Semantic Segmentation for Self Driving Cars

Eddie Poon

TABLE OF CONTENTS

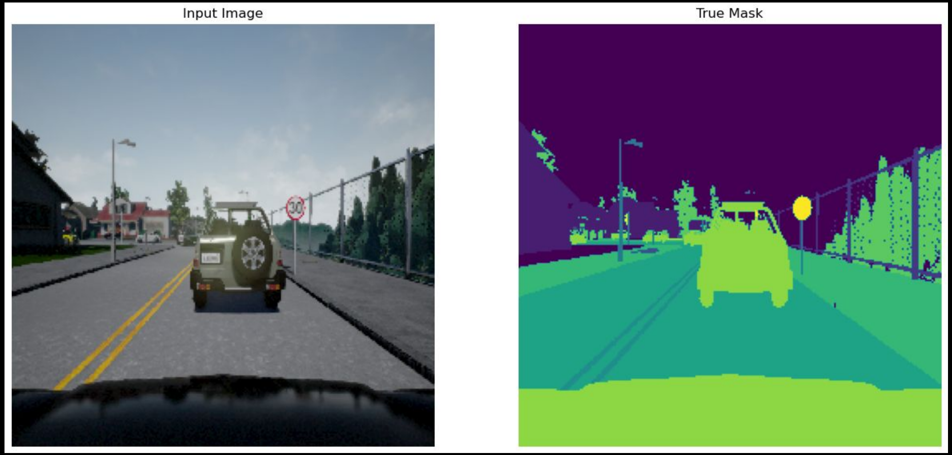
Business Understanding

Autonomous driving technologies have experienced incredible growth as electric vehicles become more mainstream. More robust systems are needed in order to prevent accidents.

Our goal is to create a model that is proficient at classification of roads, vehicles, signage, and other environmental elements.

Data

- ➔ Images come from CARLA simulator
- ➔ 5000 images, 5000 masks
- ➔ Older version with 13 classes

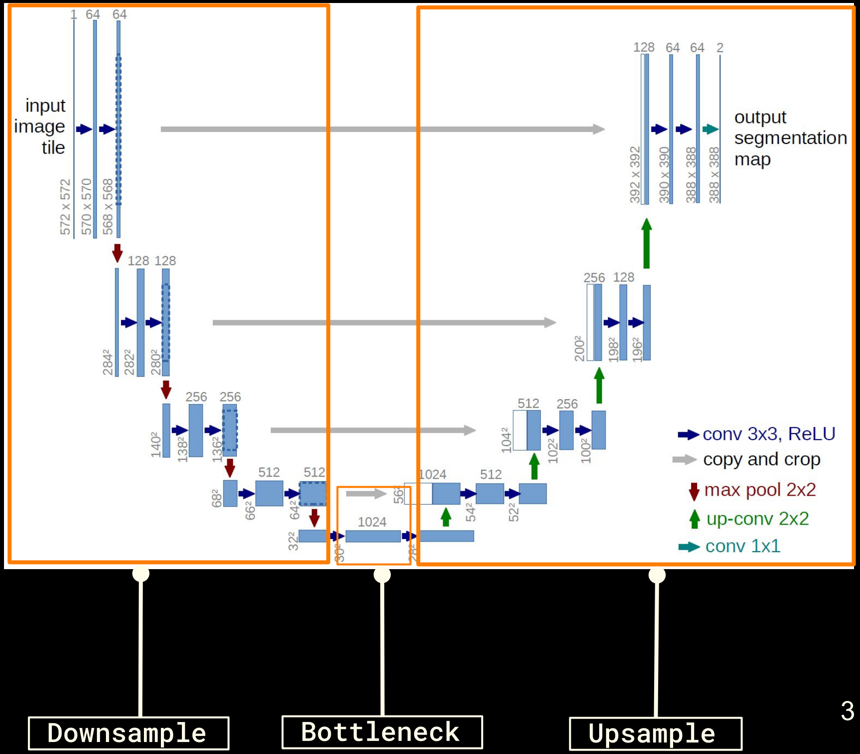


Value	Tag	Converted color
0	Unlabeled	(0, 0, 0)
1	Building	(70, 70, 70)
2	Fence	(190, 153, 153)
3	Other	(250, 170, 160)
4	Pedestrian	(220, 20, 60)
5	Pole	(153, 153, 153)
6	Road line	(157, 234, 50)
7	Road	(128, 64, 128)
8	Sidewalk	(244, 35, 232)
9	Vegetation	(107, 142, 35)
10	Car	(0, 0, 142)
11	Wall	(102, 102, 156)
12	Traffic sign	(220, 220, 0)

