# Semantic Segmentation using U-Net

Computer Vision

**Assignment 4**

Varun Edachali (2022101029)

January 2025

All of the models were trained for 100 epochs with a learning rate of 3e-4 in the final run.

This is because on running with a high learning rate (1e-3 and 5e-4) the training seemed unstable, particularly for the model with no skip connections, while for lower learning rates (1e-4) there was no sufficient convergence. These runs are available on `WandB`, but the 3e-4 epoch run is shown below.

**Note**: if checking `WandB` - the first four runs are on lr 1e-3, the next on 3e-4, then 5e-4, and back to 3e-4. I began logging the learning rate once I realised how sensitive the learning rate is to it.

The checkpoints of the final run can be seen here, and the results can be seen below:

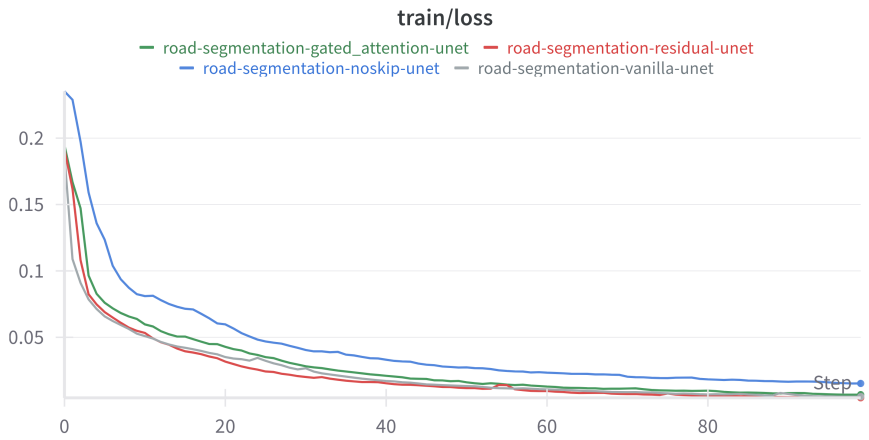


Figure 1: training loss

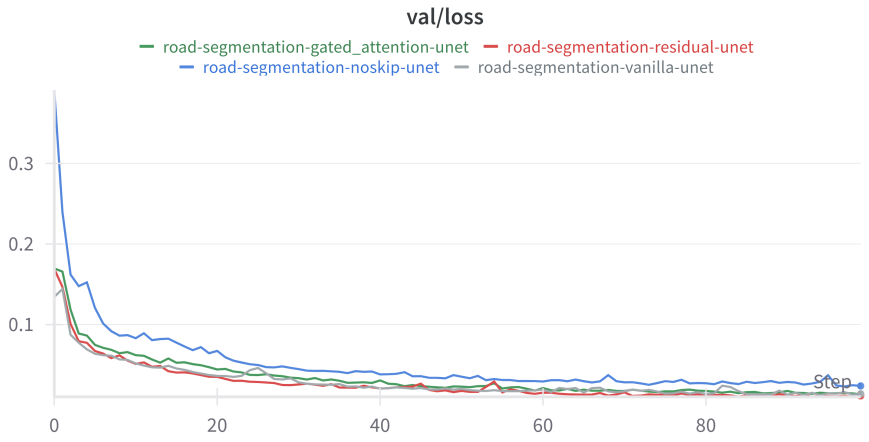


Figure 2: validation loss

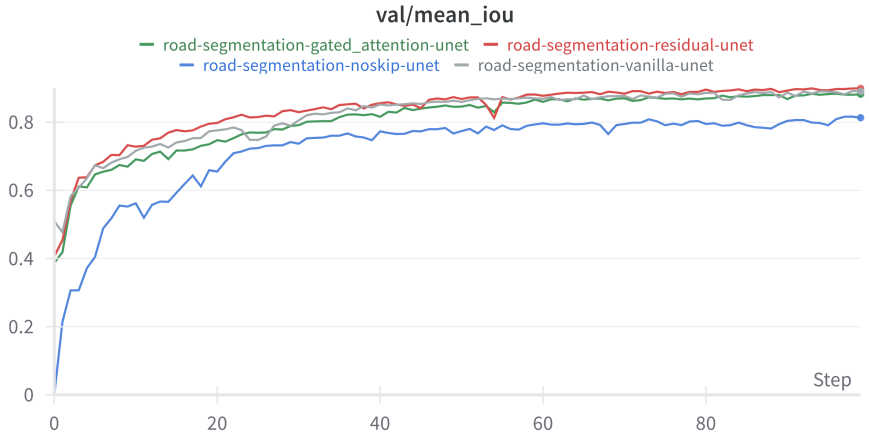


Figure 3: validation mIoU

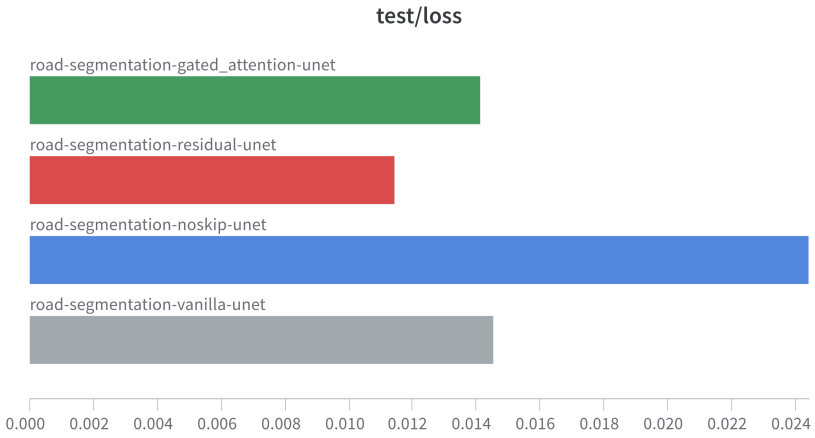


Figure 4: test loss

