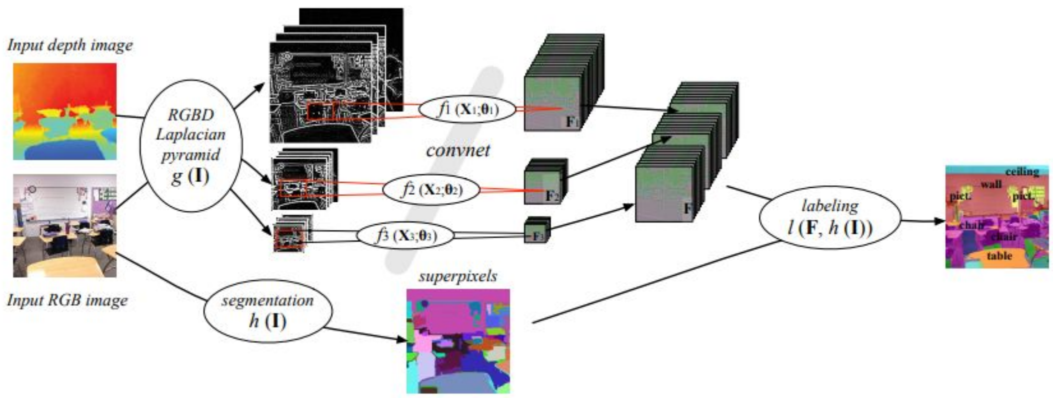
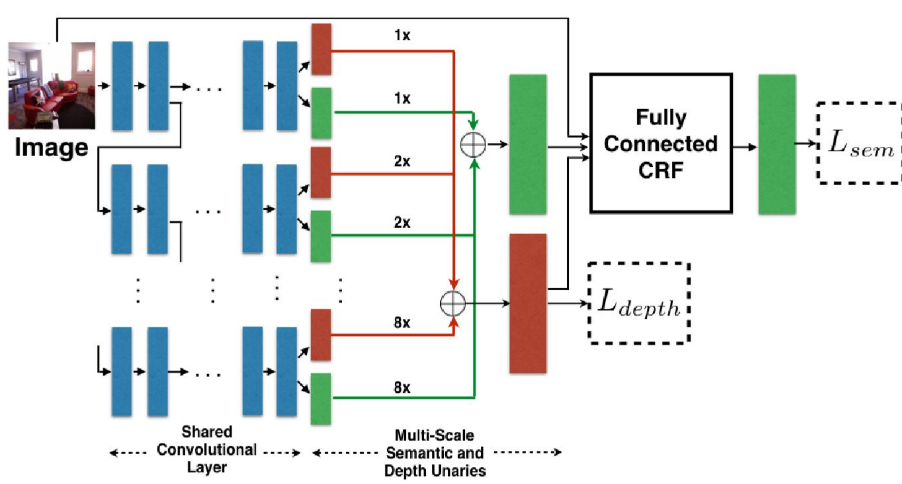
Semantic Segmentation - Depth Relationship

Nowadays, a lot of scientific effort goes into estimating the semantic segmentation or the depth using as input a RGB image, but not many think about the relationship between these two representations: depth and semantic segmentation.

In 2013 Camille Couprie, Clement Farabet, Laurent Najman and Yann LeCun presented a paper entitled “Indoor Semantic Segmentation using depth information” which proposes the use of depth alongside RGB in order to help the network estimate both semantic segmentation and labeling.



After that, in 2016 Arsalan Mousavian, Hamed Pirsiavash and Jana Kosecka proved in their paper “Joint Semantic Segmentation and Depth Estimation with DeepConvolutional Networks” that using the estimated depth and semantic segmentation from a RGB image improves the final semantic segmentation estimation.



This project aims to analyze the relationship between depth and semantic segmentation, firstly by trying to estimate semantic segmentation using a depth map as input and vice versa (experiments

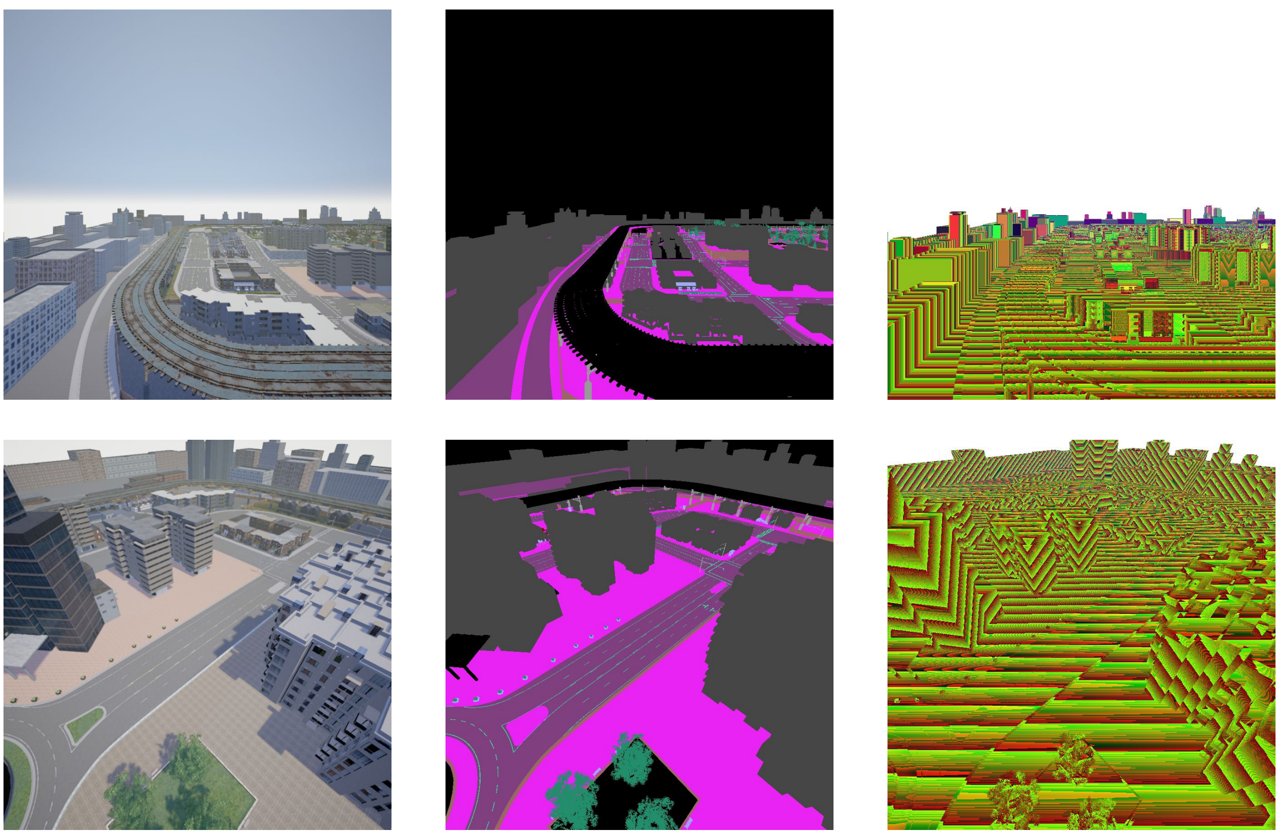
1 and 2) and secondly by trying to improve the semantic segmentation using the estimated depth map from a RGB input and vice versa (experiments 3 and 4).