U-Net Architecture

- Special type of CNN that has a U shape
- Downsample extracts features from input, decreases resolution of image
- Bottleneck stops downsampling, passes info to upsample blocks
- Upsample increases resolution of image, uses skip connections to prevent info loss during downsample

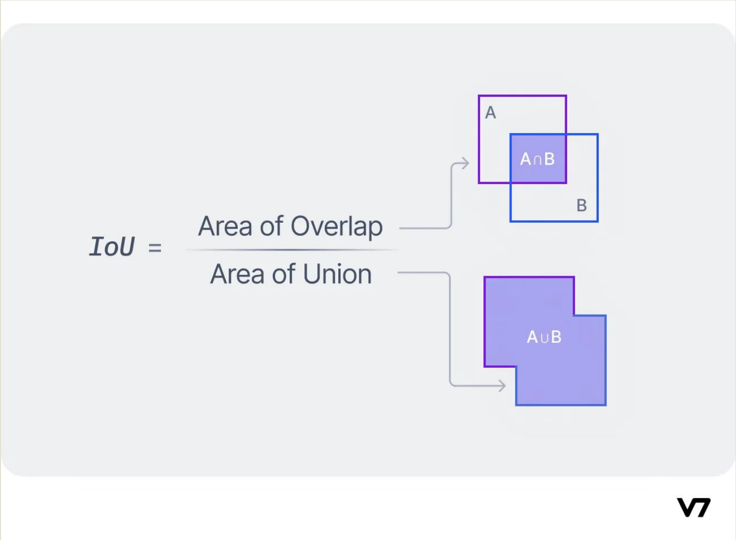
[1]

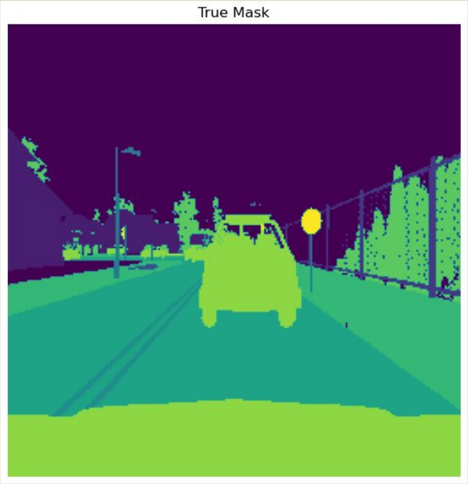


Metrics

- Accuracy is not a great measure for semantic segmentation
- Mean Intersection over Union is much better

[2]



