Semantic Segmentation using FCNs

Computer Vision

Assignment 4

Varun Edachali (2022101029)

January 2025

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

1 Dataset Visualisation

We read the segmentation mask image, extract the first channel, isolate each of the 13 classes individually by creading binary masks, and visualise each binary mask.

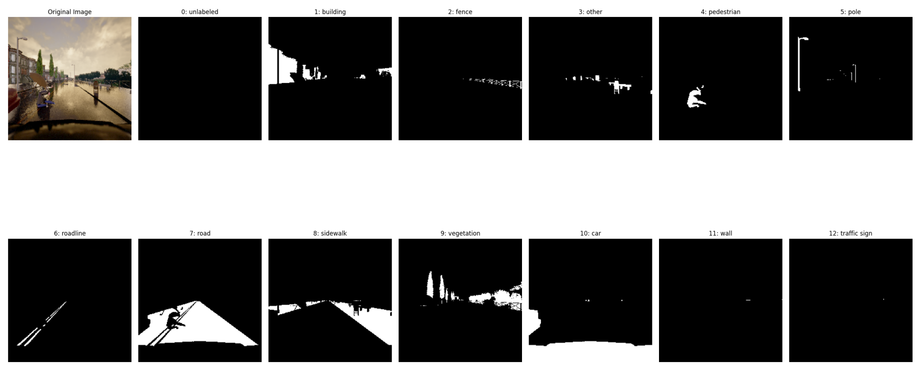


Figure 1: eda for fcn

2 FCN Variants

2.1 Frozen VGG Backbone

On freezing the VGG backbone, we see the following results:

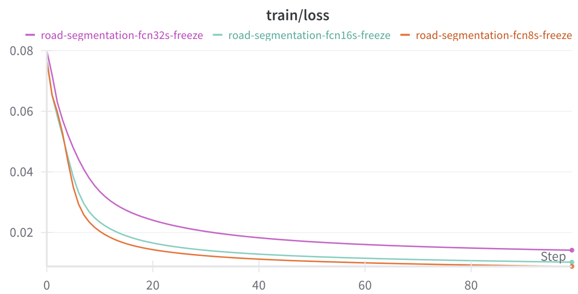


Figure 2: train loss - frozen backbone

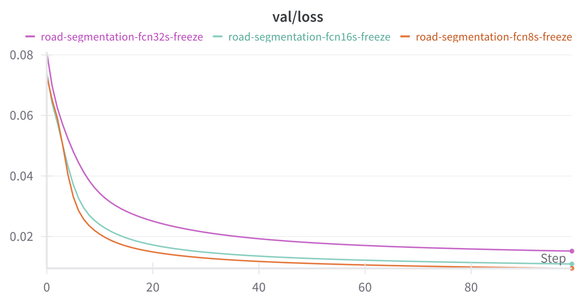


Figure 3: validation loss - frozen backbone

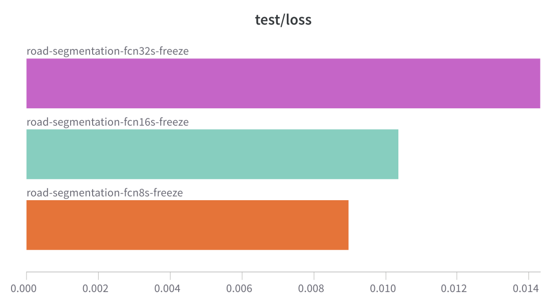


Figure 4: test loss - frozen backbone

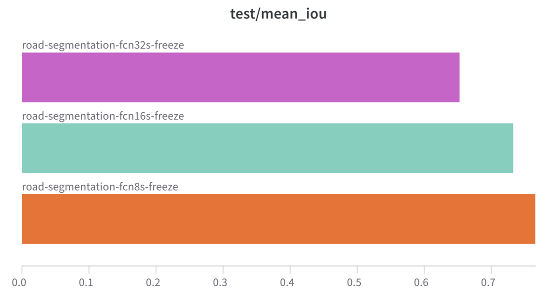


Figure 5: test mIoU - frozen backbone

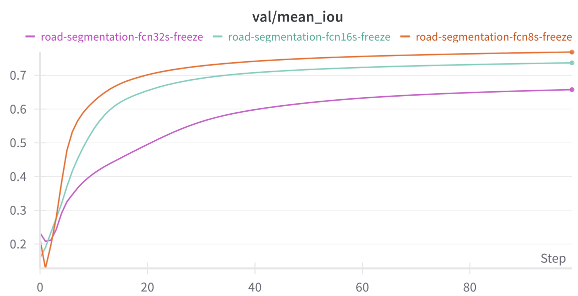
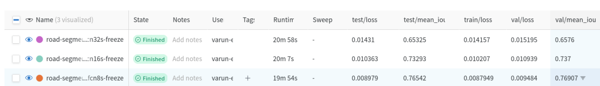


