Semantic Segmentation for a Self Driving Vehicle

Keith Chester
Bob DeMont
Zeke Flaton

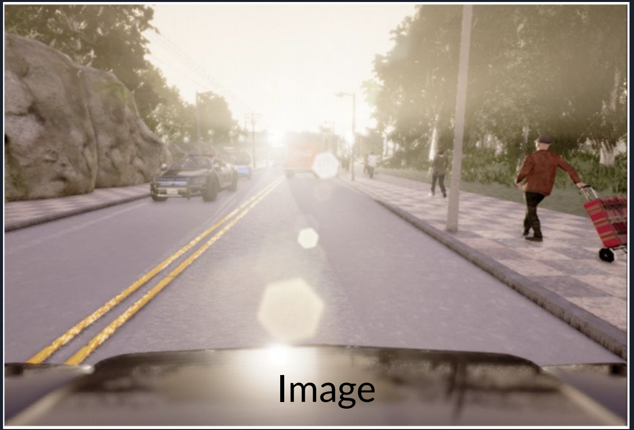
Outline

- What is semantic segmentation
- How is semantic segmentation achieved
- Datasets
- Data augmentation
- Model
- Results

What is Semantic Segmentation

- Classifying each pixel to a specific class, e.g. car, road, pedestrian, traffic light
  - Not a specific instance (a car is a car- if they overlap, they're not separate)
- As distinguished from Instance segmentation
  - where every instance of a class is identified\
- Or Panoptic segmentation
  - Which is a combination of the two

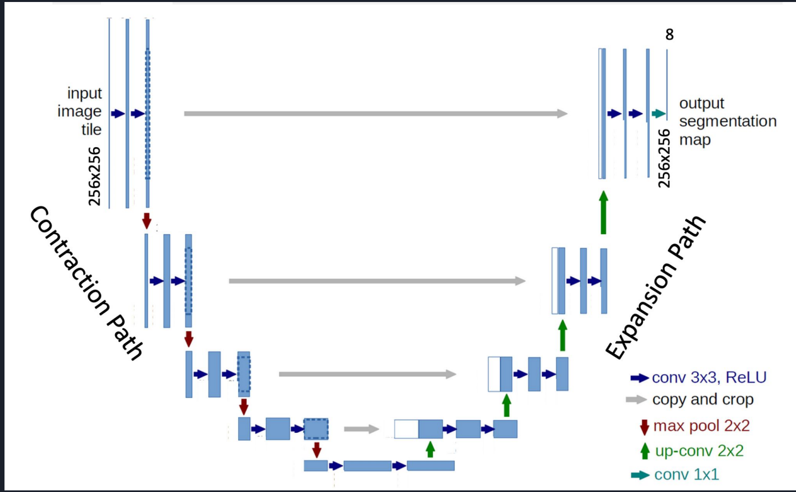
- Segmentation achieved via DNN training of images vs ground truth labelling



DNN Model: U-net

Model Architecture

- Based model on U-Net architecture
- Originally developed for medical Imaging
- Particularly adept at segmentation results
- Network **downsamples** input with ReLU convolutional layers, then **upsamples** with up-convolutions



Methods-Dataset Generation

- Datasets generated by automating **CARLA**, which gives us both RGB input and labeled images for a ground truth
- We created **tooling for programmatic control** of CARLA's maps, weather, # of vehicles, # of pedestrians, and road conditions
- Generated **15GB** of **14k+** image / label pairs

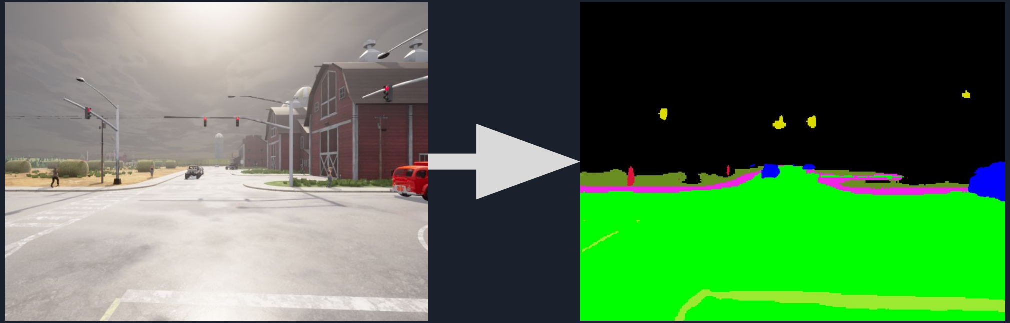
Methods - Data Augmentation

Dataset is loaded via a custom data loader that applies **data augmentation** to the training set, including:

