Semantic Segmentation for Autonomous Driving Applications

Aristos Athens, Alice Li, Amelia (Qingyun) Bian

Introduction

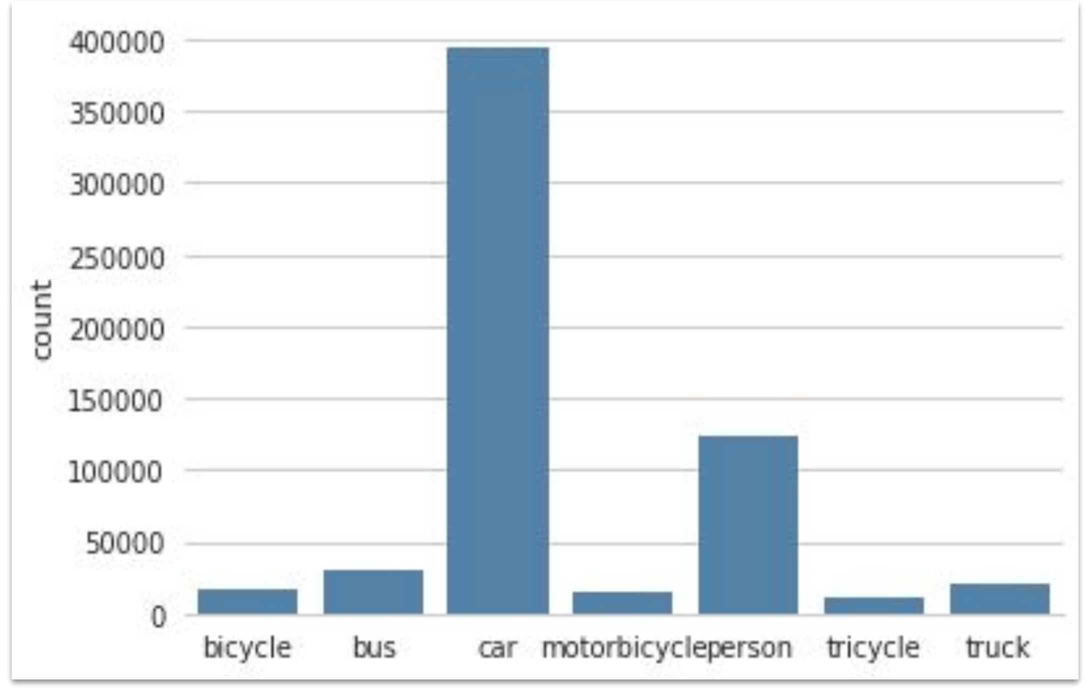
Goal: Perform segmentation to identify moving objects for autonomous driving applications.

Data

Dashcam video frames from CVPR Workshop for Autonomous Driving Kaggle challenge. Pictures were 2710 x 3384, but were cropped to 384 x 384.

Original Image

Ground Truth Label



Class Distribution: 35 classes, background omitted

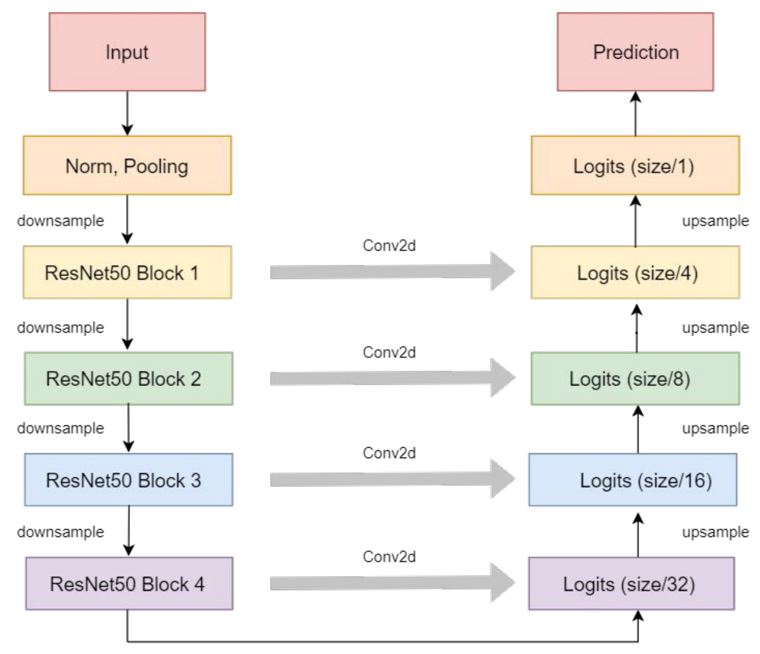
Class imbalance

- The dataset did not label sky, ground, or road pixels, leading to a large class imbalance between objects and background (~1:10).
- Models tended to converge on all-background predictions.
- To counter this, we devised two ways of weighting down the background pixels such that the model would be able to learn patterns for non-background pixels.