Figure 6: val mIoU - frozen backbone



Name (3 visualized)	State	Notes	Use	Tags	Runtim	Sweep	test/loss	test/mean_iou	train/loss	val/loss	val/mean_iou
road-segme...n32s-freeze	Finished	Add notes	varun-e		20m 58s	-	0.01431	0.65325	0.014157	0.015195	0.6576
road-segme...n16s-freeze	Finished	Add notes	varun-e		20m 7s	-	0.010363	0.73293	0.010207	0.010939	0.737
road-segme...fcn8s-freeze	Finished	Add notes	varun-e	+	19m 54s	-	0.008979	0.76542	0.0087949	0.009484	0.76907

Figure 7: compiled results - frozen backbone

In addition, we visualise these predictions on 8 images from the test set, one of which is shown here. These images are available in full resolution in the repository.

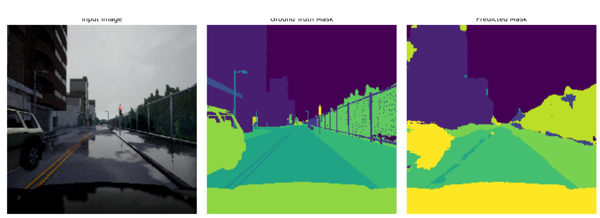


Figure 8: FCN8s - frozen backbone

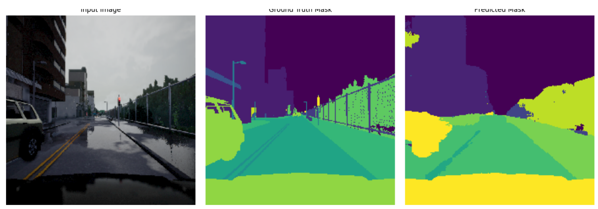


Figure 9: FCN16s - frozen backbone

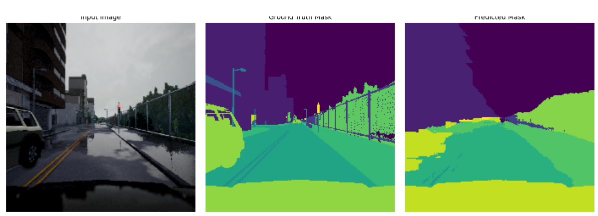


Figure 10: FCN32s - frozen backbone

2.2 UnFrozen VGG Backbone

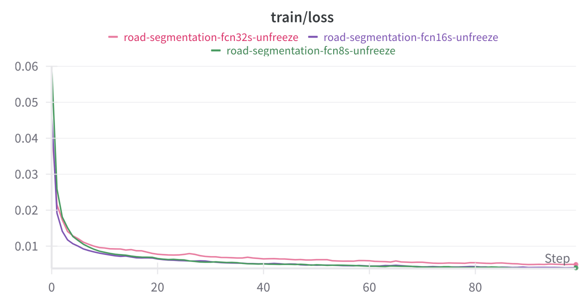


Figure 11: train loss - unfrozen backbone

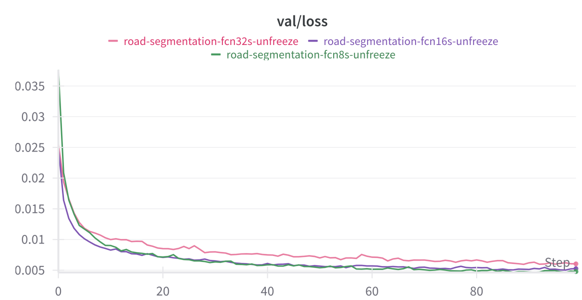


Figure 12: validation loss - unfrozen backbone

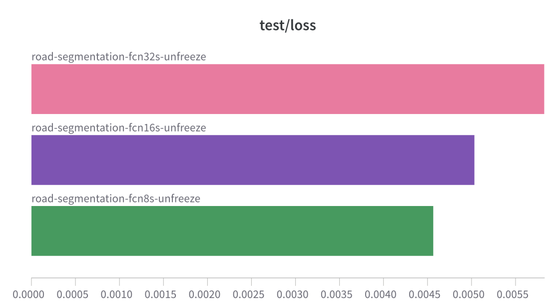


Figure 13: test loss - unfrozen backbone

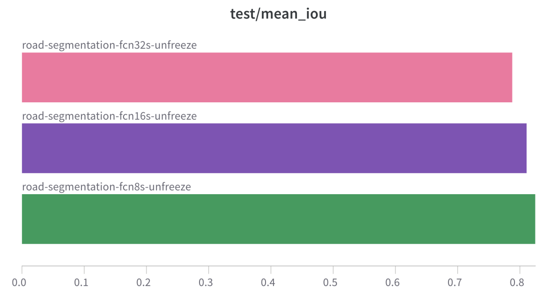


Figure 14: test mIoU - unfrozen backbone

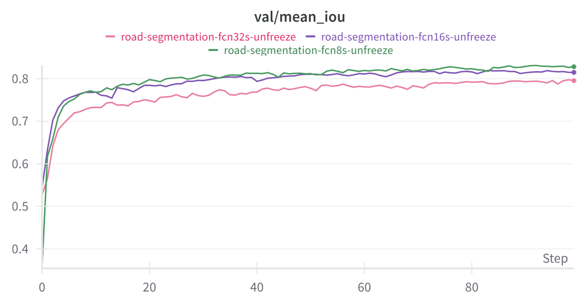


Figure 15: val mIoU - unfrozen backbone

Name (3 visualized)	State	Notes	Use	Tag:	Run	Sweep	test/loss	test/mean_iou	train/loss	val/loss	val/mean_iou
road-segmenta...cn32s-unfreeze	Finished	Add notes	varun-ε		34m 28s	-	0.0058223	0.789	0.0049128	0.006033	0.79533
road-segmenta...cn16s-unfreeze	Finished	Add notes	varun-ε		33m 39s	-	0.0050258	0.8114	0.0040014	0.0052574	0.8148
road-segmenta...fcn8s-unfreeze	Finished	Add notes	varun-ε	+	33m 39s	-	0.0045665	0.82528	0.0039601	0.0047794	0.8281

Figure 16: compiled results - unfrozen backbone

In addition, we visualise these predictions on 8 images from the test set, one of which is shown here. These images are available in full resolution in the repository.

