
BBM 418 - Computer Vision Laboratory

2022 Spring

Assignment 4

Murat Çelik
21827263
Department of Computer Engineering
Hacettepe University
Ankara, Turkey
b21827263@cs.hacettepe.edu.tr

1 Object Classification and Localization

In this section, we will deal with the classification and localization problem over image data. The image classification problem is the problem of guessing what a class object is in images. In this section, this object will be Raccoon. The localization problem is the problem of showing the location of the searched object in the picture with a frame.

The pretrained ResNet18 on ImageNet will be used in this section. Two heads were created for these two problems. The classification head will be used for classification and the regression head will be used for finding the bounding box. Softmax and CrossEntropyLoss (softmax loss) were used for the classification head, and MSELoss (L2) was used for the regression head.

2 Training Phase

The training phase took place with 150 train Racoon images and 29 validation Racoon images. At this stage, we trained our model with 2 different batch values and 3 different learning rates. We recorded all the data of 6 models in total.

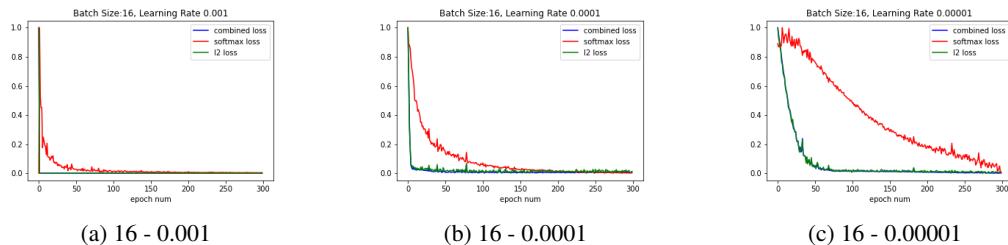


Figure 1: The Batch Size value is 16, the Learning Rates are the above-mentioned loss function values.

2.1 Loss Function

We have 3 different loss values. These are the l2 loss, the softmax loss, and the combined loss, which is the sum of the two. Below, 3 loss values are brought into the same range with the normalization

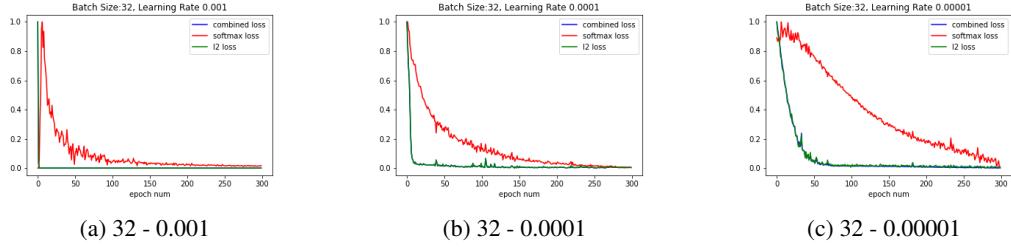


Figure 2: The Batch Size value is 32, the Learning Rates are the above-mentioned loss function values.

function and their changes are examined. In these changes, it is observed that the combined loss and the L2 loss value are closer to each other. Since the L2 loss value is much higher than the softmax loss value, this closeness is considered normal.

2.2 Batch Size

If we look at the results below and the lowest combined loss value of each model, we see that 32 batch size gives better results than 16.

2.3 Learning Rate

3 different learning rate values were used. These are: 0.001, 0.0001, 0.00001 values. These values were trained with 300 epochs and analyzed with graphs.

- It is observed that the value of 0.001 drops to minimum levels at 50 epochs with a rapid sharpness. This minimum value remains at a certain point, and does not fall any further. Therefore, this value remained a little high for the results.
 - It is observed that the value of 0.0001 reaches minimum levels at 150 epoch values. Due to the low level of this minimum, it is recorded as the most preferred learning rate value.
 - It seems that the value of 0.00001 reaches the minimum levels with 300 epochs, but this value is a small value for us as the level at which this value decreases is not enough.

2.4 Classification Accuracy

Despite the use of techniques such as Dropout due to the scarcity of the data set, it is seen that the accuracy value reaches 100% for each model in the first few values.

3 Testing Phase

A trained model with 32 batch values and 0.0001 learning rate was used for test operations. With the Sliding Window technique, the picture was divided into parts and a box estimate was made for each picture. These boxes are indicated in the pictures below. When we add all these boxes on the original image, a large number of boxes appear. We solve this problem with the non-maximum suppression technique. We calculate our final box and ground truth values with the Intersection over Union(IoU) method. Below are 3 sample test data attached. For more data, check the drive folder.