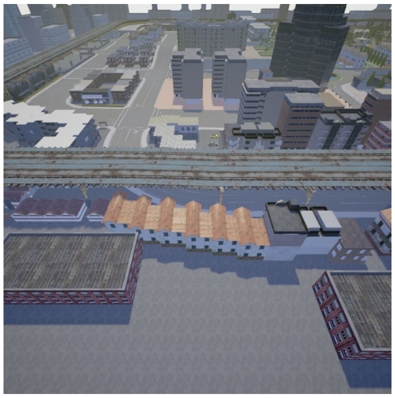
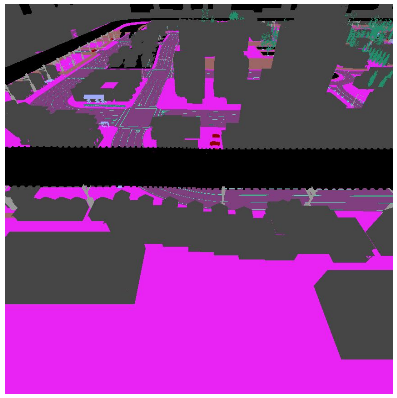
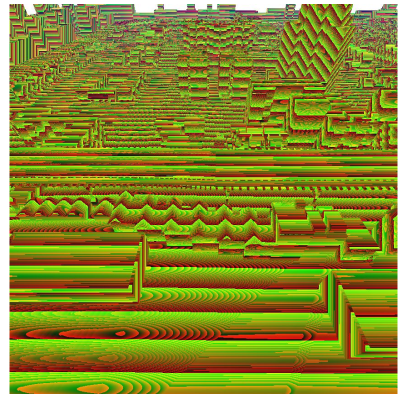
The dataset

The dataset used contains 12360 tuples of RGB, Depth and Semantic Segmentation extracted from a simulator on a city map taken from an airplane. For faster computation, the dataset was reduced to a smaller dataset taking frames with a step of 5 (frame_*****0 and frame_*****5). This reduced dataset was split into training: all frames < 10000 and testing: all frames > 10000.

Below we can see some tuples:

RGB Semantic Segmentation Depth



The framework

For training I used a framework developed in PyTorch which supports representations as nodes (RGB, SS, Depth) linked together by encoder - decoder models. The technical details of the training are:

- Optimizer: Adam
- Learning rate: 0.001
- Betas: (0.9, 0.999)
- Eps: 1e-08
- EarlyStopping on loss with patience 5 epochs

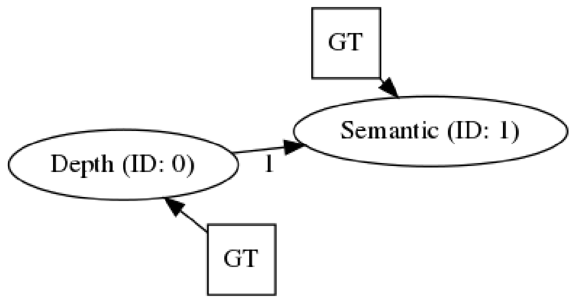
Experiments 1 and 2

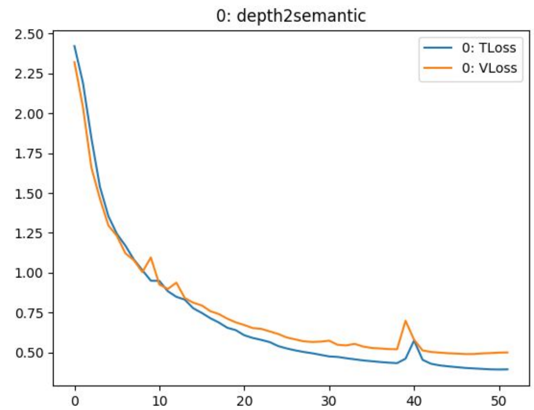
As stated in the introduction, the project started with two experiments:

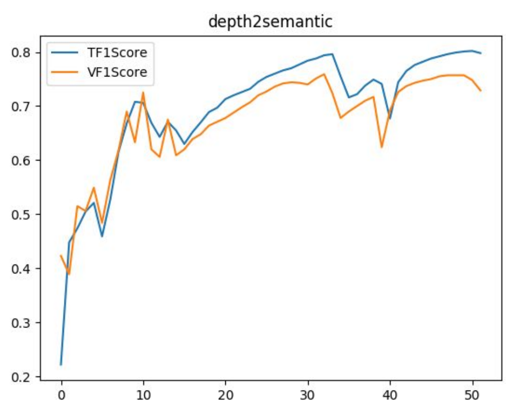
1. Depth -> Semantic
2. Semantic -> Depth

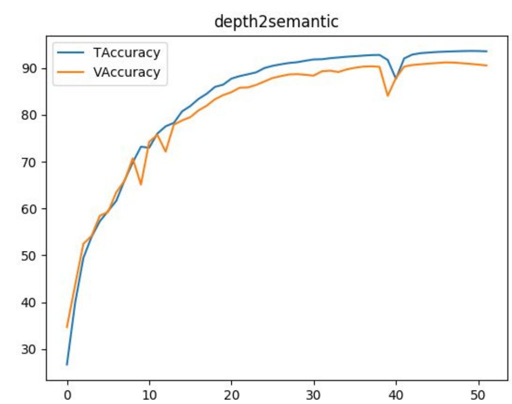
Experiment 1

To do the first experiment, the graph needs to be configured as in the image below. We can see both of the nodes having ground truth and the directed linked form Depth to Semantic. Charts for loss, F1Score and accuracy are also presented below.

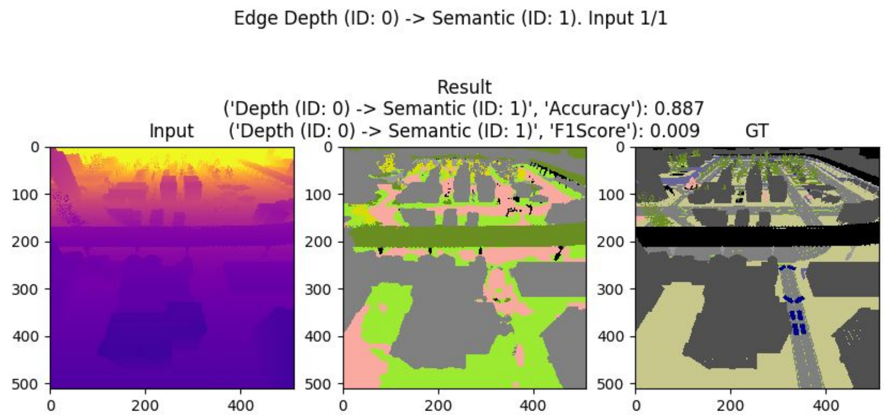






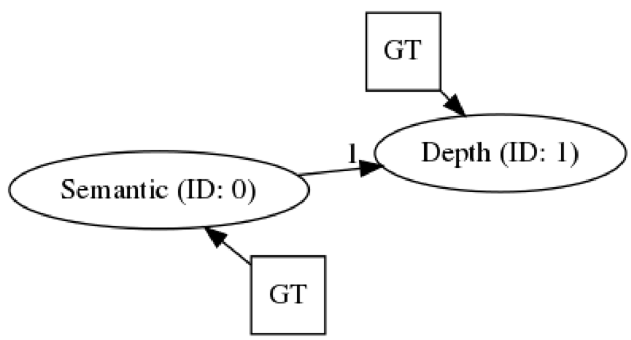


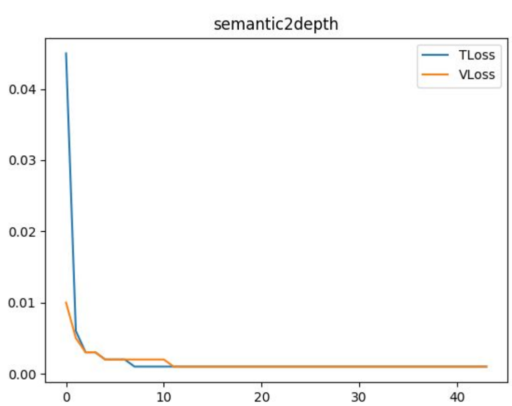
The result on the testset was 91.190 accuracy with 0.755 F1Score. In the image below we can see the input, the result and the ground truth for one frame. (The colors are not the same but they represent the same classes)

