using optical flow

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

(d)



(a) The original tensor without shift.

(b) Offline temporal shift (bi-direction).

(c) Online temporal shift (uni-direction).

Temporal Shift Module (TSM) uses channel shifting operation. It shifts part of the channels along the temporal dimension, thus facilitating information exchange among neighboring frames. This is much more efficient than 2 stream approach or 3DCNN.

TSM (Temporal Shift Module) 使用信道移位操作。

它沿着时间维度移动部分信道，从而促进相邻帧之间的信息交换。

这比2流方法或3DCNN更有效。

