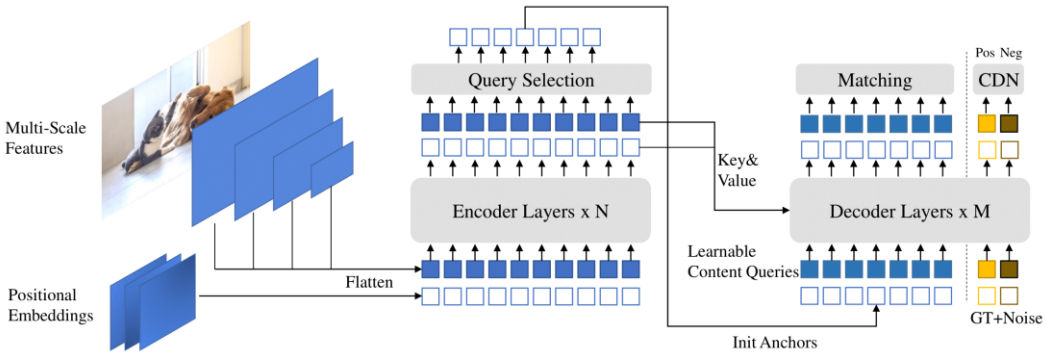
CVPDL HW1 report

1. Framework



Backbone: ResNet-50

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

2. 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 $\alpha = 0.25$, $\gamma = 2$ 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”

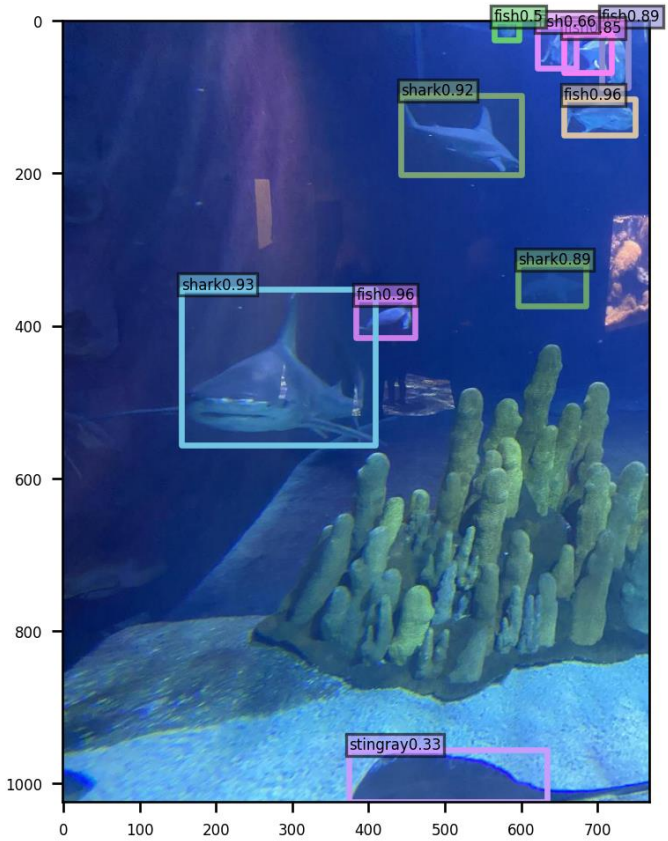
Set_cost_class: 2.0

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

3. 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)}
```

4. Visualization



Reference:

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