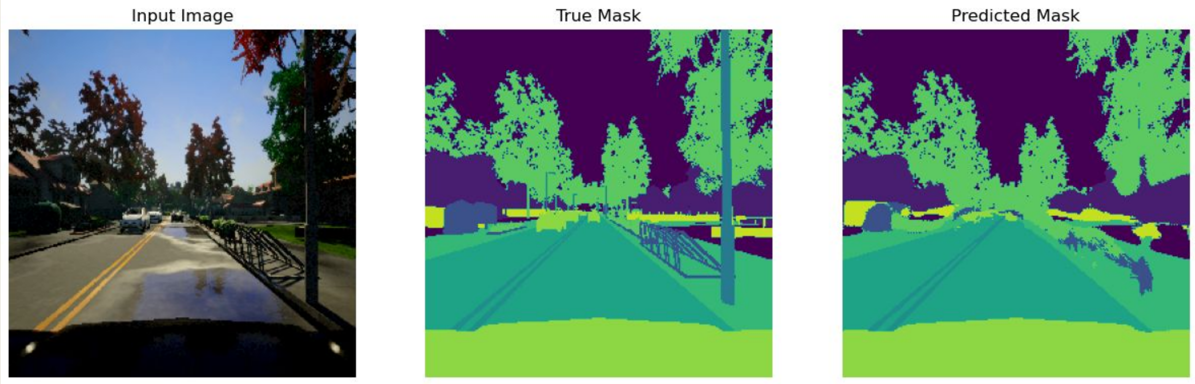


Baseline

- Decent prediction for a beginning model

- Simple architecture, lacked features for more robust learning

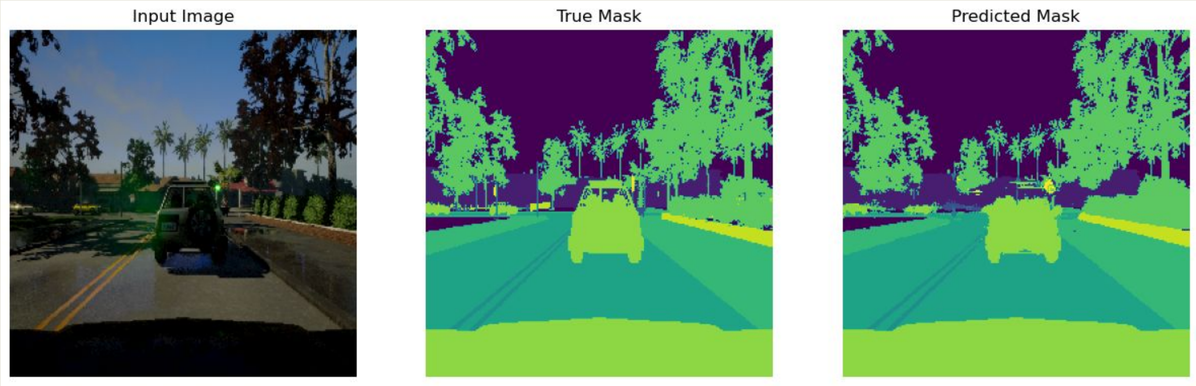
- Struggles on smaller objects



Conv2DTranspose

- Considerably better segmentation of smaller objects, but still struggles

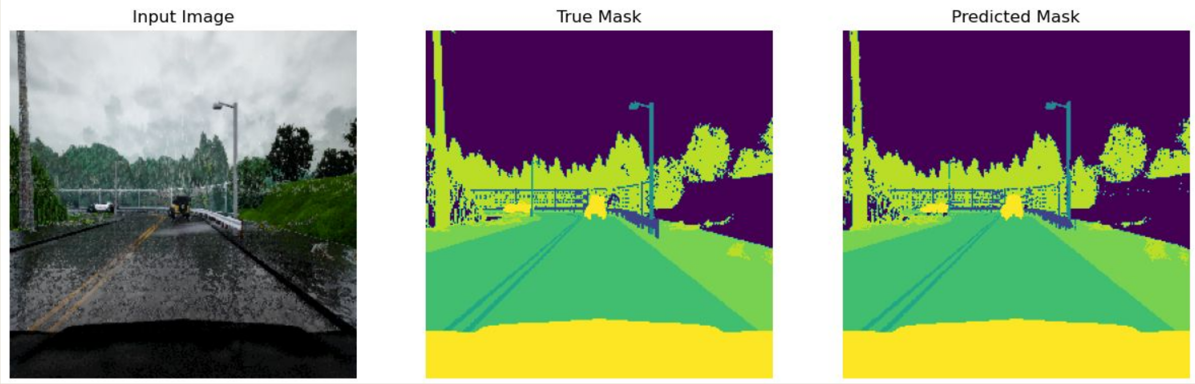
- Added regularization, activation functions to improve learning



Upsampling2D

- Similar output compared to second model, but fares better on smaller objects

- Very similar architecture as previous model, replaced with simpler deconvolution



