Computer Vision - Project 3

Ankur Duggal

December 9, 2024

INDIANA UNIVERSITY

Contents

1 Objective

Test out DETR and see how it works.

2 Methods

Run the code provided in the DETR paper.

3 Results

Logits

```
tensor([[[ 0.6964, -1.2092,  0.5691,  ...,  0.7348,  0.5838, -0.0035]],
        [[ 0.7684, -1.1451,  0.5113,  ...,  0.7312,  0.5660,  0.0856]],
        [[ 0.7116, -1.1590,  0.5353,  ...,  0.6440,  0.6295,  0.0349]],
        ...,
        [[ 0.6992, -1.1500,  0.5965,  ...,  0.7161,  0.5593,  0.0195]],
        [[ 0.7160, -1.2004,  0.5466,  ...,  0.6985,  0.5674,  0.0925]]],
       grad_fn=<ViewBackward0>)
```

Bounding Boxes

```
tensor([[[0.6426, 0.2710, 0.6302, 0.2297]],
        [[0.6597, 0.2477, 0.6202, 0.2342]],
        [[0.6449, 0.2742, 0.6227, 0.2300]],
        ...,
        [[0.6436, 0.2642, 0.6223, 0.2212]],
        [[0.6554, 0.2711, 0.6201, 0.2310]]],
       grad_fn=<SigmoidBackward0>)
```

Logit Example

```
tensor([[ 0.6964, -1.2092,  0.5691, -0.3872, -0.5159,  0.4093,  0.4286, -0.8433,
         -0.1327,  0.0074, -0.9611, -0.7159,  0.8142, -0.5632,  0.6320,  0.6878,
          0.6834, -0.5975, -0.3922,  0.0163,  0.1548, -0.3492, -0.2514, -0.7916,
          0.5532, -0.0526, -0.7709,  0.4626, -0.3595,  0.4727, -0.1911, -0.6829,
         -0.3139, -0.1079,  0.0777, -0.5370, -0.2825, -0.3070, -0.6595,  0.6055,
          0.7546, -0.1737, -0.3737,  0.0663, -0.1850, -0.0126, -0.0609,  0.1170,
         -0.3695, -0.0326,  0.0880,  0.8970, -0.9729, -0.6287,  0.0221,  0.0051,
          0.0309,  1.0344,  0.3865, -0.7079,  0.7514,  0.4516, -0.2383, -0.0289,
         -0.0648,  0.2646, -0.1535,  0.0839, -0.6175,  0.1158,  0.2083,  0.2686,
         -0.6850, -0.9527,  0.9408,  0.0086, -0.1705,  0.4390,  0.1232, -0.9033,
         -0.0750, -0.0342,  0.6741, -0.7501,  0.2086, -0.7705,  0.0712, -0.8127,
          0.1308,  0.7348,  0.5838, -0.0035]], grad_fn=<SelectBackward0>)
```

Bounding Box Example

```
tensor([[0.6426, 0.2710, 0.6302, 0.2297]], grad_fn=<SelectBackward0>)
```

4 Observations

- **Logits:** The logtis are basically the confidence that the bounding box is from a certain class. There are 92 outputs in the logits because there are 91 classes in the code and 1 background class. In the example, the logits showcase that the top 3 classes that this bounding box could relate to are:
    1. **Position 57** 1.0344
    2. **Position 74** 0.9408
    3. **Position 51** 0.8970
- **Bounding Box:** The bounding boxes are between 0-1 because they are a percentage of the whole image. The bounding box is ordered as 'center x', 'center y', 'width', 'height'. Each of these are based on the percentage of the image. In the example, that means the bounding box in the example is ordered as:
    1. **center x:** 64.26
    2. **center y:** 27.10
    3. **width:** 63.02
    4. **height:** 22.97

5 Conclusion

There are 100 bounding boxes because there are 100 object queries. The logits showcase the results. Ultimately, this works pretty well and is a very simple implementation of the DETR. Unfortunately, the results do not have much meaning as the image provided in the code is a randomly created matrix with no actual image behind it. So, the output is also random. But, this is a very simple implementation and could easily be replicated with real images.