Project Report

FMML Capstone Project

Semantic Segmentation for Indian Driving Dataset (IDD)

Name: Sadiya Maheen Siddiqui

HUB ID: HUB20240169

Problem Statement

Autonomous driving in unstructured environments, such as Indian roads, poses unique challenges due to varied traffic behaviour, inconsistent infrastructure, and diverse visual scenes. Semantic segmentation - classifying each pixel into predefined categories - is essential for understanding such complex driving environments. This project addresses the problem of pixel-level semantic segmentation for 26 Level 3 classes using the IDD dataset, contributing to improved perception systems for autonomous navigation in Indian scenarios.

Objective

- To develop a deep learning model capable of accurate semantic segmentation on the IDD dataset.
- To evaluate model performance using the Mean Intersection over Union (mIoU) metric.
- To optimize training using techniques such as model checkpointing, loss monitoring, and fine-tuning.

Datasets Used

- **IDD 20K (Part I & II)**: Comprising over 20,000 images captured across various cities and conditions in India.
- **Label Format**: Level 3 annotations with 26 classes + 1 miscellaneous class (total: 27 class IDs).
- **Resolution**: Images resized to 1280 x 720 for training and evaluation.

Methodology

1. Data Preparation

- **Resizing:** All images and their corresponding masks were resized to 256 x 256 pixels to ensure compatibility with the model and to speed up training.
- **Data Augmentation**: To enhance the model's generalization ability, several augmentation techniques were applied:
    - Random flips
    - Random rotations
    - Colour jittering
- **Data Splitting**: The dataset was split into 85% for training and 15% for validation to evaluate model performance during training.

2. Model Architecture

Model: DeepLabV3+ with MobileNetV2 Backbone

- Atrous Spatial Pyramid Pooling for multi-scale context.
- Encoder-decoder structure for refined boundary segmentation.
- Lightweight backbone for efficient training.

The model was chosen for its strong track record in semantic segmentation and its balance between accuracy and computational efficiency.

3. Training Process

a. Loss Function

- Categorical Cross Entropy with Masking

b. Optimizer and Learning Rate

- **Optimizer:** Adam
- **Initial Learning Rate:** 1e-3

c. Batching and Epochs

- **Batch Size:** 32

- **Total Epochs: 10**
- **Training Samples:** 372 steps/epoch (85% of ~ 6993 images)
- **Validation Steps:** 66 (15% of ~ 6993 images)

4. Evaluation

a. Metric: Mean Intersection over Union (mIoU)

The model's performance was evaluated using mIoU, which measures the overlap between predicted masks and ground truth:

- **Per-Class IoU:** Computed for each of the 26 classes.

$$\text{IoU} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive} + \text{False Negative}}$$

- **mIoU:** The average of all per-class IoUs.
- **Best validation mIoU:** 0.8520

b. Qualitative Evaluation

In addition to numerical metrics, the following qualitative evaluations were conducted:

- Visualized predictions to ensure the model correctly segmented critical classes like roads, pedestrians, and vehicles.
- Analysed failure cases to identify common errors (e.g., confusion between pedestrians and riders).

5. Prediction

Training Performance:

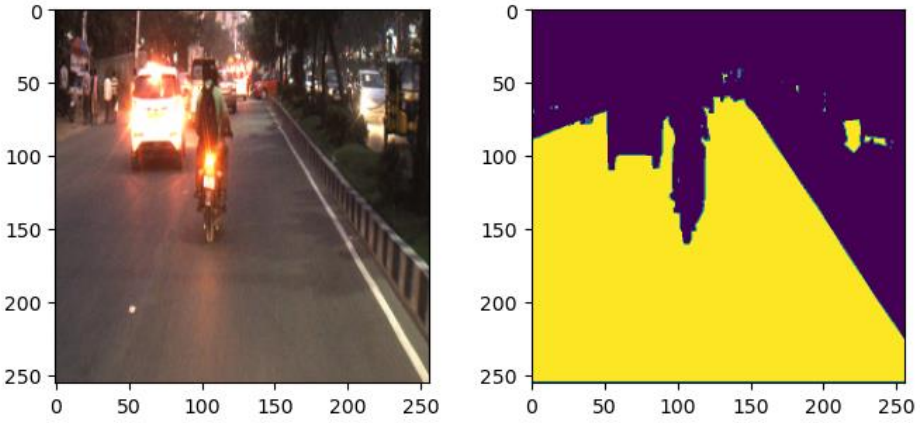
Observations:

- Peak validation mIoU achieved at Epoch 8.
- Slight overfitting after Epoch 8 as validation loss increases despite decreasing training loss.

