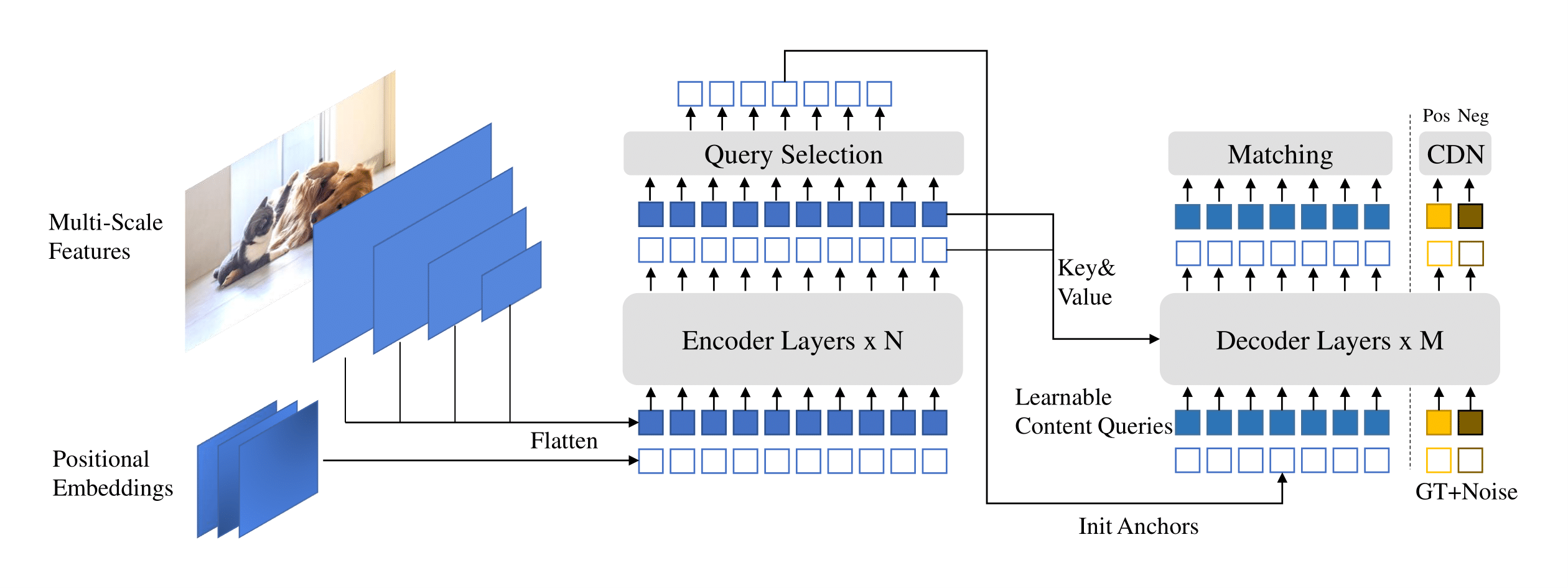
CVPDL HW1 report

1. Framework



Backbone: ResNet-50

Pre-trained weights: Pre-train on COCO 2017 object detection dataset.

1. Implement details:

Data augmentation: Horizontal Flip, resize images to different scales, crop images to random size and locations.

Loss function: L1 loss and GIOU loss for box regression and focal loss with α = 0.25, γ = 2 for classification

Batch size: 1 (due to GPU memory limit)

Epochs: 24

Learning rate: 0.0001

Lr\_backbone: 1e-05

Weight\_decay: 0.0001

Clip\_max\_norm: 0.1

Pe\_temperature: 20

Enc\_layers: 6

Dec\_layers: 6

Dim\_feedforward: 2048

Hidden\_dim: 256

Dropout: 0.0

Nheads: 8

Num\_queries: 900

Enc\_n\_points: 4

Dec\_n\_points: 4

Transformer\_activation: “relu”

Batch\_norm\_type: “FrozenBatchNorm2d”

