Semantic Segmentation on MUAD Dataset: Experiments and Analysis

Team: Yulong MA, Boyuan ZHANG, Huanshan HUANG

Abstract

This report summarizes the experiments conducted on Semantic Segmentation using the MUAD dataset. The primary objective was to train a neural network to segment urban driving scenes into semantic categories. The experiments involved training a UNet model, analyzing its performance, and investigating the use of Deep Ensembles for uncertainty estimation and Out-of-Distribution (OOD) detection. A key focus of the analysis is the observation of overfitting during the training process.

1 Experiment Background

The dataset used for these experiments is the MUAD (Multiple Uncertainties for Autonomous Driving) dataset, specifically the small version configured for this task. MUAD is a synthetic dataset designed to simulate diverse and challenging driving conditions, providing ground truth for semantic segmentation tasks. The dataset follows the standard Cityscapes label set with 19 semantic classes.

2 Technical Detail

2.1 Model Architecture

The segmentation model employed is a UNet architecture. UNet is a fully convolutional network consisting of an encoder (contracting path) and a decoder (expanding path).

- **Encoder:** Captures context via downsampling convolutional blocks.
- **Decoder:** Enables precise localization via upsampling and skip connections that concatenate features from the encoder.
- **Dropout:** A dropout rate of 0.1 was applied for regularization.

Figure 1: UNet Architecture Diagram

2.2 Training Setup

The model was trained using the following hyperparameters and configuration:

- **Optimizer:** Stochastic Gradient Descent (SGD) with a momentum of 0.9 and weight decay of 1×10^{-4}.
- **Loss Function:** Cross Entropy Loss.
- **Learning Rate Scheduler:** StepLR with a step size of 10 epochs and a decay factor (gamma) of 0.1. The initial learning rate was 0.01.
- **Epochs:** The training was conducted for 30 epochs.
- **Metric:** Mean Intersection over Union (mIoU) was calculated for both training and validation sets.

3 OOD Methods

To improve robustness and estimate uncertainty, a Deep Ensemble was constructed using 5 independently trained UNet models. The ensemble prediction was obtained by averaging the softmax probabilities of the individual models.

Predictive entropy was used as a metric for uncertainty estimation. The entropy maps generated from the ensemble predictions highlighted high uncertainty in two key areas:

1. **Object Boundaries:** High entropy was observed at the edges of objects, representing aleatoric uncertainty due to the inherent ambiguity in transition regions.
2. **OOD Regions:** Ideally, out-of-distribution objects should exhibit higher entropy compared to known classes. This property is leveraged for OOD detection, where a threshold on entropy can be used to reject uncertain predictions.

Figure 2: Predictive Entropy Map for OOD Detection

4 Neural Collapse (NC)

Towards the end of the training process in classification and segmentation tasks, feature representations often exhibit Neural Collapse (NC). Given that the model heavily overfitted on the training data (reaching a training mIoU of approximately 0.75), we hypothesize that features belonging to the same semantic class collapsed towards their respective class centers in the feature space. However, due to the lower validation mIoU, this structural collapse did not generalize effectively to unseen data. Future research could analyze feature-level variability to verify how NC impacts the setting of OOD detection thresholds.

5 Results Analysis

The training process was monitored using Cross Entropy Loss and Mean Intersection over Union (mIoU).

- **Loss:** Both training and validation losses showed a downward trend, stabilizing around 0.3.
- **mIoU and Overfitting:** A significant discrepancy was observed between the training and validation performance. The training mIoU reached approximately 0.75, whereas the validation mIoU plateaued at around 0.65.

Comment on Overfitting: This gap of approximately 0.10 in mIoU indicates that the model is suffering from overfitting. It has effectively memorized patterns specific to the training data that do not generalize well to the unseen validation set. While the validation performance is reasonable, the divergence suggests that further regularization or more training data could improve generalization.

Ensemble Performance: The Deep Ensemble achieved a Mean IoU of 0.6558 on the In-Distribution (ID) validation set. This performance is consistent with the single-model validation results, illustrating that ensembling maintains predictive accuracy while providing the added benefit of uncertainty quantification.

Figure 3: Training and Validation mIoU indicating Overfitting

6 Conclusion

The experiments successfully demonstrated the application of UNet for semantic segmentation on the MUAD dataset. While the model achieved a respectable validation mIoU, the analysis of training curves revealed clear signs of overfitting. The implementation of Deep Ensembles provided a robust mechanism for uncertainty estimation, yielding an ID mIoU of 0.6558 and enabling potential Out-of-Distribution detection through predictive entropy. Future work could focus on stronger regularization techniques to mitigate overfitting, further refining OOD detection thresholds, and investigating the specific impacts of Neural Collapse on feature distribution.