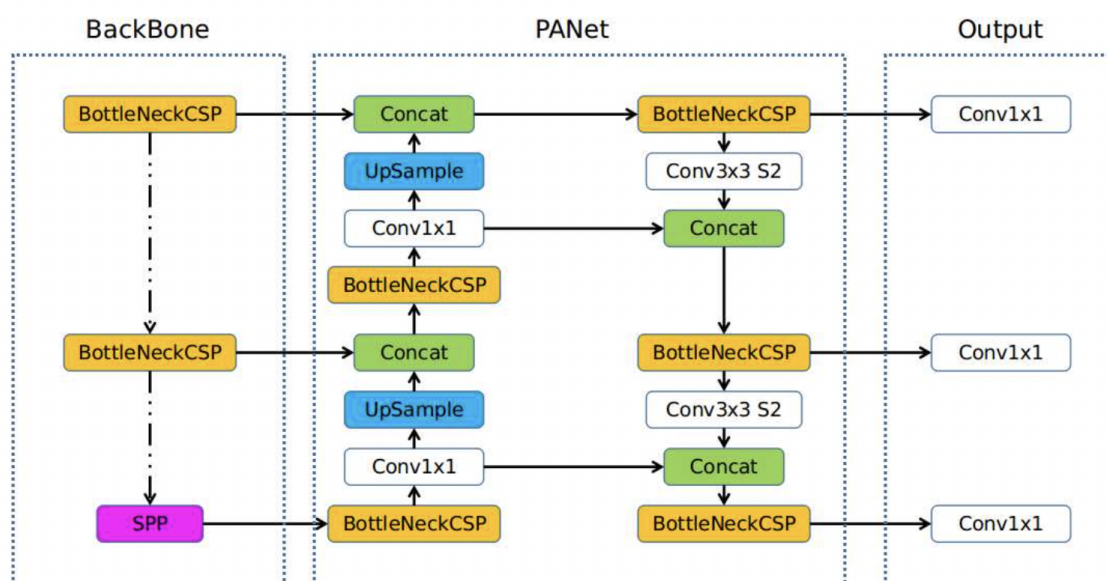


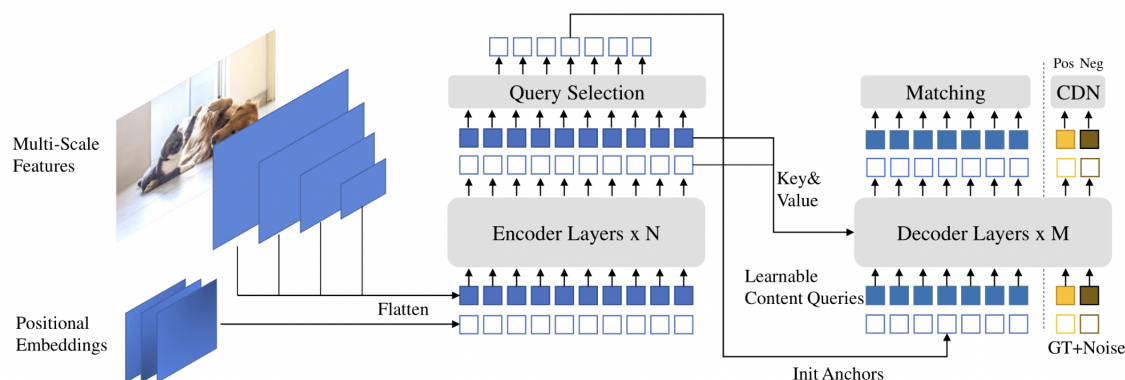
CVDPL - HW1

R11922a05 資工AI碩一 林聖硯

1. (5%) Draw the architectures for both CNN-based and Transformer-based methods
 - CNN-based model: YOLOv5-medium
 - Reference: <https://github.com/ultralytics/yolov5>, [YOLOv5 - paper](#)



- Transformer-based model: DINO (DINO-4scale + RestNet50)
 - Reference: <https://github.com/open-mmlab/mmdetection>, DINO paper



2. (10%) Report and compare the performance of two methods on validation set

Models	mAP@[50:5:95]	mAP@50	mAP@75
DINO 4 scale + ResNet 50	0.5183	0.8031	0.5377
YOLOv5 - medium	0.4716	0.6791	0.4924

3. Report the implementation details of both methods

Models	Data Augmentation	Loss function	Batch size	Optimizer
Dino 4 scale + backbone = ResNet 50	RandomChoiceResize (scales = [(400, 4200), (500, 4200), (600, 4200)]) + RandomCrop (384, 600)	IoU loss: GloULoss classifier loss: binary cross entropy	2	AdamW - lr = 0.0001 - weight decay = 0.0001 clip gradient: max norm = 0.1
YOLOv5 - medium	Image Scaling (50%) Image translation Random flip (probability = 0.5) Mosaic augmentation (probability = 1.0)	IoU loss: CIoU loss classifier loss: binary cross entropy	8	SGD - lr = 0.0001 - weight decay = 0.0005 - warmup epochs = 3 - warmup momentum = 0.8 - warmup bias lr = 0.1

4. Visualization: draw the bounding boxes of two methods

Dino 4 scale + ResNet 50



YOLOv5 - medium