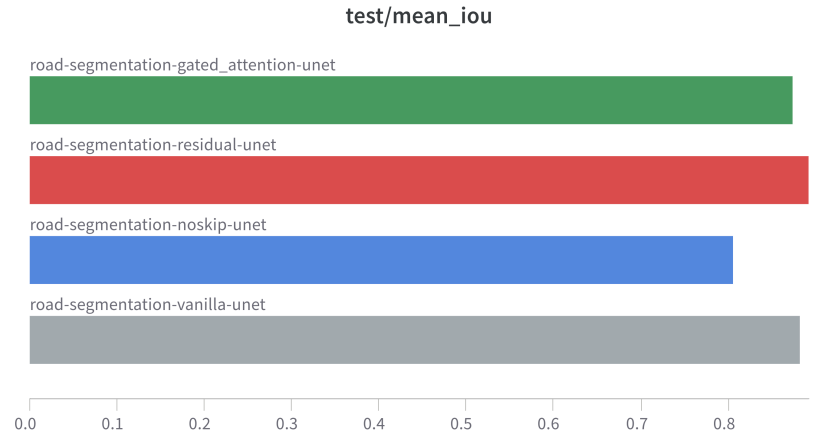


Figure 5: test mIoU

| Name (4 visualized) | State | Use | Runtir | learning rate | test/loss | test/mean_iou | train/loss | val/loss | val/mean_iou |
|---|---|---|---|---|---|---|---|---|---|
| road-segme...ention-unet | Finished | varun-e | 1h 44m 39 | 0.0003 | 0.014122 | 0.87399 | 0.006795 | 0.013398 | 0.88206 |
| road-segme...sidual-unet | Finished | varun-e | 1h 48m 52 | 0.0003 | 0.011432 | 0.89218 | 0.0044184 | 0.011457 | 0.89943 |
| road-segme...noskip-unet | Finished | varun-e | 1h 21m 7s | 0.0003 | 0.024418 | 0.80512 | 0.015276 | 0.023871 | 0.81303 |
| road-segme.../anilla-unet | Finished | varun-e | 1h 35m | 0.0003 | 0.014524 | 0.88269 | 0.0053194 | 0.014134 | 0.89048 |

Figure 6: consolidated results

For each model, we visualise the predicted and ground truth masks for 8 images from the test set. Two of each is shown here, while the rest can be seen in the repository.

# 1 Vanilla U-Net

The model was implemented as per the paper and some online references shown in the code and the `docs` directory.

