Capstone 3 Final Presentation

Semantic Segmentation of Aerial Cityscapes

Table of Contents

Problem identification (1 of 2)

Joby Aviation and Archer Aviation have teamed up to develop aerial vehicles capable of transporting either passengers or packages.

They must demonstrate to the Federal Aviation Administration (FAA) additional safety measures in regards to obstacle avoidance, in the event that the pilot becomes incapacitated.

Problem identification (2 of 2)

The goal of this project is to create a computer vision model that can be deployed onboard the vehicle with the ability to distinguish between eight different classes.

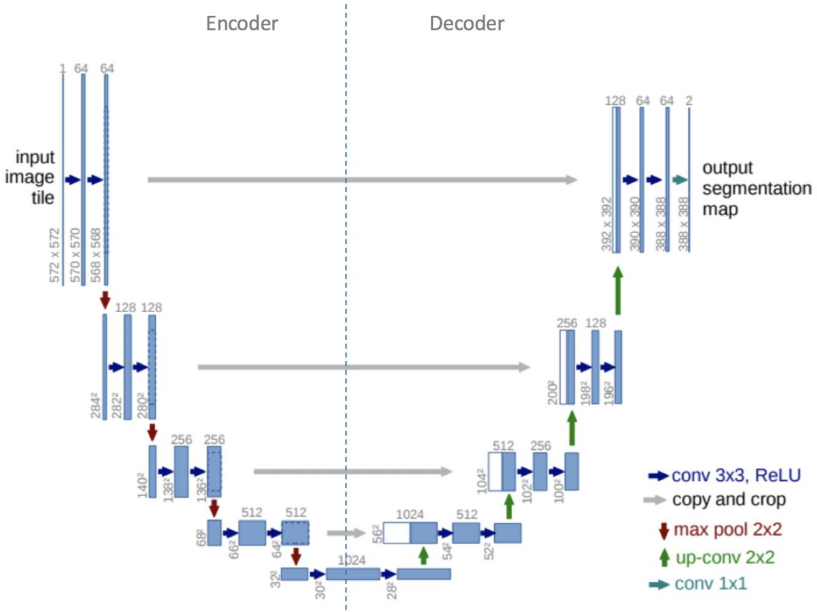
The setting that the model must be familiar with are cityscapes and urban landscapes from an aerial perspective.

The data science approach (1 of 2)

Computer vision is one use case of deep learning neural network models. In particular convolutional neural networks.

U-Net is a model architecture that became famous for its ability to outperform other models on biomedical image segmentation.

A modified version of this architecture can be used to segment images of cityscapes taken from aerial perspectives.



The data science approach (2 of 2)

Two datasets were chosen for training such a model:

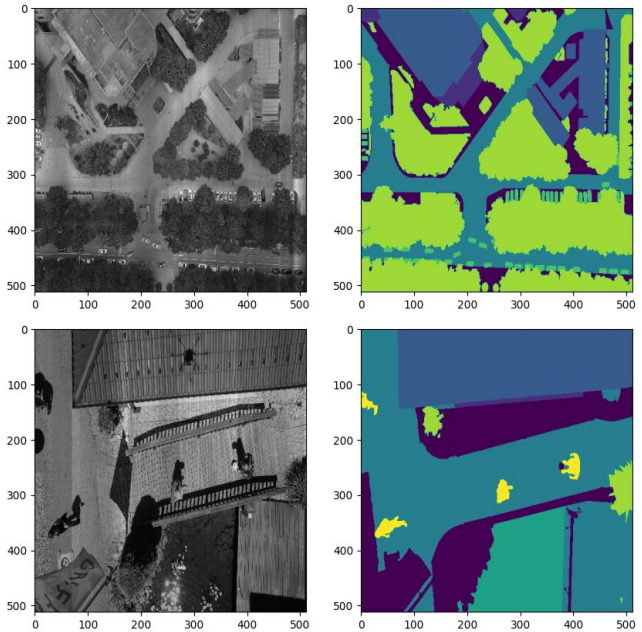
1. Varied Drone Dataset for Semantic Segmentation (VDD)
2. Semantic Drone Dataset (SDD)

Each dataset contains original images and corresponding masks where each pixel is labeled as a particular class.

