UNet Model Training/Finetuning (Subhasis Panda)

1 Model Architecture

1.1 UNet from Scratch

• The architecture follows the classic UNet design with an encoder-decoder structure and skip connections.
• The encoder downsamples using Conv2D layers followed by ReLU activations and max-pooling.
• The decoder upsamples using transposed convolutions and concatenates features from the encoder to preserve spatial information.
• UNet was chosen for its ability to accurately capture both spatial and contextual information through symmetric skip connections, making it highly effective for pixel-wise segmentation tasks..

1.2 Key Hyperparameters

• **Input size**: 256x256.
• **Number of classes**:7 in our case.
• **Padding**: 'same' to preserve input-output shape.

2 Model Training

• **Dataset**: Processed Cityscapes images and remapped masks.
• **Loss Function**: CrossEntropyLoss.
• **Optimizer**: Adam with a learning rate of 0.001.
• **Training Time**: 2 hours and 16 mins.

3 Model inference

• Inference was done using a subset of the validation dataset.
• Visualizations compared predicted masks vs. ground truth using matplotlib.
• Test Accuracy and per class IOUs are as follows: As we can see(from figure 1 and 2), due to the

```
Test Accuracy: 0.8887
Class 0: IoU = 0.6536
Class 1: IoU = 0.8968
Class 2: IoU = 0.7898
Class 3: IoU = 0.8034
Class 4: IoU = 0.4864
Class 5: IoU = 0.8467
Class 6: IoU = 0.8761
```

Figure 1:

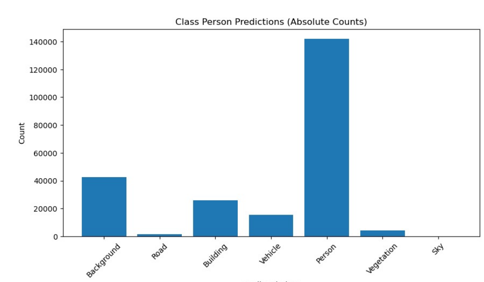

Figure 2:

class imbalance problem, the IOU for person is much less than other classes. So finetuning is needed to overcome this.

4 Model Fine-tuning

4.1 Weighted Cross-Entropy + Occlusion Augmentation

• Implemented class imbalance correction by assigning higher weights to under-represented classes.
• Added occlusion-based augmentations (random black boxes) to improve robustness.

- Observed improved generalization and stability in validation loss.(as in the below figure)

```
Test Accuracy: 0.8988
Class 0: IoU = 0.6801
Class 1: IoU = 0.9066
Class 2: IoU = 0.8078
Class 3: IoU = 0.8139
Class 4: IoU = 0.4792
Class 5: IoU = 0.8595
Class 6: IoU = 0.8701
```

Figure 3: Test Acuracy and Per class IOUs after fine-tuning(Weighted CE and Occlusion Augmentation)

Problem Encountered:Though the accuracy increased for this fine tuning but per class IOU(for person) reduced.

4.2 Combined CE + Dice Loss + Occlusion Augmentation

- Combined CrossEntropy with DiceLoss to improve boundary detection and handle class imbalance.
- Augmentations further improved edge segmentation.
- Notable improvement in Intersection over Union (IoU) scores across classes.(as in the below figure)

```
Class 0: IoU = 0.6810
Class 1: IoU = 0.9072
Class 2: IoU = 0.8072
Class 3: IoU = 0.8128
Class 4: IoU = 0.5176
Class 5: IoU = 0.8585
Class 6: IoU = 0.8828
```

Figure 4: Per Class IOUs with aan accuracy of 0.8999

5 Computational Resources

System Specs: • OS: Windows 11 • GPU: GPU P100(kaggle GPU) • RAM: 16 GB **Training Time**: • UNet from scratch:2 hr 16 mins • Finetuning with weighted class Cross entropy and OCclusion Augmentation : 48 mins 32 seconds • Finetuning with Combined CE abd DICE loss with occlusion Augmentation : 53 minutes 52 seconds.