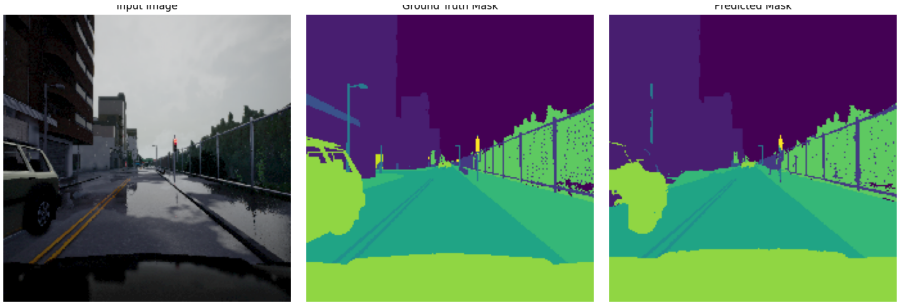

Figure 7: prediction on image 0 - vanilla

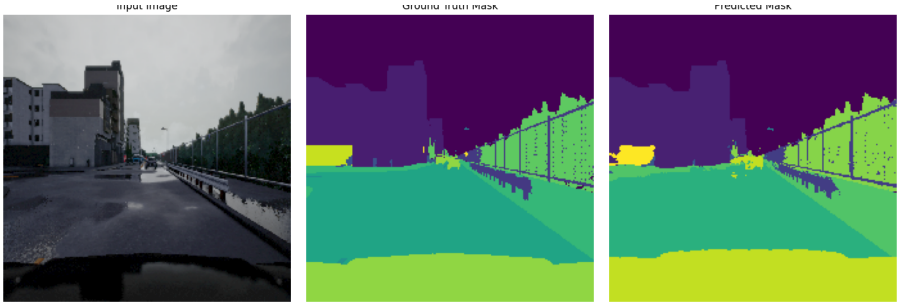

Figure 8: prediction on image 7 - vanilla

# 2 U-Net without Skip Connections

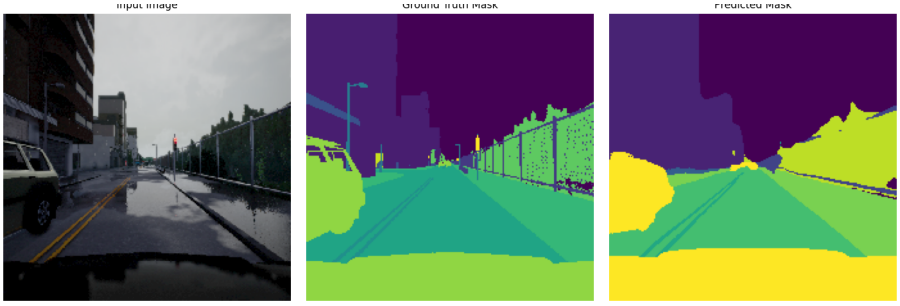

Figure 9: prediction on image 0 - noskip

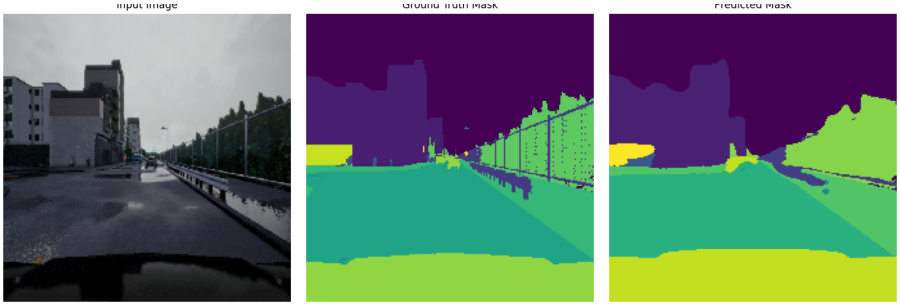

Figure 10: prediction on image 7 - noskip

## 2.1 What differences do you observe in the visualised results compared to the standard U-Net results?

The U-Net without skip connections struggles to capture fine-grained spatial detail (see, for example, the fence on the right) as compared to the one with them. Also, the performance is worse.