Data wrangling and key takeaways (1 of 2)

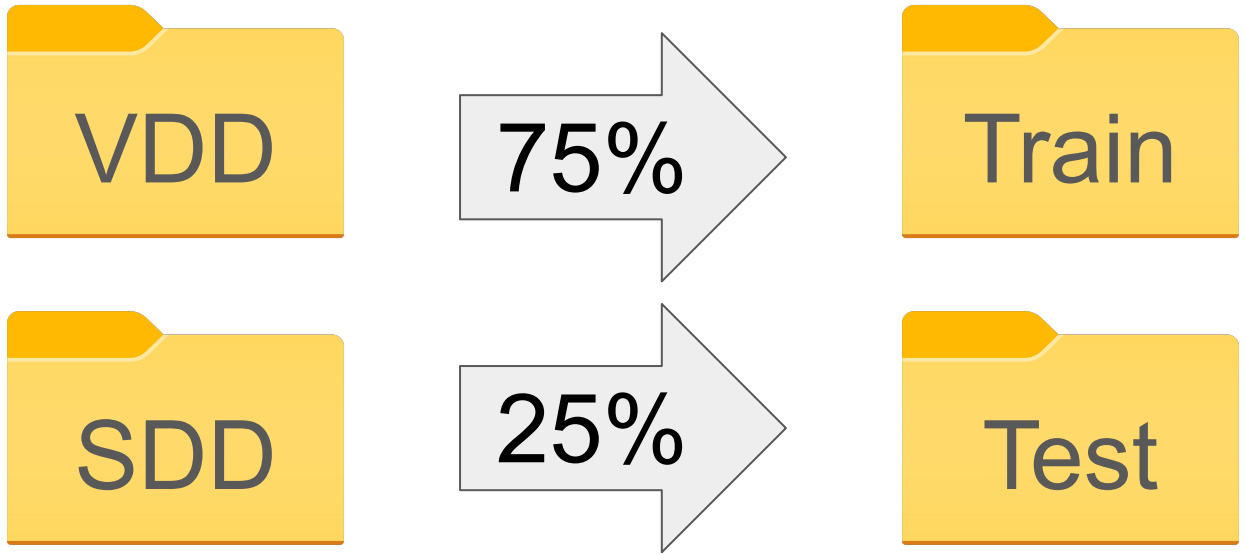
Combining two datasets requires remasking them each into common classes.

Master Class ID (Class)
0 (unlabeled)
1 (wall)
2 (roof)
3 (road)
4 (water)
5 (car)
6 (vegetation)
7 (person)



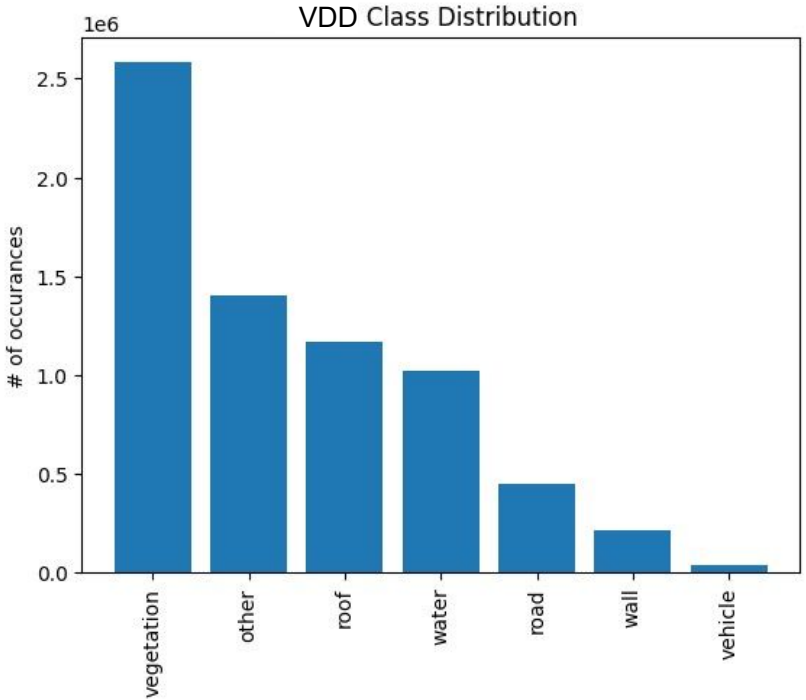
Data wrangling and key takeaways (2 of 2)