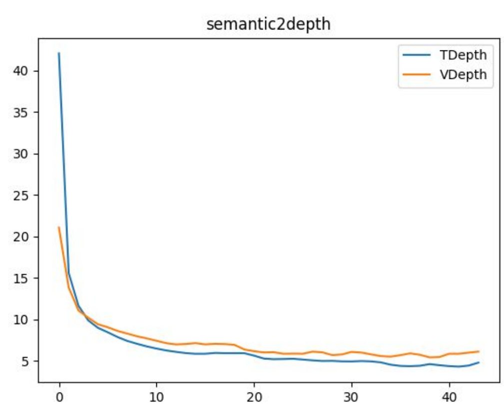


Experiment 2

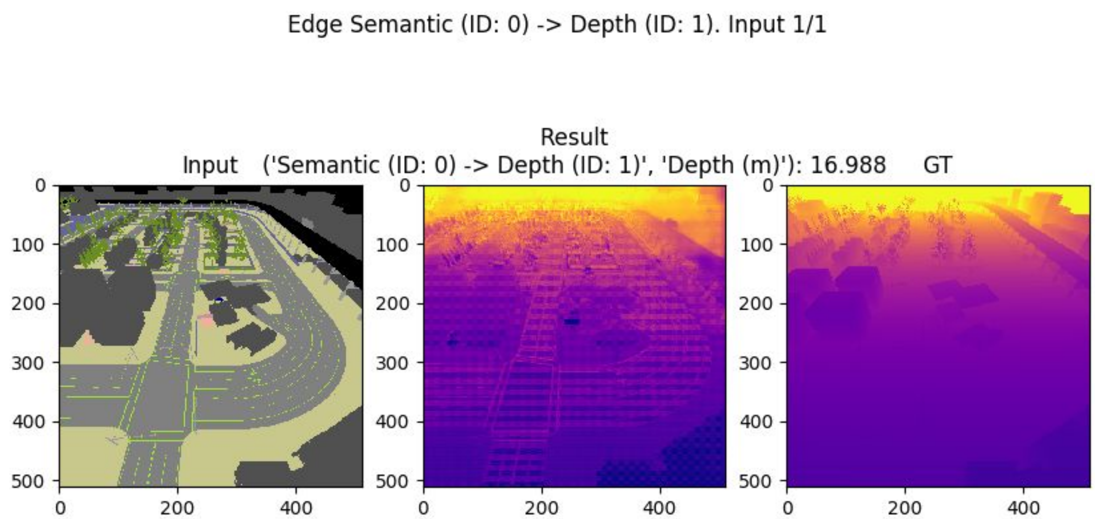
To do the second experiment, the graph needs to be configured as in the image below. We can see both of the nodes having ground truth and the directed linked form Semantic to Depth. Charts for loss and mean depth error are also presented below.







The result on the testset was 5.435 (m) mean depth error. In the image below we can see the input, the result and the ground truth for one frame.



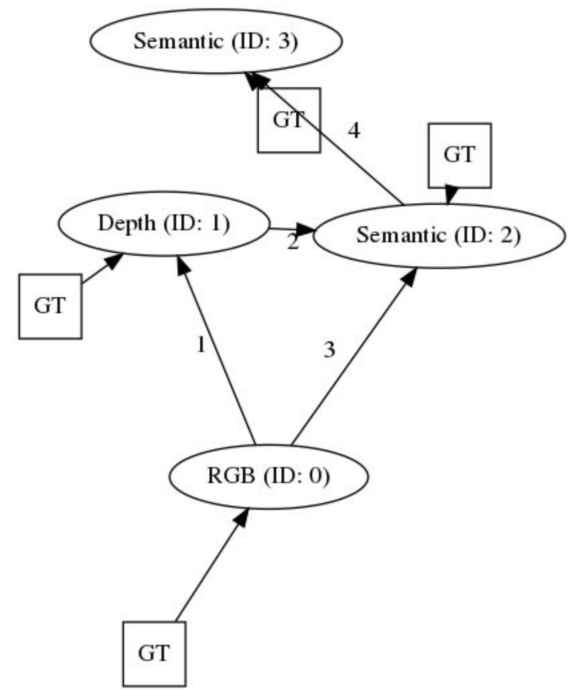
The conclusion after these two experiments is that there is a strong link between depth and semantic segmentation, but more information is given by depth for a semantic segmentation.

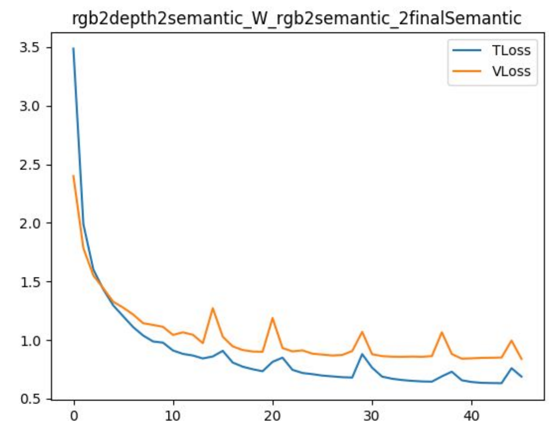
Experiments 3 and 4

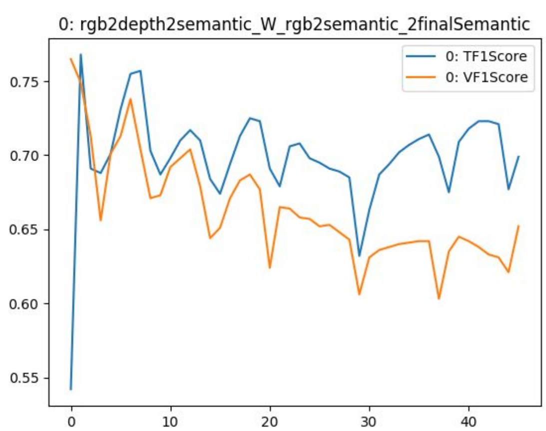
The second series of experiments were limited due to the structure of the framework. The third desired graph would use rgb as input and estimate a depth map which would be concatenated to the rgb input and then passed forward to estimate a semantic segmentation. Because the framework does not support yet a concatenation node the graph used a rgb input to estimate a depth map which would be used to estimate a segmentation map. Another branch used rgb input to estimate a segmentation map, then both the outputs would be used to produce the final segmentation map. For the fourth graph the same improvisation was used.

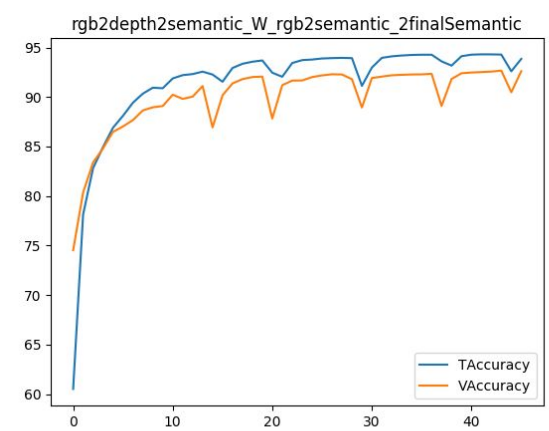
Experiment 3

The graph for experiment three can be seen below, alongside charts for loss, F1Score and accuracy.