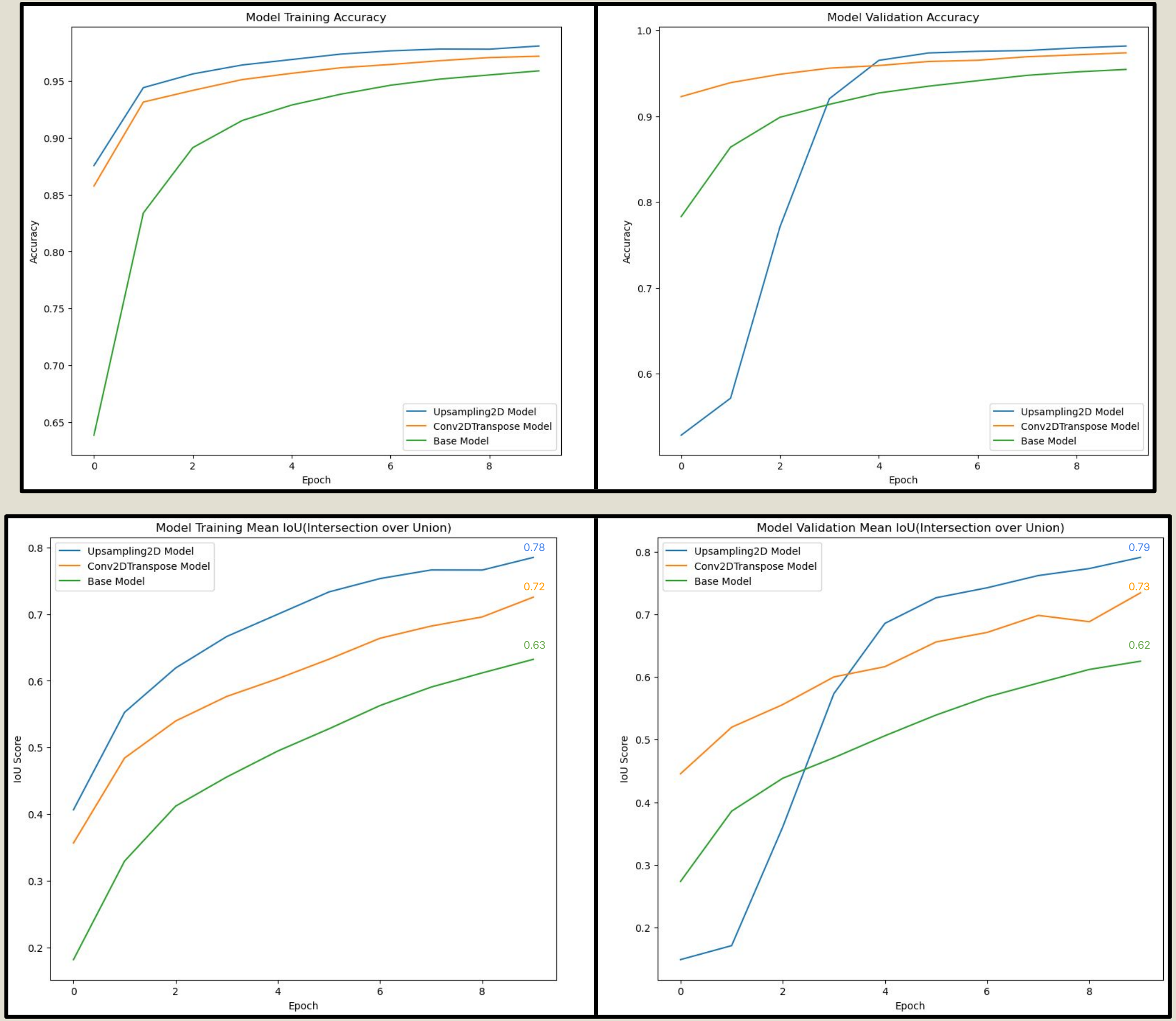
Performance

- According to accuracy, all models performing very well

- Using a better metric, baseline struggles to keep up with other models

- Upsampling2D is best performing model

- Only ran for 10 epochs, could do better with more time

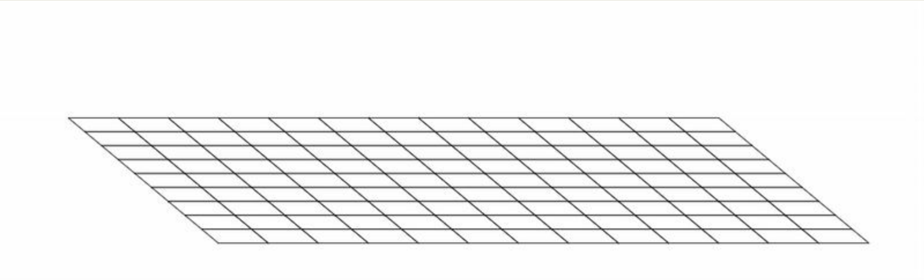


Conv2DTranspose

- Despite being more complex than Upsampling2D, performed worse

- Frequently creates artifacts from filters overlapping unevenly across output

[3]