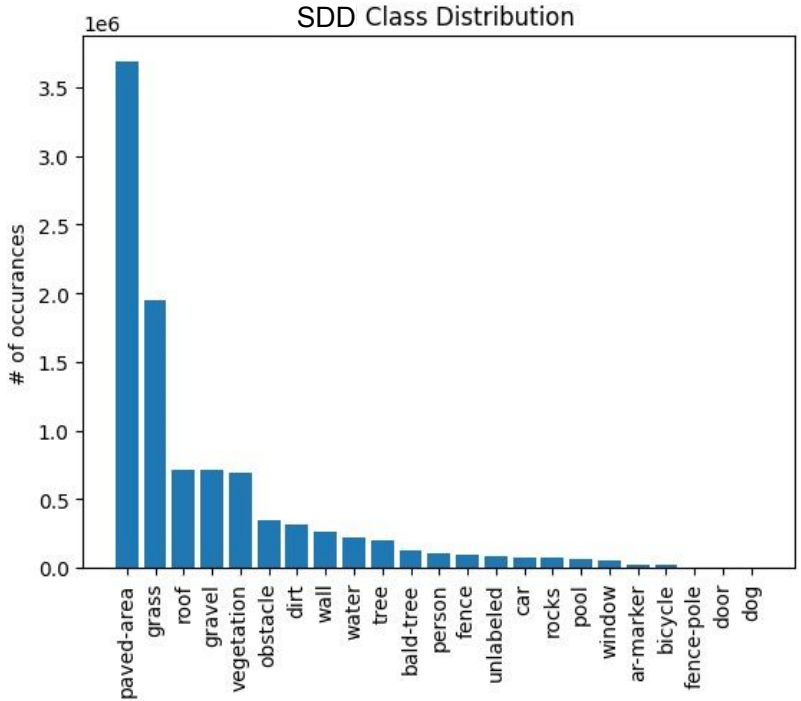
Equal ratios of each dataset were dedicated to the final train and test datasets.



Exploratory analysis and key findings (1 of 2)

Both VDD and SDD datasets experience class imbalance. Class weights required.





Exploratory analysis and key findings (2 of 2)

The VDD and SDD datasets have different sized images with different number of channels.

This is something to be aware of, but all images will be resized upon training.

	VDD		SDD	
	Images	Masks	Images	Masks
# of Channels	3	1	3	3
Resolution (height by width)	4000x3000	4000x3000	6000x4000	6000x4000

Modeling (1 of 2)

Zero padding used to preserve image size through convolutions.