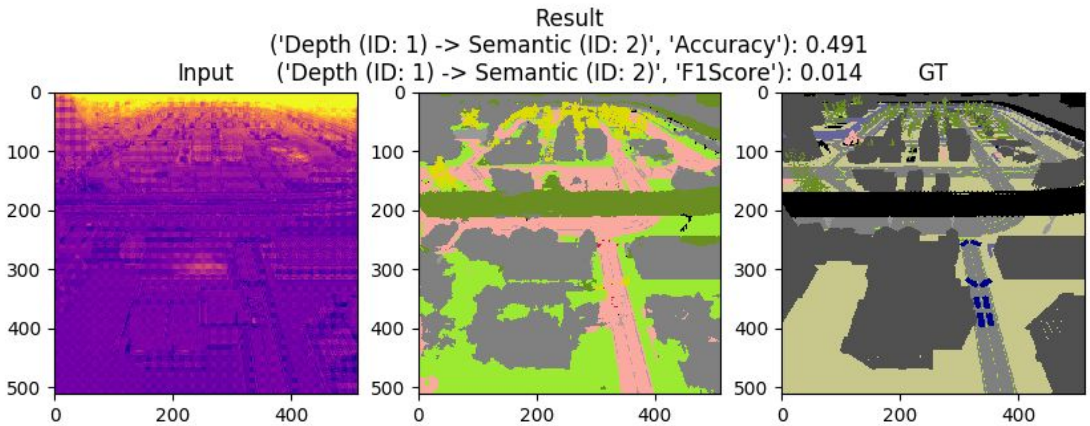




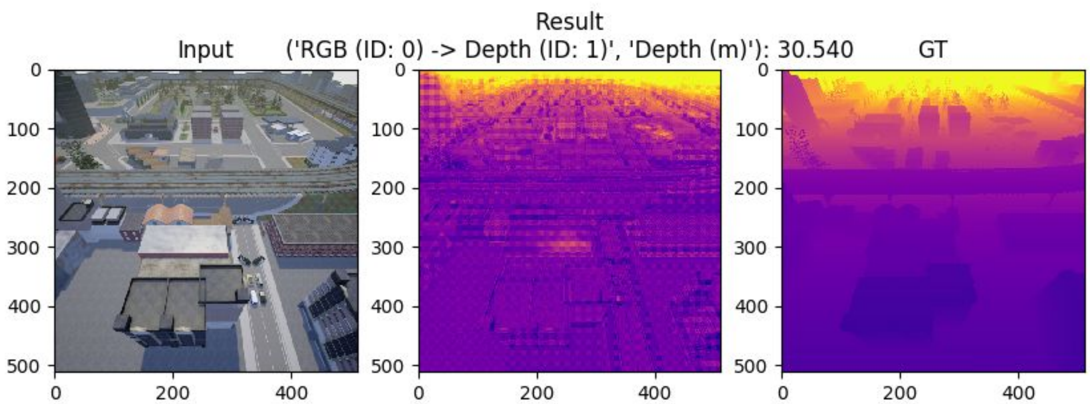


The result on the testset was 92.476 accuracy with 0.642 F1Score. In the images below we can see the inputs, the results and the ground truths for one frame.

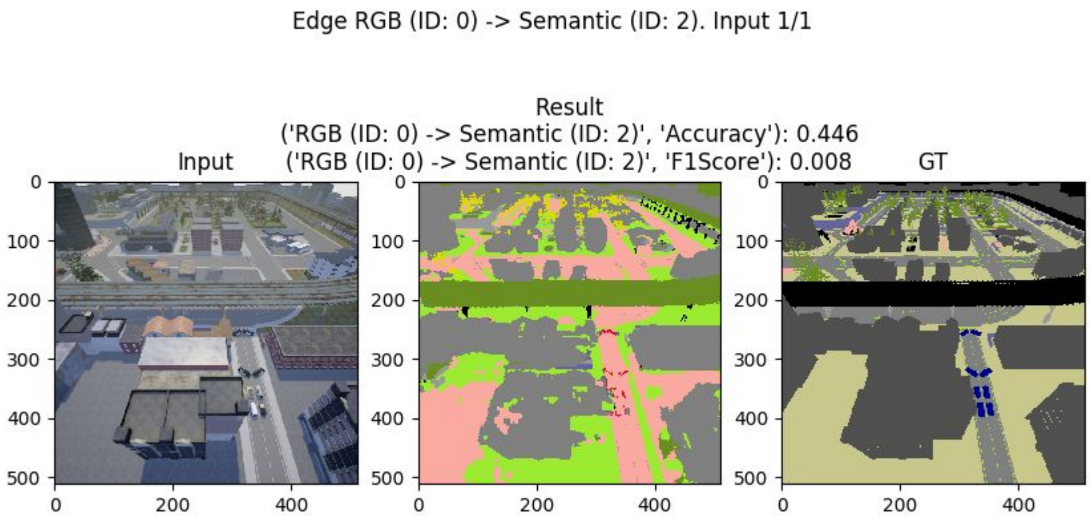
Edge Depth (ID: 1) -> Semantic (ID: 2). Input 1/1