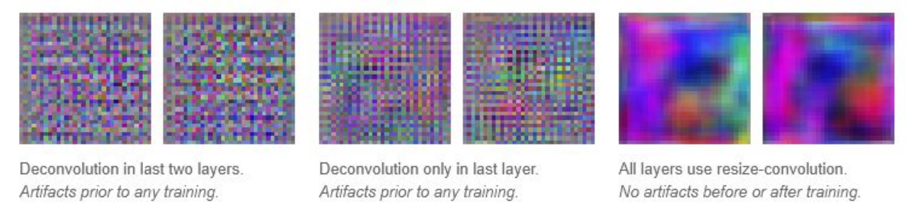


Upsampling2D+Conv2D

- Uses simpler interpolation methods like nearest neighbor

- Separates upsampling and feature transformation steps

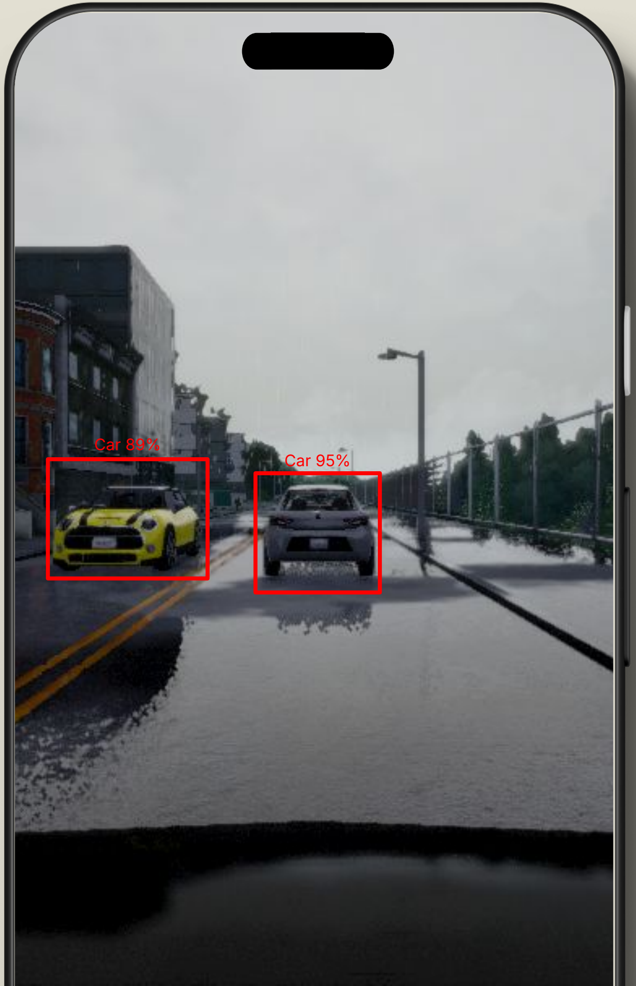
[3]



Next Steps

- Evolve model to work on video segmentation

- Create lightweight app to allow user to submit images, viewing output



References

1. Ronneberger, Olaf, Fischer, Philipp, & Brox, Thomas. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*.
2. Shah, D. (2023, May 30). *Intersection over union (IOU): Definition, calculation, code*. V7. https://www.v7labs.com/blog/intersection-over-union-guide
3. Odena, A., Dumoulin, V., & Olah, C. (2016, October 17). Deconvolution and checkerboard artifacts. Distill. https://distill.pub/2016/deconv-checkerboard/
4. S, P. (2024, November 20). *A comprehensive guide to UNET architecture*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2023/08/unet-architecture-mastering-image-segmentation/