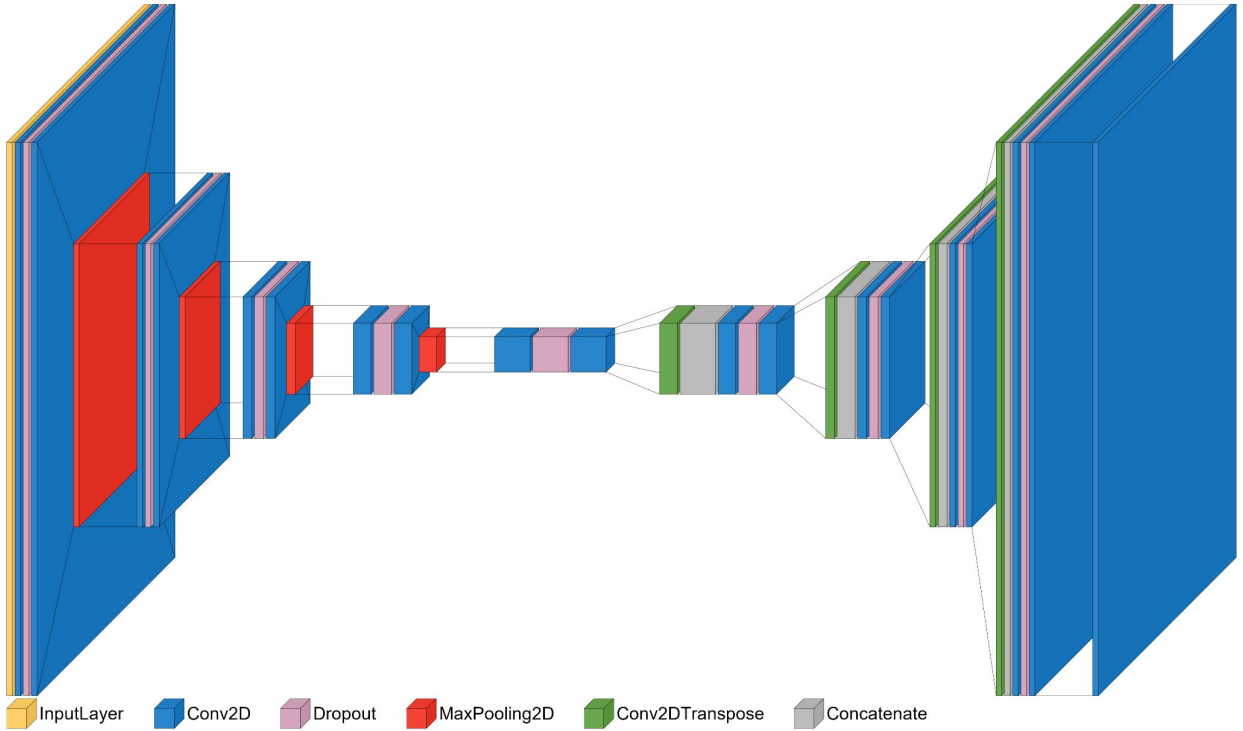
Dropout layers introduced.

Input size 512x512.

Adam Optimizer.

Categorical Cross Entropy Loss.

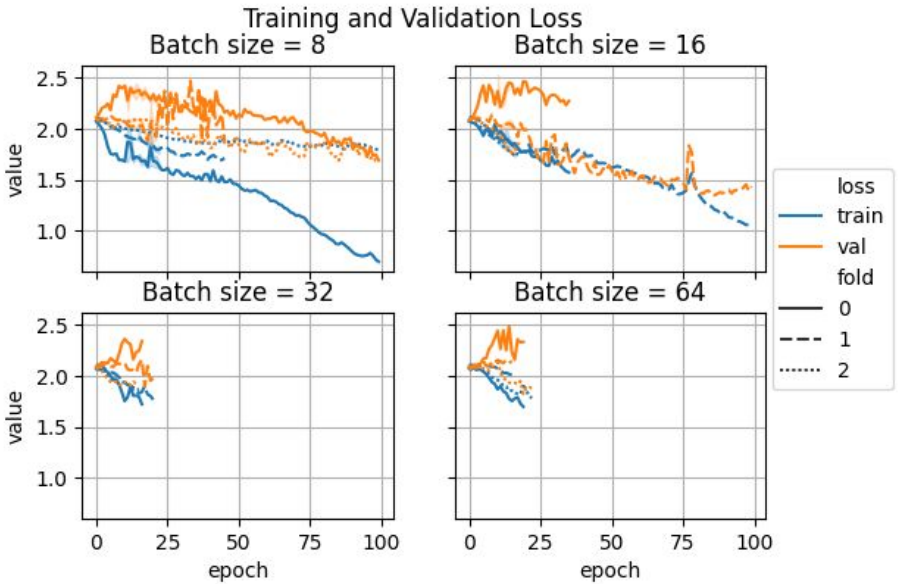
1,940,936 trainable parameters.