6 Reproducibility

GitHub Repo Linkhttps://github.com/subhasisp1/Semantic-Segmentation-Cityscapes-Training-Finetuning
Instructions to reproduce:
• **Clone the repo**:
git clone https://github.com/your-username/your-repo-name.git
cd your-repo-name
• Create a virtual environment:
python3 -m venv venv
source venv/bin/activate
pip install -r requirements.txt
• Run preprocessing python script from https://github.com/subhasisp1/Semantic-Segmentation-Cityscapes-Prepr
• Run training notebook: **(Model Training UNet From scratch.ipynb)**

• Run fine-tuning notebooks: **(Finetuning with Weighted CE and Augmentation.ipynb)**, **(Fine tuning with combined CE and DICE Loss.ipynb)**

7 Output Visualization

Below are some sample outputs for visualization produced after final fine-tuning.

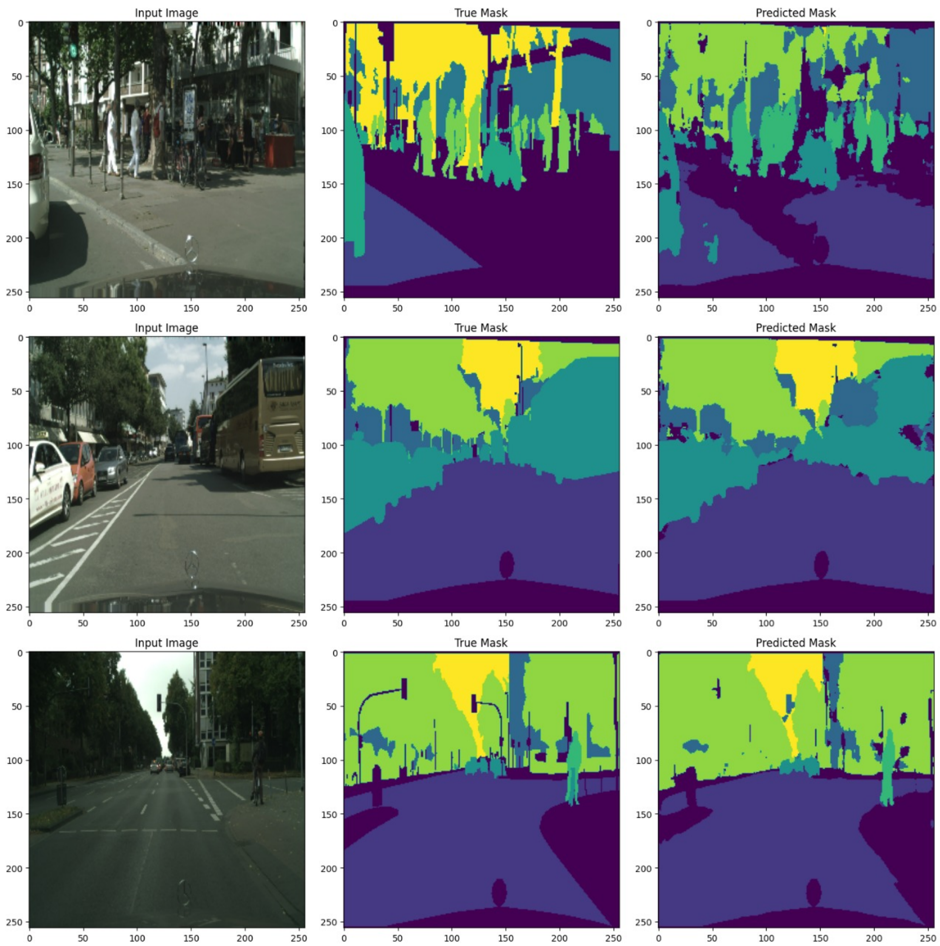

Figure 5: Test Image, GT Mask, and Predicted ask

8 References

- R. Fong and A. Vedaldi, "Occlusions for Effective Data Augmentation in Image Classification," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), 2019, pp. 4158-4166, doi: 10.1109/ICCVW.2019.00511.
- Claude, Deepseek