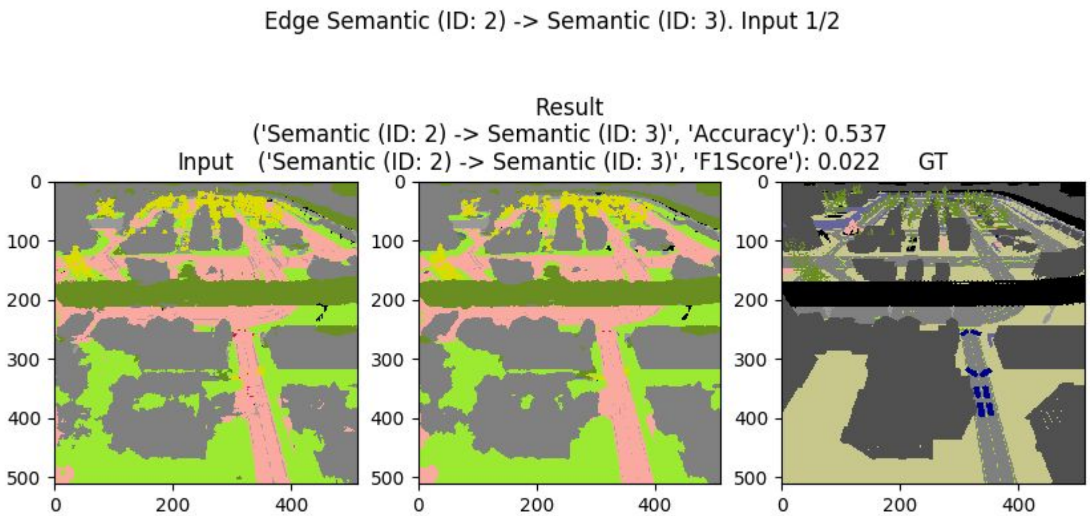


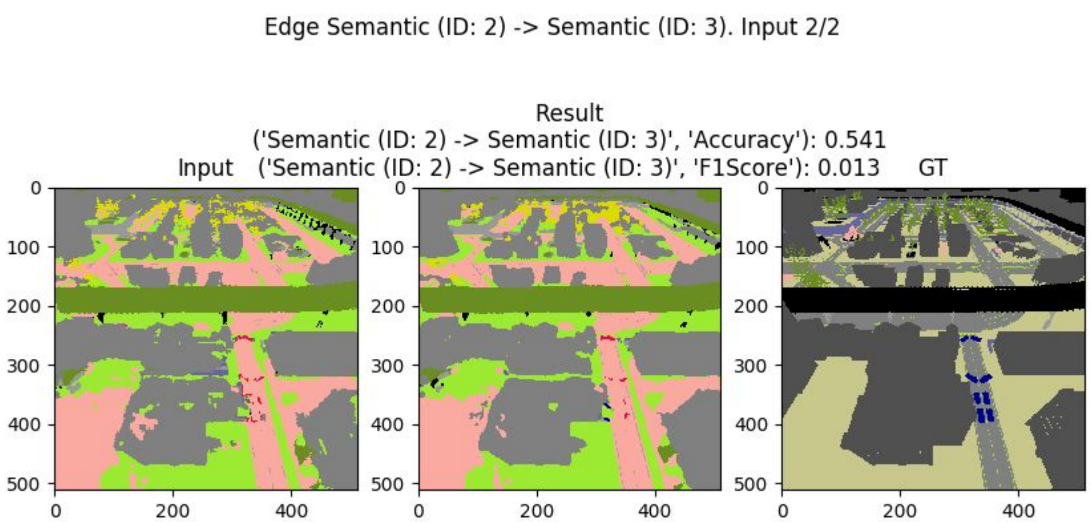
Edge RGB (ID: 0) -> Depth (ID: 1). Input 1/1



hjvhjb

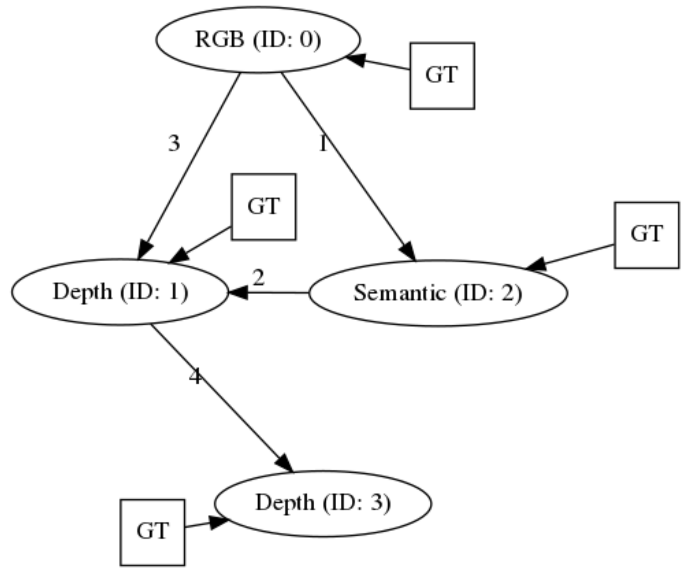

Edge RGB (ID: 0) -> Semantic (ID: 2). Input 1/1
Input
Result
('RGB (ID: 0) -> Semantic (ID: 2)', 'Accuracy'): 0.446
('RGB (ID: 0) -> Semantic (ID: 2)', 'F1Score'): 0.008
GT