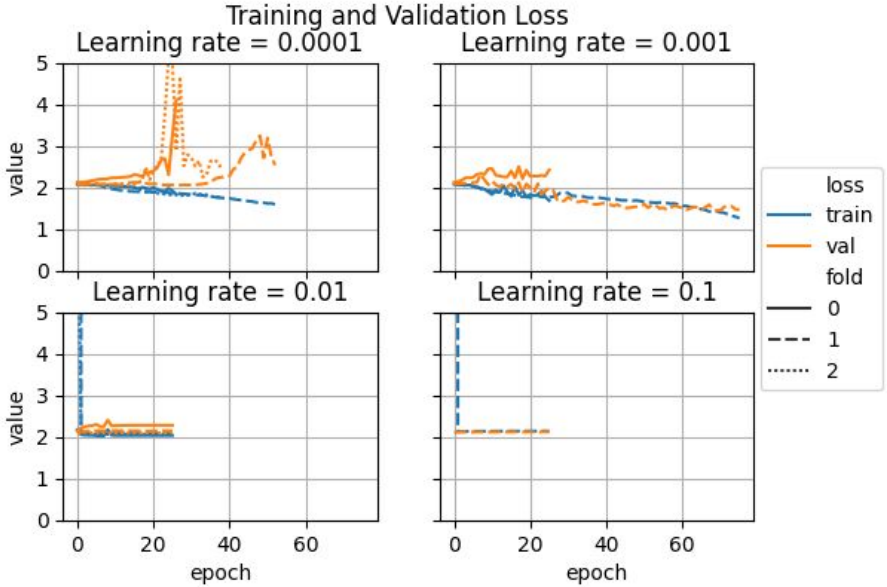
Modeling (2 of 2)

Two hyperparameter studies performed: batch size and learning rate.

Best batch size = 16



Best learning rate = 0.001



Analysis results (1 of 2)

Intersection over union scores are the metric for analyzing each class.

$$IoU_i = \frac{True\ Positives_i}{\left(True\ Positives_i + False\ Positives_i + False\ Negatives_i\right)}$$

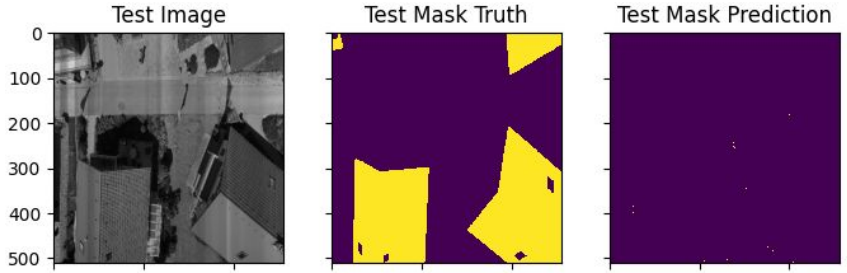
Note that the model being analyzed was trained to 50 epochs and has not yet converged.

Class	IoU Score (%)
0 (unlabeled)	16.64
1 (wall)	10.25
2 (roof)	0.13
3 (road)	31.48
4 (water)	47.88
5 (car)	4.65
6 (vegetation)	32.06
7 (person)	2.80

Analysis results (2 of 2)

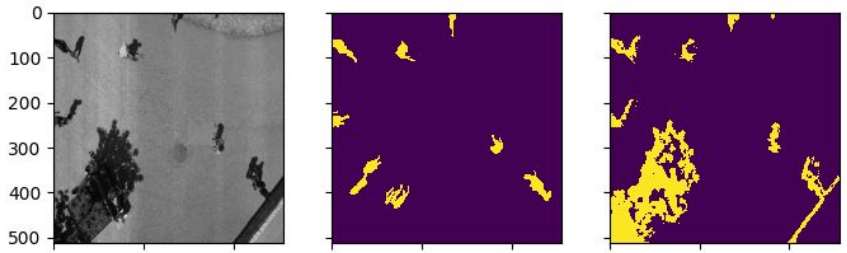
Roofs often mislabeled as roads, as in this example.