Set\_cost\_class: 2.0

Set\_cost\_bbox: 5.0

Set\_cost\_giou: 2.0

Cls\_loss\_coef: 1.0

Bbox\_loss\_coef: 5.0

Giou\_loss\_coef: 2.0

Focal\_alpha: 0.25

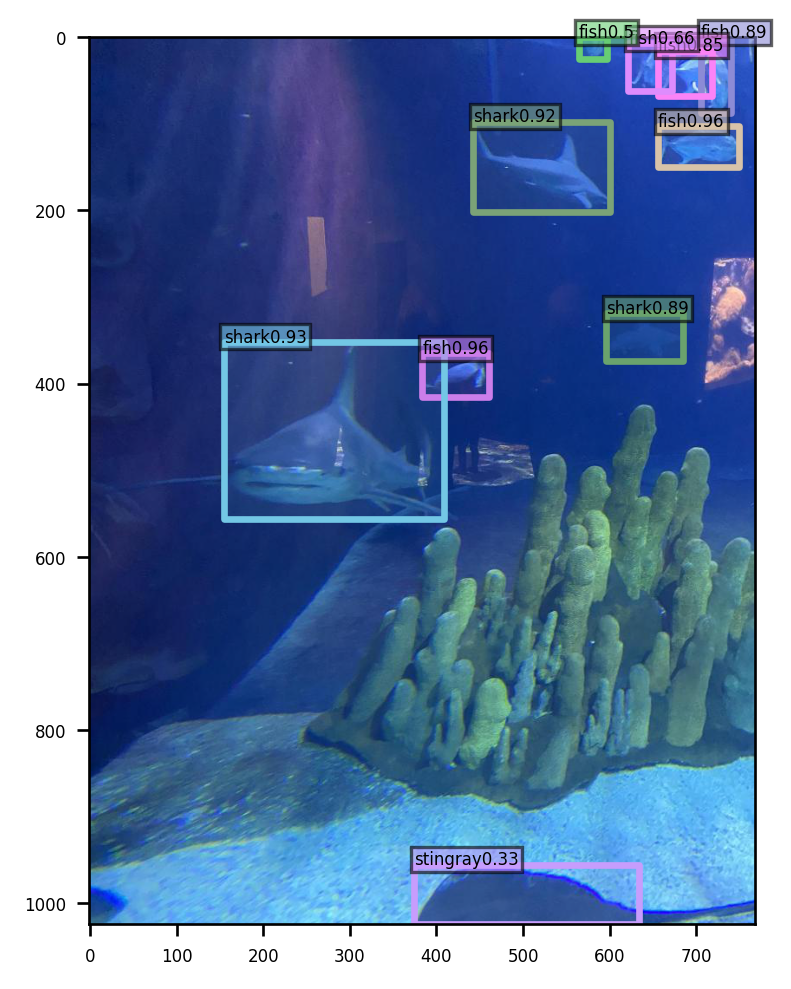
Dn\_box\_noise\_scale: 0.4

Dn\_label\_noise\_ratio: 0.5

1. Table of your performance for validation set (AP, AP50, AP75):

{'map': tensor(0.5213), 'map\_50': tensor(0.8101), 'map\_75': tensor(0.5314), 'map\_small': tensor(0.2119), 'map\_medium': tensor(0.4267), 'map\_large': tensor(0.6666), 'mar\_1': tensor(0.2520), 'mar\_10': tensor(0.5507), 'mar\_100': tensor(0.6654), 'mar\_small': tensor(0.4225), 'mar\_medium': tensor(0.5920), 'mar\_large': tensor(0.7660), 'map\_per\_class': tensor(-1.), 'mar\_100\_per\_class': tensor(-1.), 'classes': tensor([0, 1, 2, 3, 4, 5, 6, 7], dtype=torch.int32)}

1. Visualization



Reference:

1. DINO paper <https://arxiv.org/abs/2203.03605>
2. Github repo <https://github.com/IDEA-Research/DINO>