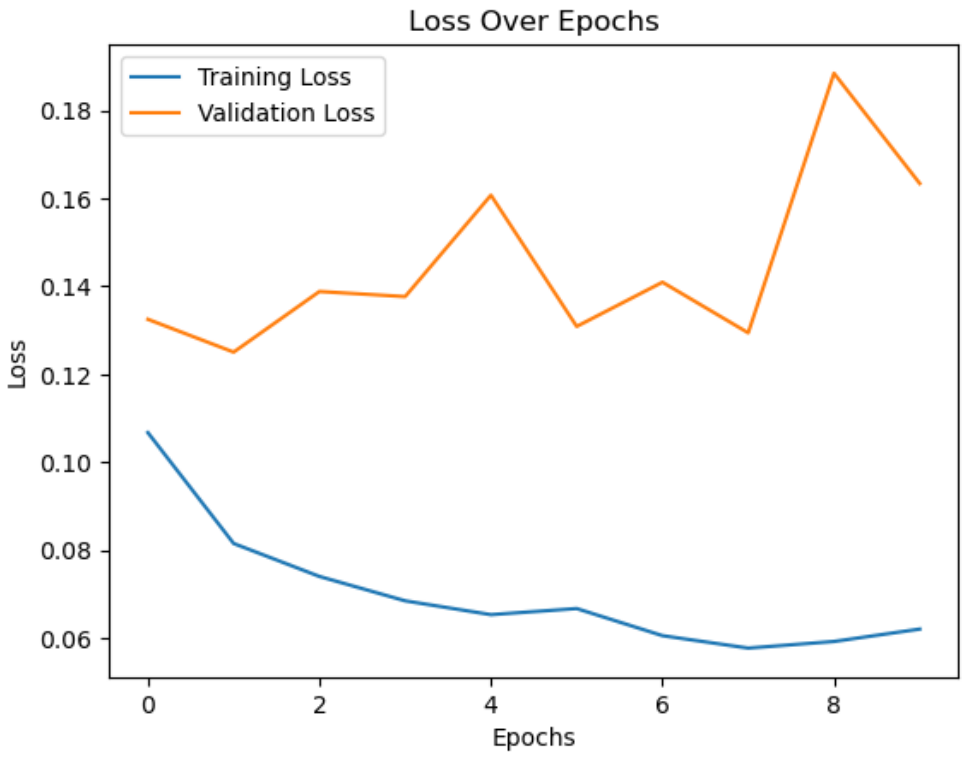
Epoch	Training Loss	Accuracy	mIoU	Val Loss	Val Accuracy	Val mIoU
1	0.1067	0.9559	0.8587	0.1325	0.9518	0.8396
5	0.0653	0.9714	0.9073	0.1607	0.9501	0.8368
8	0.0577	0.9746	0.9185	0.1294	0.9563	0.8520

Visual Results

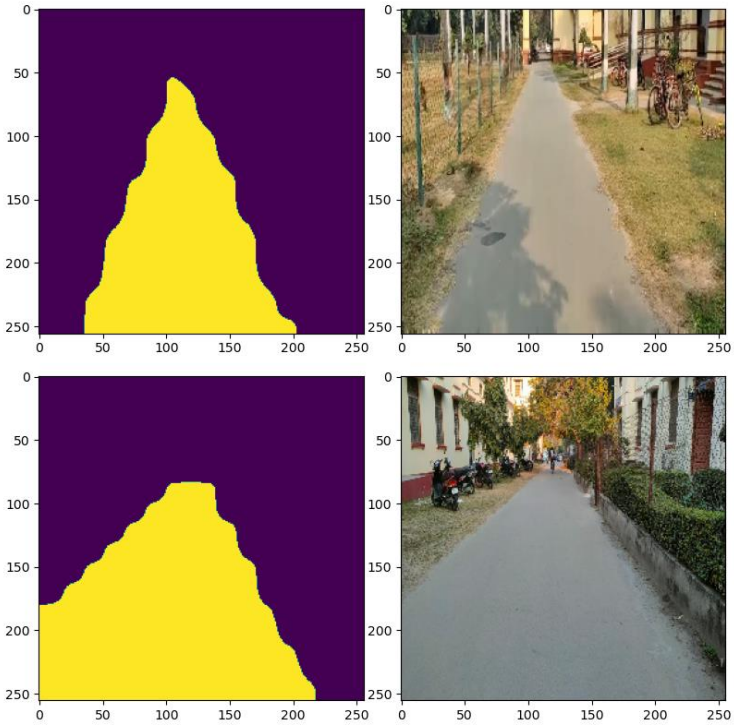
- **Training Image + Ground Truth Mask**



- **Graph: Training and Validation Loss vs Epochs**



- **Test Image + Predicted Mask**



Future Work

- Fine-tune on higher resolution images (512x512 or full 1280x720).
- Experiment with transformer-based architectures (e.g., SegFormer, Mask2Former).
- Perform domain adaptation across other driving datasets like Cityscapes or BDD100K for generalized performance.

Conclusion

This project effectively applied semantic segmentation techniques to the Indian Driving Dataset using DeepLabV3+. The model achieved a validation mIoU of 0.852, indicating strong performance in understanding complex road scenes. The modular pipeline ensures easy extension to additional backbones and datasets.