Roof



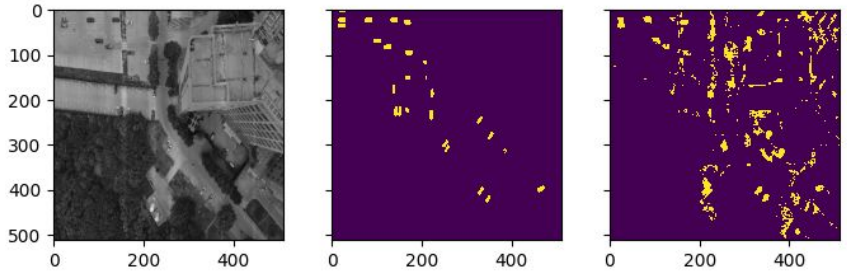
Persons labeled correctly, but object such as trees and shadows also labeled as persons.

Person



Less-intuitive features in the images are being labeled as cars.

Car

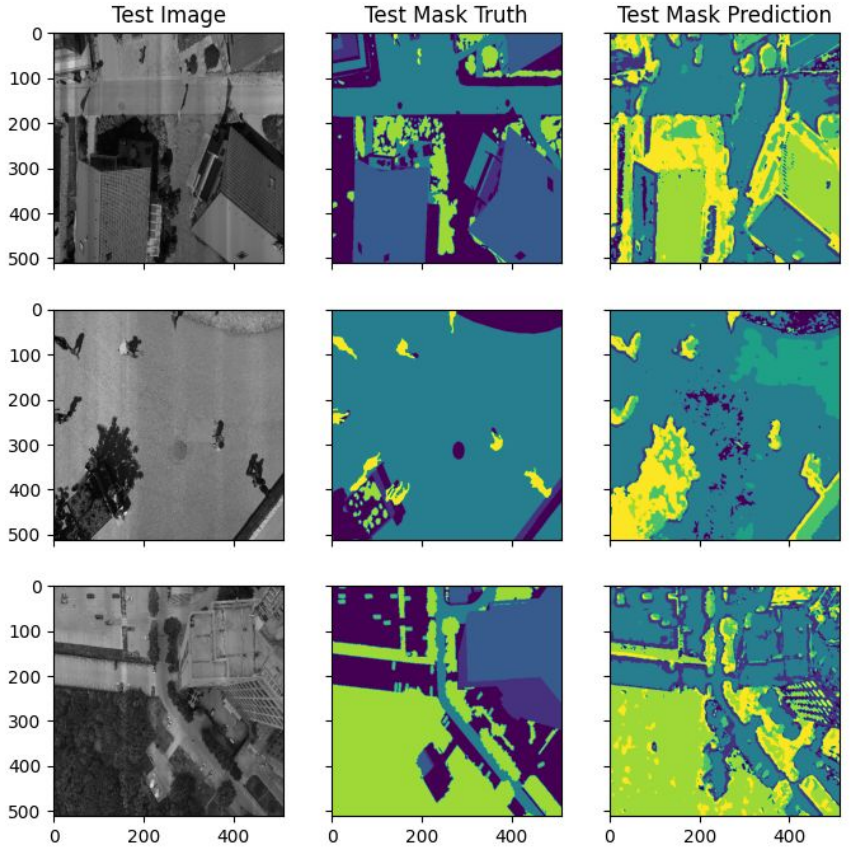


Recommendations to client

Until the model can reliably predict classes common in the cityscape environment, one should not be relied upon for safety measures.

Model shows ability to classify clusters of pixels that belong to particular classes, even before higher accuracies were achieved.

Recommend reducing classes to those critical for operation and start with achieving a high accuracy model before expanding.



Suggestions for improvement/ Future work

Recommend training the model on individual datasets first before combining datasets to establish baseline performance. It would provide insight on how mixing datasets affect class accuracies.

Implement a MobileNetV2 model combined with DeepLabV3, as described by the MobileNetV2 paper, for high performing semantic segmentation on edge devices such as mobile or embedded applications that would be implemented on the aerial vehicle.