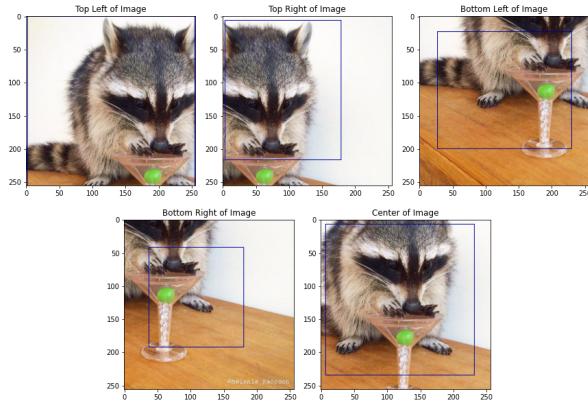


Figure 3: Sliding window frames of the raccoon dataset 59

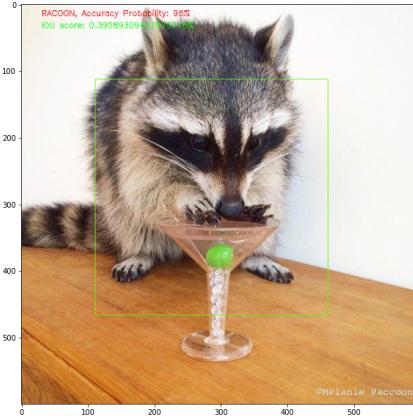


Figure 4: Result of raccoon dataset 59

4 Conclusion

When we look at the results above, we see our accuracy value as high. When we look at our IoU value, it is seen that it intersects with the real values at a high rate, but the estimations are small in terms of box area. It can be said that the results are positive.

There are two weaknesses in this project. Due to the scarcity of the data set, adequate development cannot be achieved during the training phase. If the data set is diversified, more effective and more successful results can be obtained. Secondly, in the non-maximum suppression algorithm, the frames should be merged on a small scale with an efficient algorithm. This will increase the iou score by enlarging the predicted boxes.

In this assignment process, we discussed both classification and localization problems together. We saw the performance of the model we developed with two heads on the Racoon dataset. We saw how well the model was trained with the correct number of epochs, number of learning rates and batch size. Positive comments can be made for results and gains.

Note : All data can be found in the drive folder.



Figure 5: Sliding window frames of the raccoon dataset 63



Figure 6: Result of raccoon dataset 63

5 References

- [1] https://pytorch.org/hub/pytorch_vision_resnet/
- [2] <https://public.roboflow.com/object-detection/raccoon/2/download/yolov4pytorc>
- [3] https://www.youtube.com/watch?v=nDPWwyWRIRo&list=PLf7L7Kg8_FNxHATtLwDceyh72QQL9pvpQ&index=12

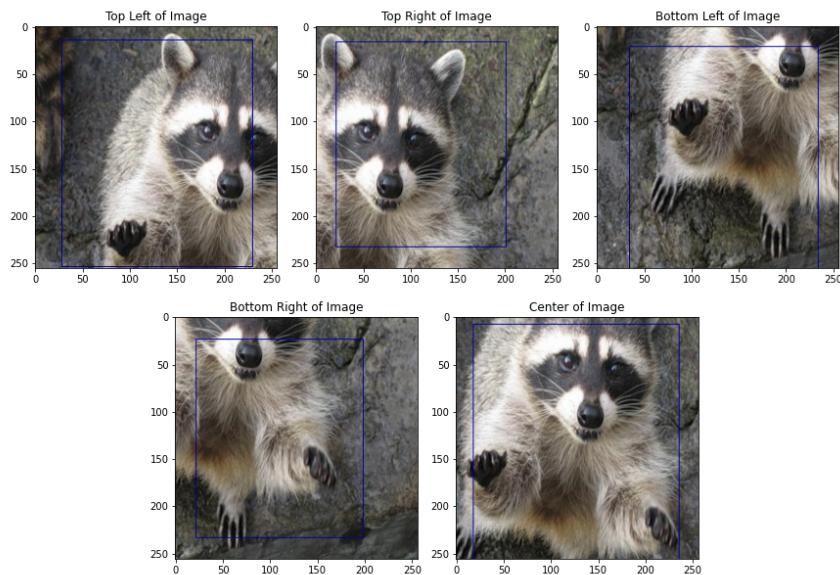


Figure 7: Sliding window frames of the raccoon dataset 70

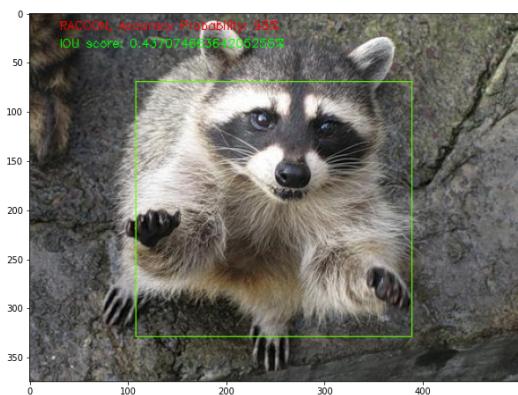


Figure 8: Result of raccoon dataset 70