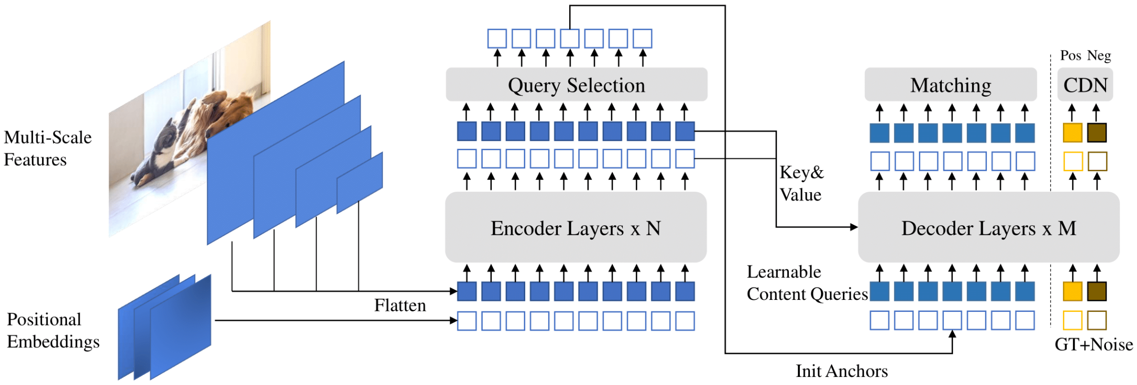
CSIE5428 Computer Vision Practice with Deep Learning

Homework 1 Report

Name: 高榮浩 **ID:** R12922127

1. The Architecture of Object Detector

I've chosen the DINO model. The diagram below illustrates the DINO framework.



2. Implement Details

I predominantly opt for the original DINO model. The pre-trained checkpoint I leverage from the model zoo is specifically DINO-4scale, with the Swin-L backbone. Consequently, this implementation involves the utilization of two pre-trained weights, namely those for DINO-4scale and Swin-L, as indicated in the following table.

Name	*File*	*Dataset*	*Source*
DINO-4scale (36 epoch setting)	`checkpoint0029_4scale_swin.pth`	COCO 2017	Link
Swin-L	`swin_large_patch4_window12_384_22k.pth`	ImageNet-22K	Link

Please note that I have deliberately avoided using the DINO-5scale (36 epoch setting) pre-trained checkpoint. While it does offer a higher box AP, my GPU (GeForce RTX™ 2080 Ti 11G) lacks the capacity to accommodate it. This is due to a `RuntimeError` that occurs, specifically "`CUDA out of memory`," even with the batch size initially set to 1.

In that case, I attempted to use DINO-4scale (36 epoch setting) with an initial batch size of 2, but my GPU still couldn't handle it. As a result, I had to reduce the batch size to 1, and that ultimately led to successful execution.

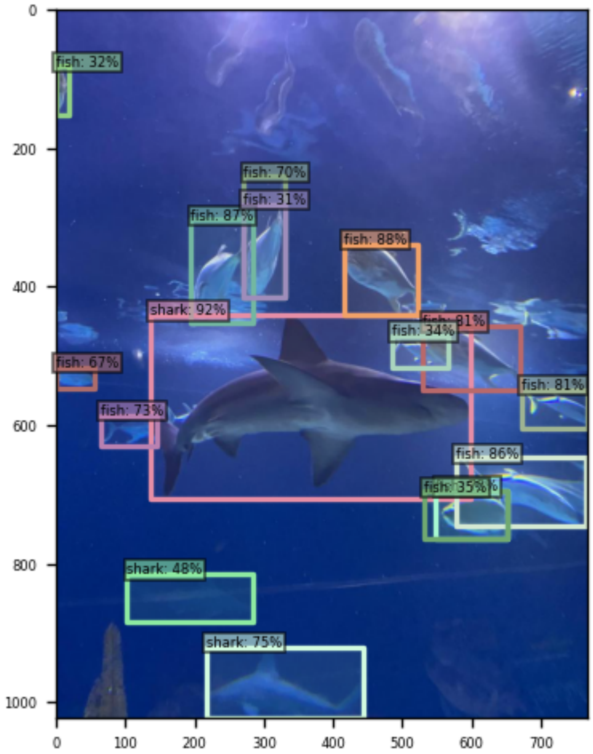
Subsequently, I embarked on a series of experiments involving various combinations of epoch numbers and learning rates, the results of which are outlined in the following section.

3. Performance for Validation Set

#	*Learning Rate*	*Epoch*	*Best Epoch*	*AP*	*AP_{50}*	*AP_{75}*
1	0.0001	12	11	0.582	0.863	0.608
2	0.0001	12	11	0.574	0.850	0.610
3	0.0001	12	6	0.577	0.858	0.611
4	0.0001	12	11	0.578	0.852	0.609
5	0.0001	12	11	0.574	0.863	0.587
6	0.0001	18	12	0.578	0.848	0.606
7	0.0001	18	17	0.580	0.866	0.605
8	0.0001	24	16	0.587	0.863	0.609
9	0.0001	24	18	0.581	0.859	0.604
10	0.0001	36	19	0.584	0.864	0.618
11	0.00005	48	36	0.588	0.866	**0.621**
12	0.000025	36	23	0.587	**0.867**	0.596
13	0.000025	96	12	**0.591**	**0.867**	0.616
14	0.0000125	36	20	0.589	0.856	0.604
15	0.00000625	48	31	0.581	0.858	0.601

This table displays the performance results for the validation set. The highest AP score achieved is 0.591, which was obtained with a learning rate of 0.000025 during the 13th epoch. It's important to note that the index for the 'Best Epoch' in this table begins at 0.

4. Visualization



The image above showcases the detection results for `IMG_2570_jpeg_jpg.rf.ed40900b657a5b23d92cb2d296ad2dbc.jpg` in the testing set.