## 2.2 Discuss the importance of skip connections in U-Net. Explain their role in U-Net's performance.

Skip connections primarily help in preserving spatial detail by transferring high-resolution feature maps from the encoder directly to the corresponding decoder layers. This helps restore fine details. This can be observed in the above image: the U-Net with skip connections captures much more fine-grained details than the one without (see: the fence on the right, for example).

Trivially, they also help improve gradient flow by mitigating the vanishing gradient problem by providing alternative paths for the gradient to propogate through the network. This facilitates the training of deeper networks.

# 3 Residual U-Net

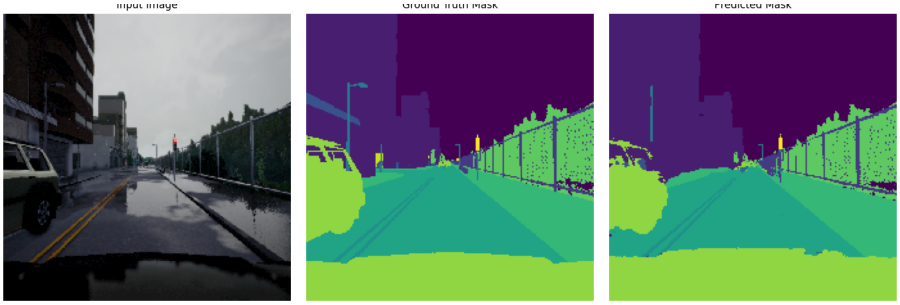

Figure 11: prediction on image 0 - residual

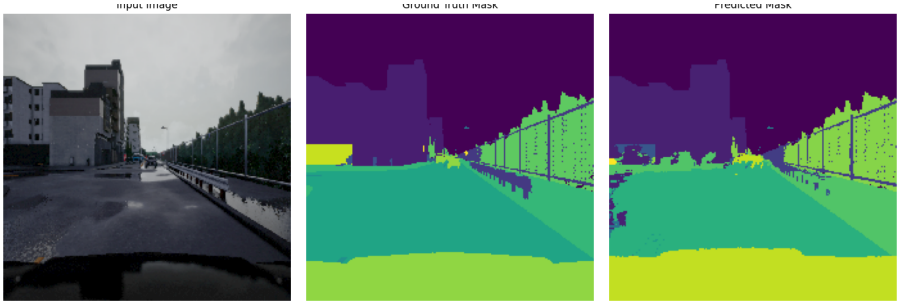

Figure 12: prediction on image 7 - residual

# 4 Gated Attention U-Net

The vanilla U-Net was used as the base for simplicity. The model was implemented as in the image and paper with the help of some online references shown in the code and the `docs` directory.

Figure 13: prediction on image 0 - gated attention

Figure 14: prediction on image 7 - gated attention

## 4.1 What are some advantages of using Attention gates as per the paper?

As in the `Introduction` of the paper:

- CNN models with AGs can be trained from scratch in a standard way similar to the training of a FCN model.
- AGs automatically learn to focus on target structures without additional supervision.
- At test time, these gates generate soft region proposals implicitly on-the-fly and highlight salient features useful for a specific task. Moreover, they do not introduce significant computational overhead and do not require a

large number of model parameters as in the case of multi-model frameworks.

- In return, the proposed AGs improve model sensitivity and accuracy for dense label predictions by suppressing feature activations in irrelevant regions. In this way, the necessity of using an external organ localisation model can be eliminated while maintaining the high prediction accuracy.

## 4.2 How does gating signal at skip connections help in improved performance?

- The gating signal for each skip connection aggregates information from multiple imaging scales which increases the grid-resolution of the query signal and achieve better performance (see: paper, page 5). In effect, the attention mechanism uses the gating signal to modulate the features coming through the skip connections - enhancing those that align with the higher-level context and suppressing others.

- the gating signal contains contextual information collected at a coarser scale (see: paper, page 4). This can help to prune low-level features (see: paper, page 4).

## 4.3 What differences in results do you observe as compared to the standard U-Net results? Discuss.

The most significant difference is that the gated attention model could predict some important items (eg: the car on the left of image 0) far more accurately than the standard U-Net model. This could be because the attention mechanism enhanced the focus on it.