ELEC 474 Lab 4 - YODA

ELEC 475 Prof. Michael Greenspan

Monday, November 20th, 2023

Mile Stosic (20233349)

Kieran Cosgrove (20226841)

Table of Contents

Table of Figures

1 Model

In the third step of the lab, a ResNet-18 model, sourced from the PyTorch model library, was employed for the task of car detection. The loss plot, which illustrates the model's learning curve over 40 epochs, displays a significant decrease in the training loss, indicative of the model's improving performance on the training set.

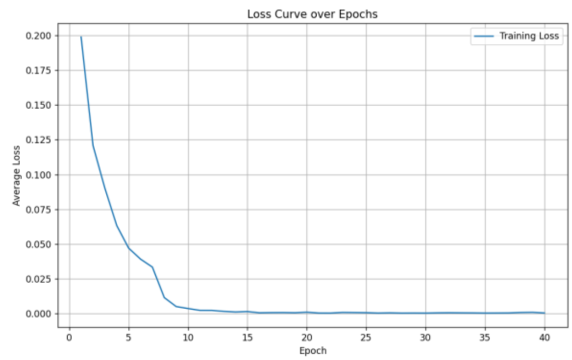

Figure 1: Loss Plot

The model had an overall accuracy of approximately 89.85%, which is a strong indicator of its capability to distinguish between the 'Car' and 'NoCar' classes. The accompanying confusion matrix provides a more detailed view of the model's performance, showcasing the number of true positives, false positives, true negatives, and false negatives.

```
Accuracy:  0.8985342111186135
```

Figure 2: Overall Accuracy

Specifically, the confusion matrix and the classification report reveal that the model has a precision of 0.89 when predicting the 'NoCar' class and a significantly higher precision of 0.96 for the 'Car' class. However, the recall for the 'Car' class is lower at 0.61, suggesting that the model is more conservative when predicting this class, potentially leading to a higher number of false negatives. The F1 scores, which balance precision and recall, are 0.94 for 'NoCar' and 0.75 for 'Car', further supporting this observation.

```
F1 score:  0.7497137305250009
Confusion matrix:
[[53072   421]
 [ 6792 10803]]
Classification report:
              precision    recall  f1-score   support

           0       0.89      0.99      0.94     53493
           1       0.96      0.61      0.75     17595

    accuracy                           0.90     71088
   macro avg       0.92      0.80      0.84     71088
weighted avg       0.91      0.90      0.89     71088
```

Figure 3: Confusion Matrix

This analysis is critical as it indicates areas where the model excels and where it may require further tuning or additional training data to improve its predictive capabilities, especially in reducing the number of false negatives for car detection.

2 Results

Despite the high overall accuracy in classifying 'Car' versus 'NoCar', the application of the model to the Kitti dataset revealed challenges in spatially localizing cars in the images. The bounding boxes, when overlayed on the test images, were not consistently accurate. As observed in the figures below, the bounding boxes were often able to identify the cars, but the boxes were significantly larger than the car or sometimes would entirely miss the cars. These qualitative results suggest a discrepancy between the model's classification accuracy and its ability to precisely localize objects in a real-world scenario.