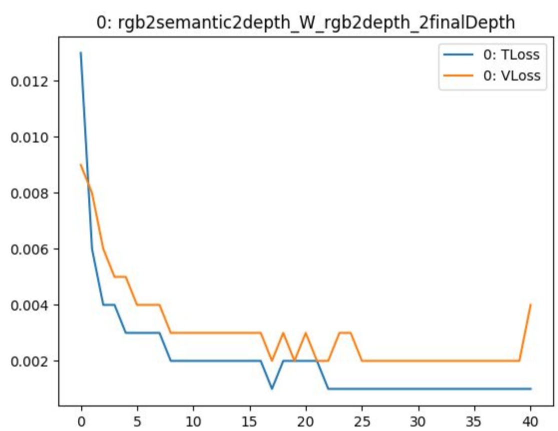


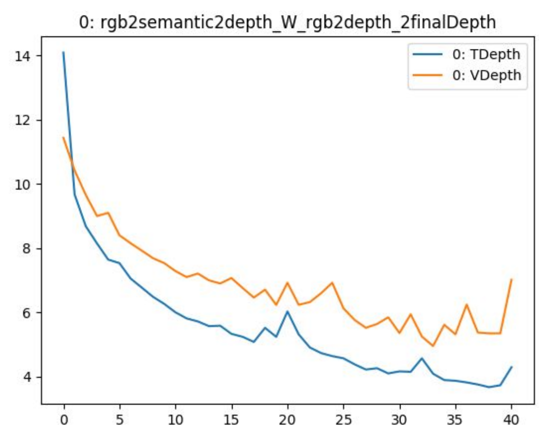


Experiment 4

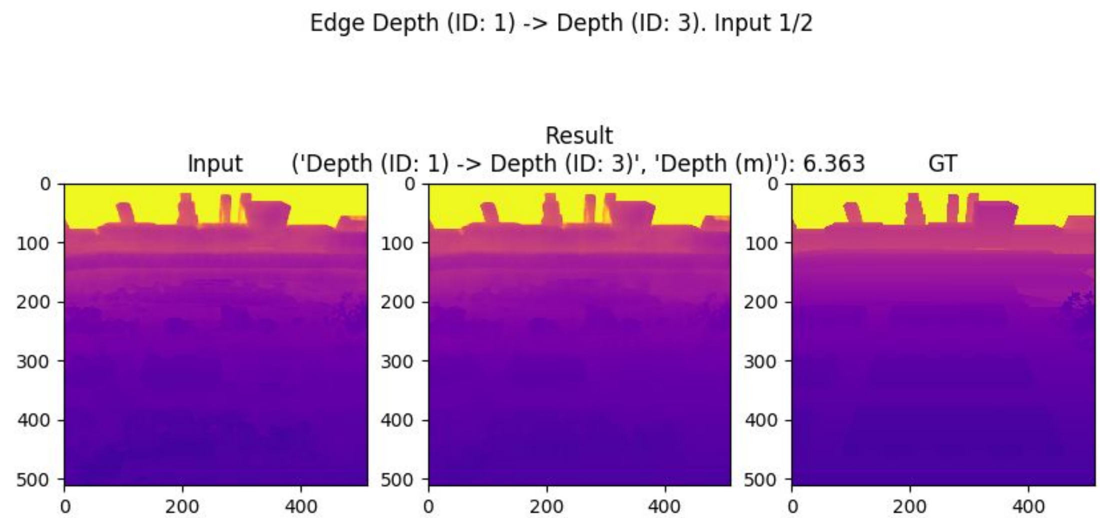
The fourth experiment tries to improve depth estimation using semantic segmentation from rgb. The graph for experiment three can be seen below, alongside charts for loss and mean depth error.