- Horizontal mirroring
- Zoomed-in subsections of the image being focused
- Image blurring
- Noise
- Brightness
- Contrast

14,000 images to more than 2.4million options

Methods - Training



Model architecture and training built in Tensorflow with Keras

Batch sizes were selected to maximize memory

- 32 batch size for *256x256*
- 8 batch size for *512x512*
- Both used ~9GB of GPU RAM on an *NVIDIA RTX 3080* GPU

Methods - Training Challenges

To **combat significant categorical imbalances** in both presentation and pixel representation multiple techniques were utilized:

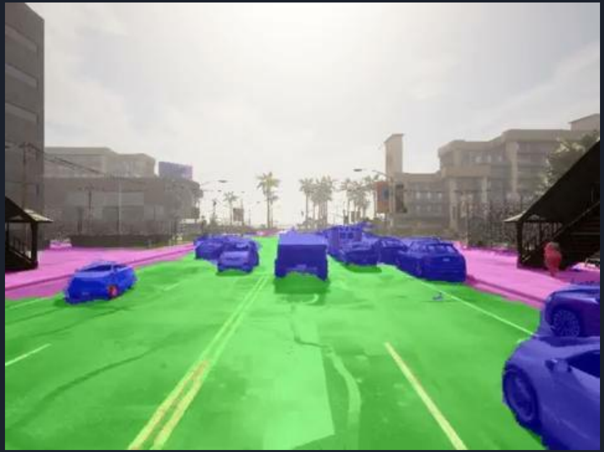
- We **increased resolution** of input images
- Swapped to **sparse categorical cross-entropy focal loss**
- **Calculated loss weights**, which effectively modify learning rates for specific categories
- **Sliding-window inference** per frame

Results

Videos/images were processed by the model and outputs overlaid on the original video/image with 40% transparency

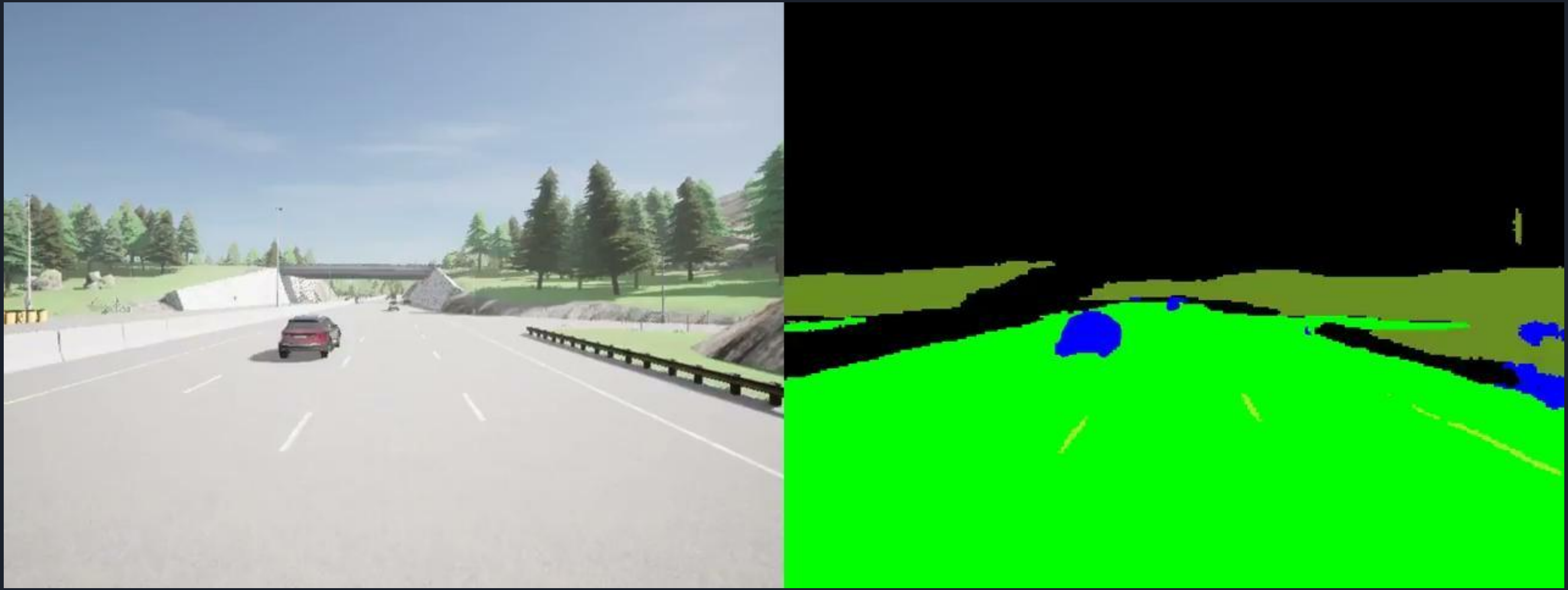
Label colors:

- Blue - Car
- Green - Road
- Light Green - Road lines
- Purple - Sidewalk
- Red - Pedestrian
- Yellow - Traffic lights/signs



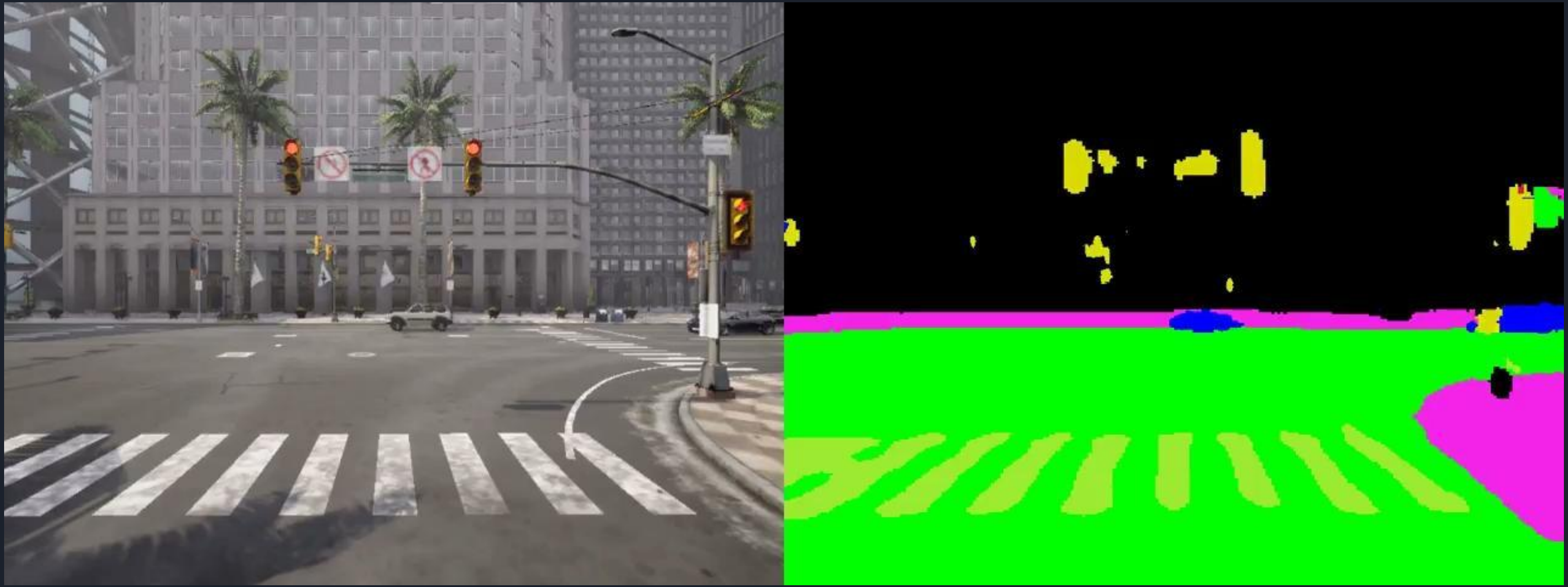
Results continued

Side-by Side Simulated image and Trained model inferred segmentation



Results continued

Side-by Side Simulated image and Trained model inferred segmentation



References

[1] Liu, X., Deng, Z. & Yang, Y. Recent progress in semantic image segmentation. Artif Intell Rev 52, 1089–1106 (2019). https://doi.org/10.1007/s10462-018-9641-3

[2] S. Malec, 'Semantic Segmentation with Carla Simulator', Dissertation, 2021.

[3] D. R. Niranjan, B. C. VinayKarthik and Mohana, "Deep Learning based Object Detection Model for Autonomous Driving Research using CARLA Simulator," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), 2021, pp. 1251-1258, doi: 10.1109/ICOSEC51865.2021.9591747.

[4] C. Wang, '2D object detection and semantic segmentation in the Carla simulator', Dissertation, 2020.