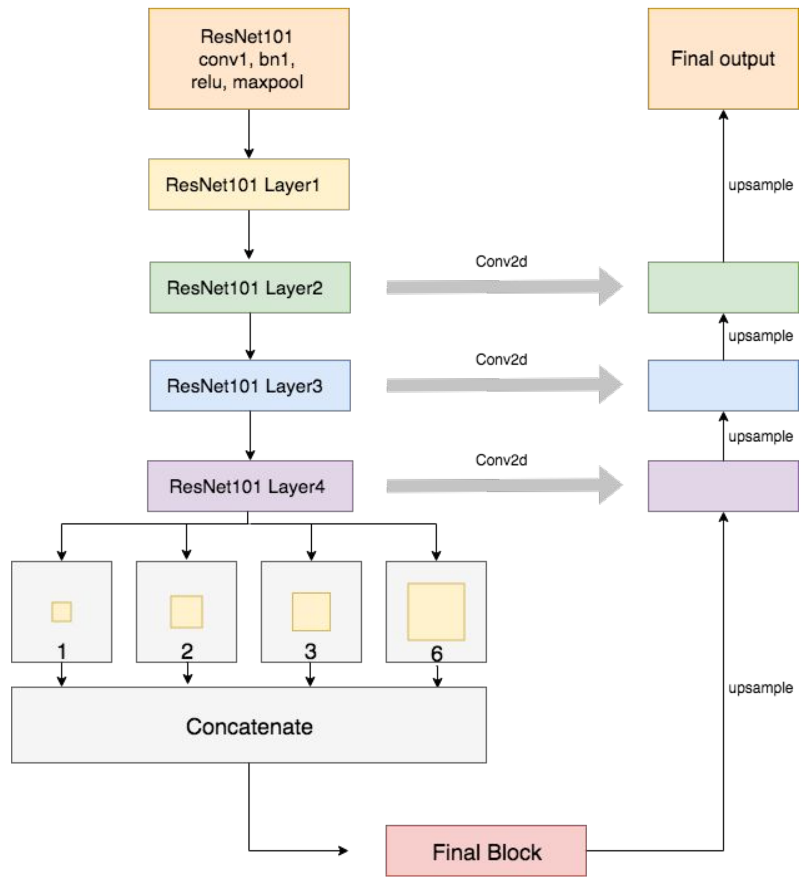
Methods

Model Architecture

Our best performing models were two architectures using ResNet50 and Spatial Pyramid Pooling (SPP) for downsampling and UNet for upsampling.



ResNet50-UNet



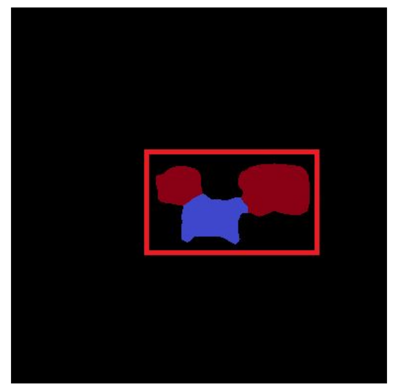
SPP-UNet

Loss Function

- Weighted each class *c* in cross entropy loss function by proportion of pixels in class and hyperparameter α.

$$loss(x,c) = \alpha * \frac{N_{c\ pixels}}{N_{total\ pixels}}\left(-x[c] + log\left(\sum e^{x[j]}\right)\right)$$

- Created bounding box around all non-background pixels in each image. Downweighted loss for every pixel *x* outside of bounding box by reducing α. This method produced sharper images in less time.



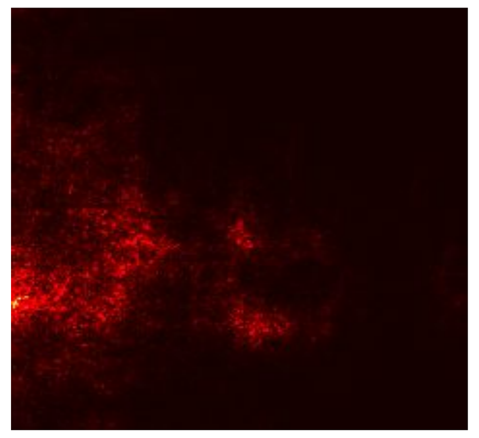
Label with bounding box

Saliency Map

Constructed Saliency Maps to visualize the effectiveness of our targeted loss strategy.

Original image

Ground truth label



Saliency map

Results

Accuracy

Pixel-wise accuracy for test data.

Model	Test Accuracy
SPP-UNet (trained on 200 batches of 20)	97.9%
ResNet50-Unet (trained on 500 batches of 50)	98.0%

Sample Results

Top row: SPP-UNet
Bottom row: ResNet50-UNet

Original image **Prediction** **Ground truth label**

Conclusion

- ResNet50-UNet with the bounding box loss performed best with a test set accuracy of 98%.
- Bounding box loss weighting improved our performance by concentrating only on relevant pixels and classes.
- SPP-UNet architecture has more parameters, but may exceed the ResNet50-UNet performance with more training time/data.