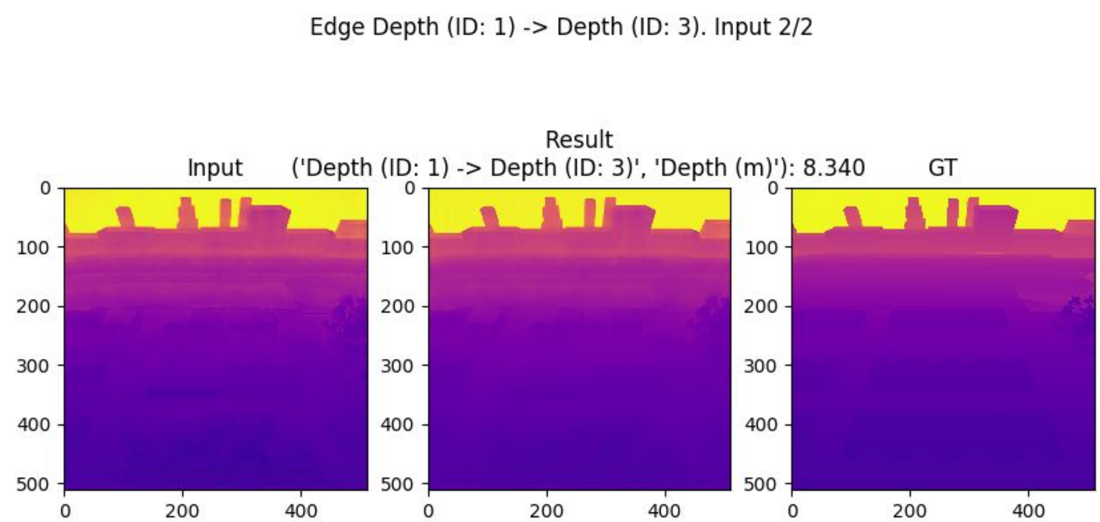


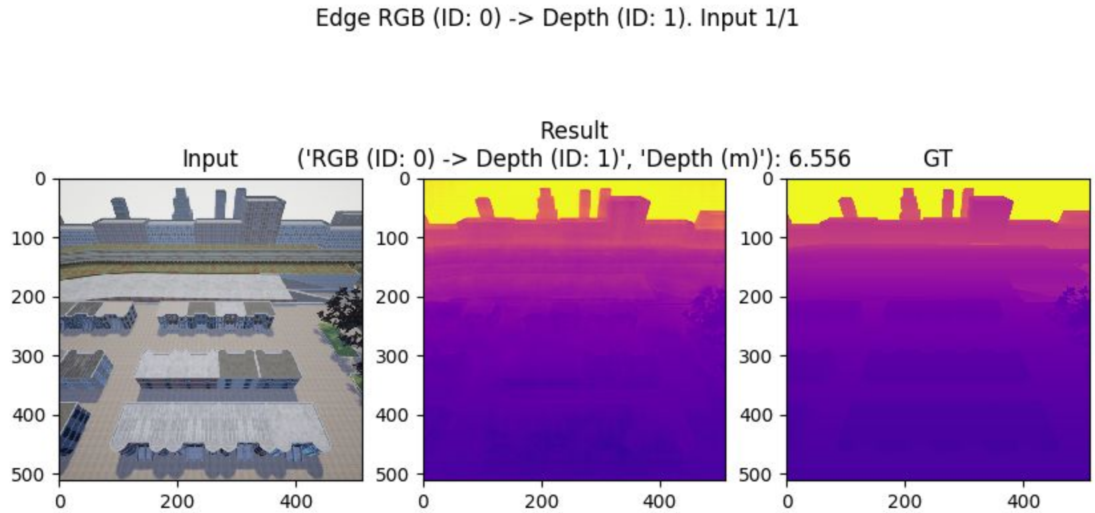


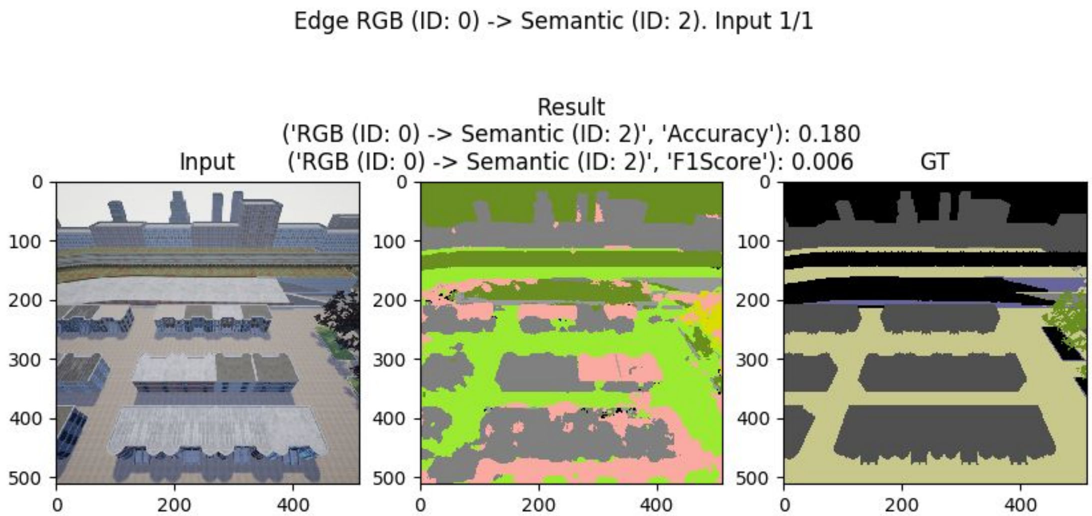


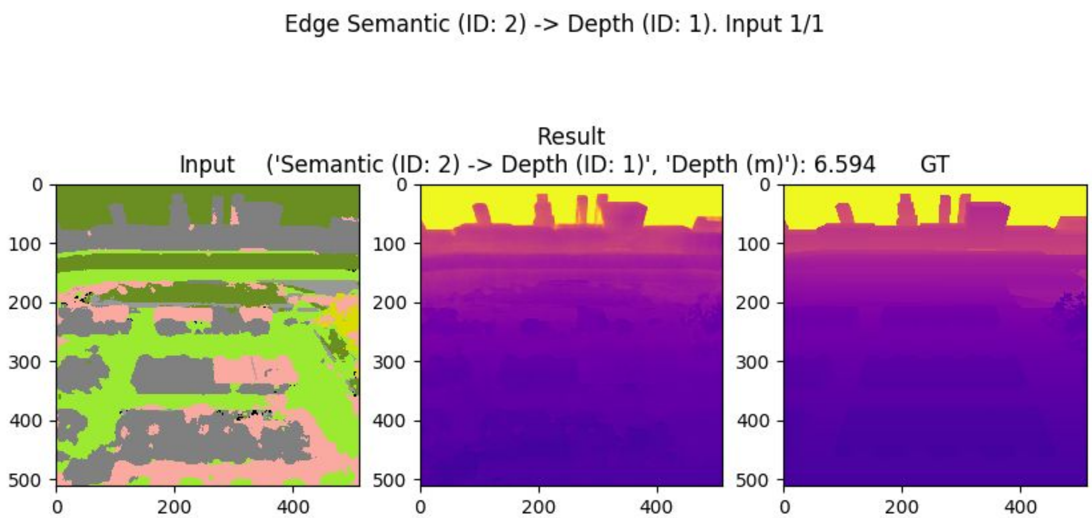
The result on the testset was 5.316 mean depth error. In the images below we can see the inputs, the results and the ground truths for one frame.


Edge Depth (ID: 1) -> Depth (ID: 3). Input 1/2
Input
Result
('Depth (ID: 1) -> Depth (ID: 3)', 'Depth (m)'): 6.363
GT


Edge Depth (ID: 1) -> Depth (ID: 3). Input 2/2
Input
Result
('Depth (ID: 1) -> Depth (ID: 3)', 'Depth (m)'): 8.340
GT


Edge RGB (ID: 0) -> Depth (ID: 1). Input 1/1
Input
Result
('RGB (ID: 0) -> Depth (ID: 1)', 'Depth (m)'): 6.556
GT
0
100
200
300
400
500
0
200
400





Benchmark

In order to compare with the normal rgb -> depth and rgb -> semantic experiments were also done on these graphs and the final results can be seen in the tables below.

	Accuracy	F1Score	Loss
rgb2semantic	92.813	0.715	0.423
depth2semantic	91.198	0.755	0.490
rgb2depth2semantic_W_rgb2semantic_2finalSemantic	92.476	0.642	0.844

	Depth	Loss
rgb2depth	5.658	0.001
semantic2depth	5.435	0.001
rgb2semantic2depth_W_rgb2depth_2finalDepth	5.316	0.002

It can be seen that using semantic segmentation for depth estimation improved the performance, but using the depth for semantic segmentation did not produce evidence of improvement.

Conclusion

In conclusion, depth and semantic segmentation are two representations that can share information between one another and have the potential to improve performance of the desired task when used together.