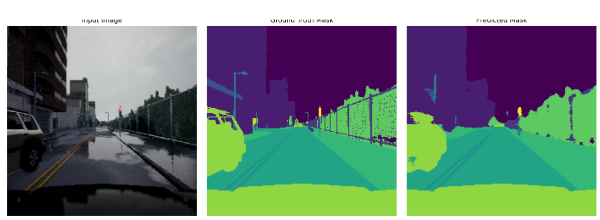


Figure 17: FCN8s - unfrozen backbone

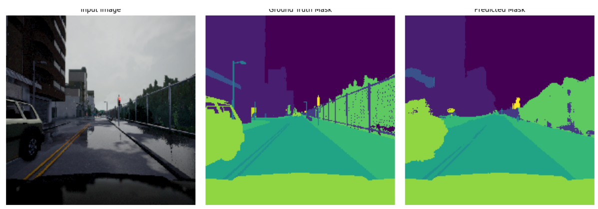


Figure 18: FCN16s - unfrozen backbone

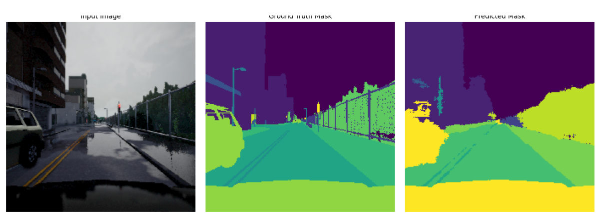


Figure 19: FCN32s - unfrozen backbone

2.3 Comparison

The core difference lies int he granularity of predictions / segmentations.

- **FCN-32s** converts the final fully connected layers into convolutional onces and directly upsamples the output in one step. This is simplest and fastest but produces course segmentations with less detail.
- **FCN-16s** improves upon FCN-32s by adding a skip connection from a shallower intermediate layer. This combination injects finer spatial details, leading to a better segmentation accuracy.
- **FCN-8s** further refines the segmentation by incorporating an additional skip connection from an even shallower layer. The fusion of features produces the finest spatial detail among the variants, leading to the best accuracy.

In addition, the rather obvious difference is in the role of the models. **FCN-8s** upsamples the features from $\frac{1}{8}$ resolution ($8\times$) to recover pixel level resolution, and so on.

We consolidate the results from above into a table to focus on the test mean IoU.

Table 1: Test Mean IoU for FCN Variants (Frozen vs. Unfrozen)

	fcn8s		**fcn16s**		**fcn32s**	
	Frozen	**Unfrozen**	**Frozen**	**Unfrozen**	**Frozen**	**Unfrozen**
test/mIoU	0.76542	0.82528	0.73293	0.8114	0.65325	0.789

We notice that the general order of performance is FCN-8s > FCN-16s > FCN-32s which aligns with our expectations. In addition, unfrozen > frozen in general, likely meaning the extracted VGG features do not sufficiently encode enough information for adequate segmentation, leading to the model learning better ones.

This aligns with the qualitative results as seen in the images in the above document and the repository.