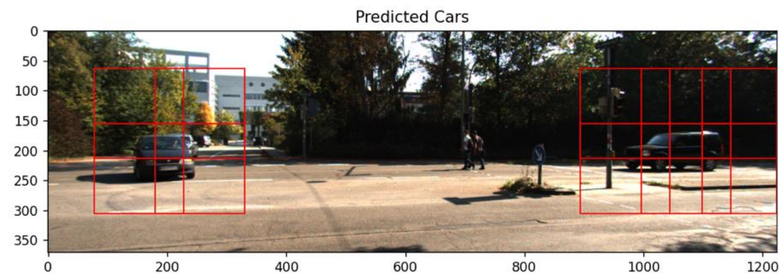

Figure 4: Qualitative Example 1

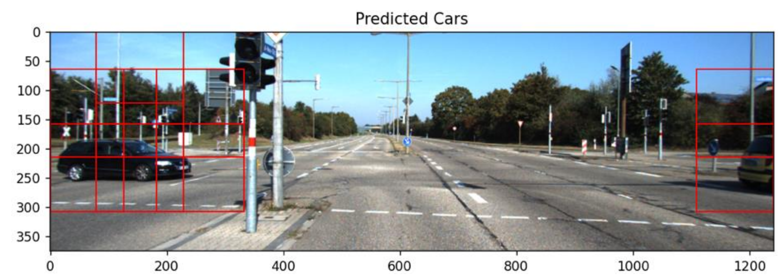

Figure 5: Qualitative Example 2

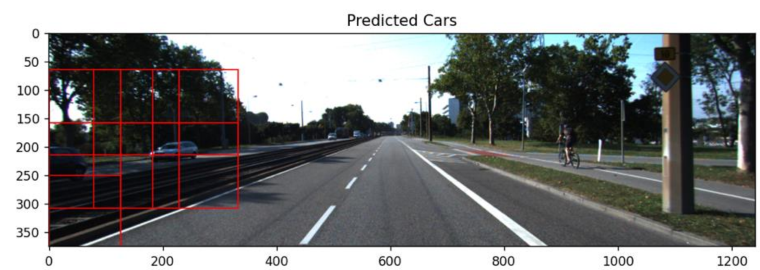

Figure 6: Qualitative Example 3

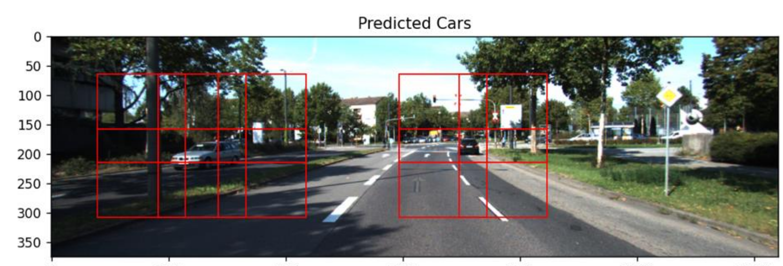

Figure 7: Qualitative Example 4

2.1 Intersection over Union Results

The IoU score for the car ROIs was 0.18. This low IoU score further corroborates the qualitative observations. It indicates that the bounding boxes predicted by the model don't align completely with the actual location and shape of the cars in the images. This low overlap unfortunately reflects a fundamental issue in the model's detection capabilities.

```
Mean IoU for 'Car' ROIs: 0.1832466270711659
```

Figure 8: Terminal Output Value of IoU

3 Discussion

In this project, we utilized a ResNet-18 model to detect cars in images. While the model showed high accuracy in classifying 'Car' versus 'NoCar', the transition to practical application, particularly in localizing objects within the Kitti dataset, revealed a significant gap in performance. The qualitative analysis, involving the overlay of bounding boxes on the test images, illustrated this challenge vividly. Despite the model's strong classification capabilities, as indicated by an accuracy of approximately 89.85%, it struggled to accurately localize cars within the images. This was further corroborated by the low IoU score of 0.1832466, indicating minimal overlap between the predicted and actual locations of the cars.

The primary challenge we faced was the model's inability to accurately delineate the boundaries of cars, suggesting a potential overfitting to the training dataset or a limitation in the model's architecture to generalize to new, real-world scenarios. In response, we experimented with different parameter adjustments, such as threshold values for bounding box predictions and increased training epochs. However, these modifications did not lead to substantial improvements, which hinted at the need for a more sophisticated model or a different approach altogether for object localization tasks.

This experience highlighted the distinction between classification and localization in computer vision. While our model could effectively differentiate between 'Car' and 'NoCar' classes, its performance in spatial detection was limited. This underscores the importance of selecting the right model and approach for specific tasks in computer vision. Incorporating a more diverse training dataset, including images from various scenarios and conditions, could help in enhancing the model's ability to generalize its detection capabilities to different environments.

Through this project, we gained invaluable insights into the complexities of computer vision tasks. We learned that high accuracy in one aspect, such as classification, does not necessarily translate to high performance in all areas, such as localization. These findings are crucial for guiding future research and applications in the field, helping to develop more accurate and versatile computer vision systems.