УДК 004.932 : 004.896

Авраменко С.Є., аспірант спеціальності 124 Системний аналіз
Науковий керівник: Желдак Т.А. к.т.н., доцент кафедри системного аналізу та управління
(Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна)

OBJECT DETECTION OF VEHICLES IN IMAGES USING THE DETR MODEL

The End-to-End Detection Transformer (DETR) [1] is a novel object detection model introduced by researchers at Facebook AI Research (FAIR). DETR represents a departure from traditional object detection methods, which typically rely on region proposal networks (RPNs) and non-maximum suppression (NMS) for bounding box prediction. Instead, DETR employs a transformer architecture to perform object detection directly in end-to-end manner.

DETR utilizes the transformer architecture, originally introduced for natural language processing tasks, for object detection. Transformers are known for their ability to capture long-range dependencies in sequences, making them well-suited for tasks like object detection where contextual understanding is crucial.

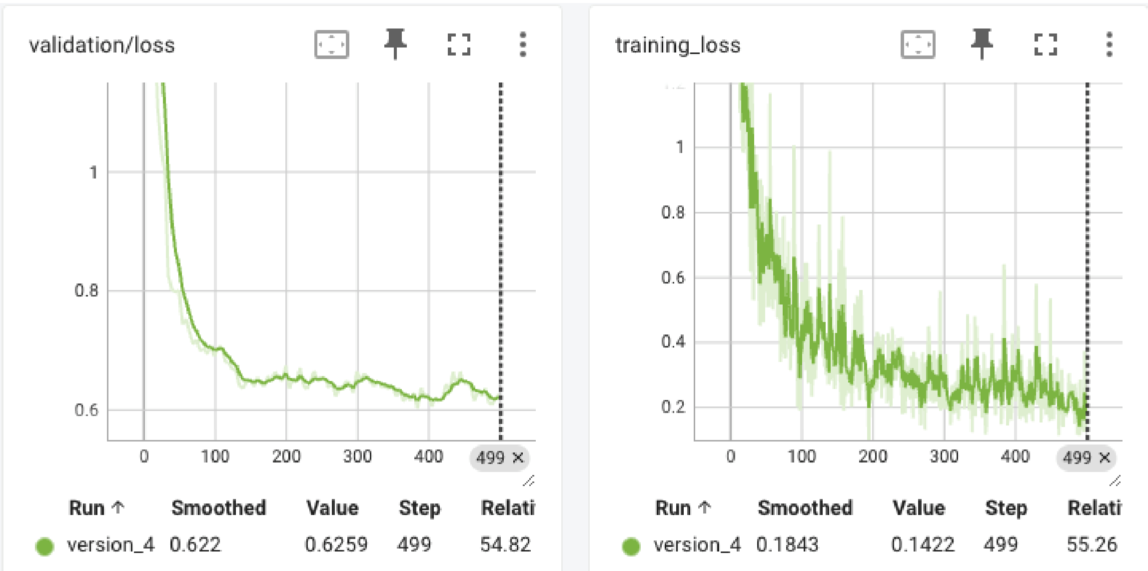
Unlike traditional two-stage object detection models, DETR directly predicts object bounding boxes and class labels in a single forward pass. This eliminates the need for separate region proposal and object classification stages, simplifying the architecture.

DETR incorporates a global context mechanism, allowing it to consider information from the entire image when making predictions for each individual object. This enables better contextual understanding and improves detection accuracy.

DETR formulates the object detection task as a set prediction problem. Instead of predicting bounding boxes individually, DETR treats object detection as predicting a set of objects, with each object represented by a bounding box and a corresponding class label.

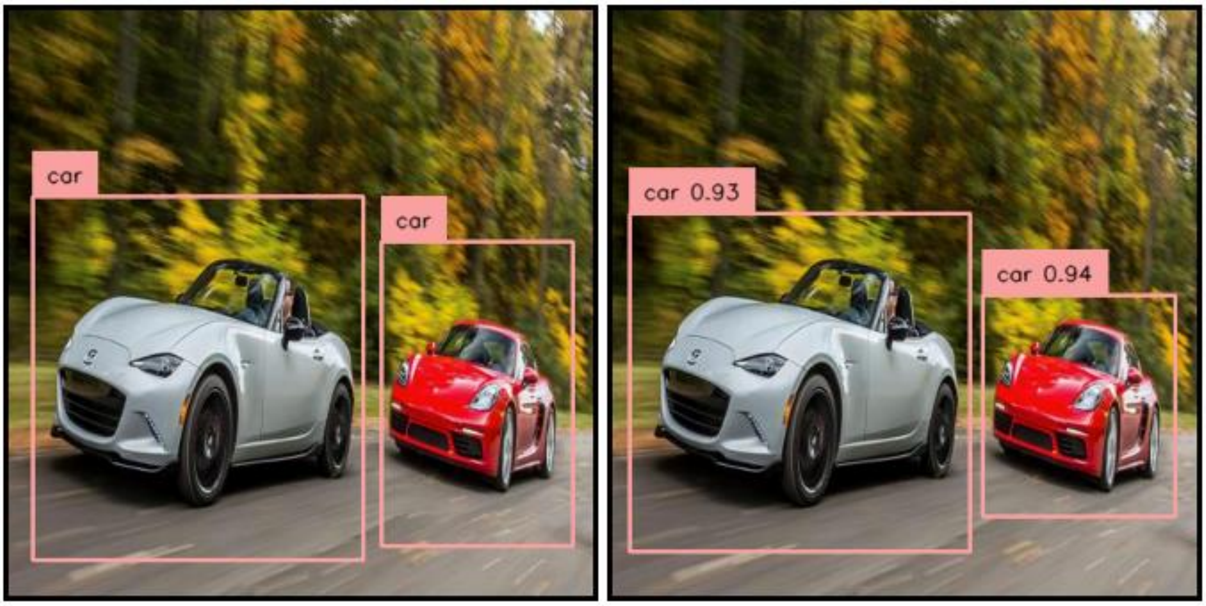
To maintain spatial information within the transformer architecture, DETR uses learnable positional encodings to represent the location of each object within the image.

It was used the DETR model to detect vehicles in images from Car dataset [2]. The model has been trained on 200 images. The dataset was split on train (130 images), validation (30 images) and test (40 images) datasets. It was trained for 500 steps so the loss could be seen on picture 1.



Picture 1 – Training and Validation Loss.

As could be seen, the model was successfully trained thus it is possible to see model performance during inference on the sample example in picture 2. On the left, it is annotated bounding boxes and on the right, is the predicted result with the class label and its score.



Picture 2 – Inference example on the sample image

According to COCO standards the model performance is illustrated on the picture 3. The DETR model achieved 27.9mAP@0.5 IoU and 31.3mAP@[0.5:0.95] Average IoU on the Car dataset.

```
Accumulating evaluation results...
DONE (t=0.01s).
IoU metric: bbox
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.185
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.279
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.192
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.256
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.193
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.139
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.249
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.251
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.313
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.268
```

Picture 3 – COCO evaluation

References:

[1] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-End Object Detection with Transformers.” arXiv, May 28, 2020. doi: 10.48550/arXiv.2005.12872.

[2] “Object detection Dataset > Overview,” Roboflow. Accessed: Mar. 11, 2024. [Online]. Available: https://universe.roboflow.com/vishi-sapra/object-detection-axukj