Analysis of Faster R-CNN for Object Detection

Nemekhbayar Nomin[1],
[1]Computer Science Engineering Dep. of PNU.

1 Introduction

Faster R-CNN is an object detection framework combining region proposal with object classification and localization. This analysis aims to evaluate the effectiveness of Faster R-CNN on the given dataset by examining key performance metrics and loss functions, such as RPN classification and regression losses as well as RoI head losses on a custom dataset, focusing on metrics such as mean Average Precision (mAP), classification accuracy, and localization error.

2 Dataset

The dataset consists of two subsets for training and testing which consisted of 4,884 images and 4,872 images respectively. Both subsets are accompanied by JSON metadata files (`train.json` and `test.json`), which include details such as file names, image dimensions, and unique identifiers for each image. These annotations are essential for the training process, allowing the model to learn object locations and classifications.

3 Network Architecture

We use the Faster R-CNN model, which integrates the Region Proposal Network (RPN) with Fast R-CNN, to detect and classify objects within an image. The key components of Faster R-CNN include:

- **Feature Extractor**: Utilizes a backbone network (e.g., ResNet-50) to extract features from input images.
- **RPN**: Proposes candidate object regions.
- **ROI Pooling**: Warps the candidate regions to a fixed size and pools features.
- **Fully Connected Layers**: Classifies regions and refines bounding box coordinates.

4 Model Configuration

The model was set up with the following hyperparameters: Learning Rate: 0.001 (decayed by 0.1 every 10 epochs), Weight Decay: 0.0005, Optimizer: SGD with momentum 0.9, Batch Size: 1 (GPU limit), Epochs: 14, RPN IoU Thresholds: Positive 0.7, Negative 0.3, RoI Positive Ratio: 0.25, Loss Function: Smooth L1 for localization, Cross-Entropy for classification.

5 Training Analysis

The training process was executed for 14 epochs, tracking four types of losses: RPN classification, RPN localization, RoI classification, and RoI localization. Below is the comparison table between first and last epochs.

RPN Loc Loss	RPN Cls Loss	ROI Loc Loss	ROI Cls Loss	Total Loss
0.1546	0.2883	0.3086	0.5285	1.2802
0.1156	0.1379	0.1726	0.2610	0.6872

Table 1: Loss values for RPN and ROI at Epoch 1 and Epoch 14.

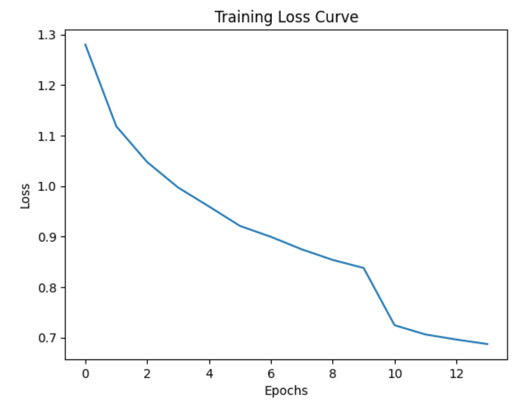

Figure 1: Training loss curves for RPN classification, RPN regression, RoI classification, and RoI regression.

6 Evaluation Metrics and Results Analysis

The evaluation of the Faster R-CNN model during training involves both classification and localization metrics:

- **Confusion Matrices:** RPN Confusion Matrix, ROI Confusion Matrix
- **Loss Metrics:** RPN Localization Loss, RPN Classification Loss, ROI Localization Loss, ROI Classification Loss

Below is the illustration of the trained model on unseen sample image.

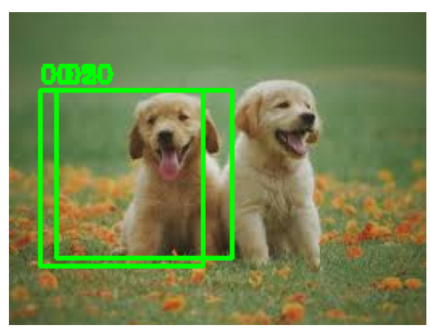

Figure 2: Example Object Detection

7 Discussion

While the Faster R-CNN model demonstrated robust classification capabilities, its localization accuracy was impacted by the complexity of the dataset. Adjustments to anchor box sizes and IoU thresholds could enhance performance. Moreover, increasing the number of proposals or using deeper backbones (like ResNet-101) may further improve mAP scores.

8 Conclusion

The Faster R-CNN model provides competitive performance for object detection on the custom dataset, achieving a satisfactory mAP and classification accuracy. However, further tuning, especially in the RPN, could reduce false positives and improve localization accuracy.