Report on Deformation-DETR

Step 1: Firstly I have divided the Dataset into two part.

i) 80 percent for the Training.
ii) 20 percent for the evaluation.

Step 2: Then I have used Deformable DETR(single scale, DC5) model.

Step 3: For visualization i have modified the main.py such that now after evaluation it draw the bounding boxes on the image and store them under the "**visual**" folder.

Step 4 : Below is the evaluation result of the pre-trained model on the evaluation dataset.

```
IoU metric: bbox
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.394
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.741
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.379
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.289
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.507
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.727
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.110
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.442
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.497
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.407
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.594
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.778
```

Fig 1: AP values using the Pre-trained Model.

Step 5: Now training the model, Hyper-parameters that I have experimented with are given below.

- ➔ Learning Rate.
- ➔ Batch-Size.
- ➔ Weight-Decay.
- ➔ Epochs.
- ➔ Learning rate drop.
- ➔ Clip-max Norm.
- ➔ Encoding layer.
- ➔ Decoding layer.

Also tried the “Frozen-weight” hyper-parameter but it generated shape error.

Step 6: Training Model on different values of Hyper-parameter.

Step 6.1: Training Model on Default Values.

```
IoU metric: bbox
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.212
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.545
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.110
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.135
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.346
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.292
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.067
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.287
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.355
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.261
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.474
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.561
```

Fig 2: AP values after training using the default values.

Step 6.2: Below is the evaluation result of the trained model set-1, hyper-parameter values are as follows:

- ➔ Learning Rate : 0.0001
- ➔ Batch-Size: 4
- ➔ Weight-Decay: 0.0001
- ➔ Epochs: 50
- ➔ Learning rate drop: 40
- ➔ Clip-max Norm: 0.1
- ➔ Encoding layer: 6
- ➔ Decoding layer: 6

```
IoU metric: bbox
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.166
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.491
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.065
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.087
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.308
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.210
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.057
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.247
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.336
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.238
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.463
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.533
```

Fig 3: AP values after training using the SET-1 values.

Step 6.3: Below is the evaluation result of the trained model set-2, hyper-parameter values are as follows:

- ➔ Learning Rate : 0.00005
- ➔ Batch-Size: 8
- ➔ Weight-Decay: 0.00005
- ➔ Epochs: 100
- ➔ Learning rate drop: 75
- ➔ Clip-max Norm: 0.1
- ➔ Encoding layer: 4
- ➔ Decoding layer: 4

```
IoU metric: bbox
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.001
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.007
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.000
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.000
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.003
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.000
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.002
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.004
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.004
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.000
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.012
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.000
```

Fig 4: AP values after training using the SET-2 values.

Step 6.4: Below is the evaluation result of the trained model set-3, hyper-parameter values are as follows:

- ➔ Learning Rate : 0.0002
- ➔ Batch-Size: 2
- ➔ Weight-Decay: 0.0001
- ➔ Epochs: 75
- ➔ Learning rate drop: 60
- ➔ Clip-max Norm: 0.1
- ➔ Encoding layer: 8
- ➔ Decoding layer: 8

```
IoU metric: bbox
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.214
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.600
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.116
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.132
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.358
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.292
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.069
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.295
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.362
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.266
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.496
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.500
```

Fig 5: AP values after training using the SET-3 values.

Step 6.5: Below is the evaluation result of the trained model set-4, hyper-parameter values are as follows:

- ➔ Learning Rate : 0.0001
- ➔ Batch-Size: 2
- ➔ Weight-Decay: 0.00005
- ➔ Epochs: 60
- ➔ Learning rate drop: 45
- ➔ Clip-max Norm: 0.1
- ➔ Encoding layer: 6
- ➔ Decoding layer: 6

```
IoU metric: bbox
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.043
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.210
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.004
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.022
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.077
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.102
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.016
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.115
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.205
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.114
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.307
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.461
```

Fig 6: AP values after training using the SET-4 values.

Step 6.6: Below is the evaluation result of the trained model set-5, hyper-parameter values are as follows:

- ➔ Learning Rate : 0.00015
- ➔ Batch-Size: 4
- ➔ Weight-Decay: 0.0001
- ➔ Epochs: 80
- ➔ Learning rate drop: 70
- ➔ Clip-max Norm: 0.1
- ➔ Encoding layer: 5
- ➔ Decoding layer: 5

```
IoU metric: bbox
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.021
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.100
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.001
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.005
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.054
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.041
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.018
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.058
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.087
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.008
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.209
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.150
```

Fig 7: AP values after training using the SET-5 values.

Evaluation of the Result:

After training the model on the training data and then done the evaluation on the evualtion dataset the AP and AR values are deteriorated.

Reasons can be :

- ➔ Overfitting.
- ➔ Insufficient Data for Fine-Tuning.
- ➔ Less data for evaluation.

In visualization we can see before training and after training the model is creating additional boxes example image 1904.jpg , we can see there are overlapping boxes. This can be removed using the NMS(Non-Maximal Suppression).

Also in example 11613.jpg the bounding boxes are not correct. Detecting pillar as pedestrian in example 13187.jpg this can be caused by overfitting or insufficient data which can be rectified by training it on large-dataset and using the cross-validation technique.

Drive Link :
https://drive.google.com/drive/folders/148S4LtVZxceICeC9Y8OGMg16UjoVTxgB?usp=sharing