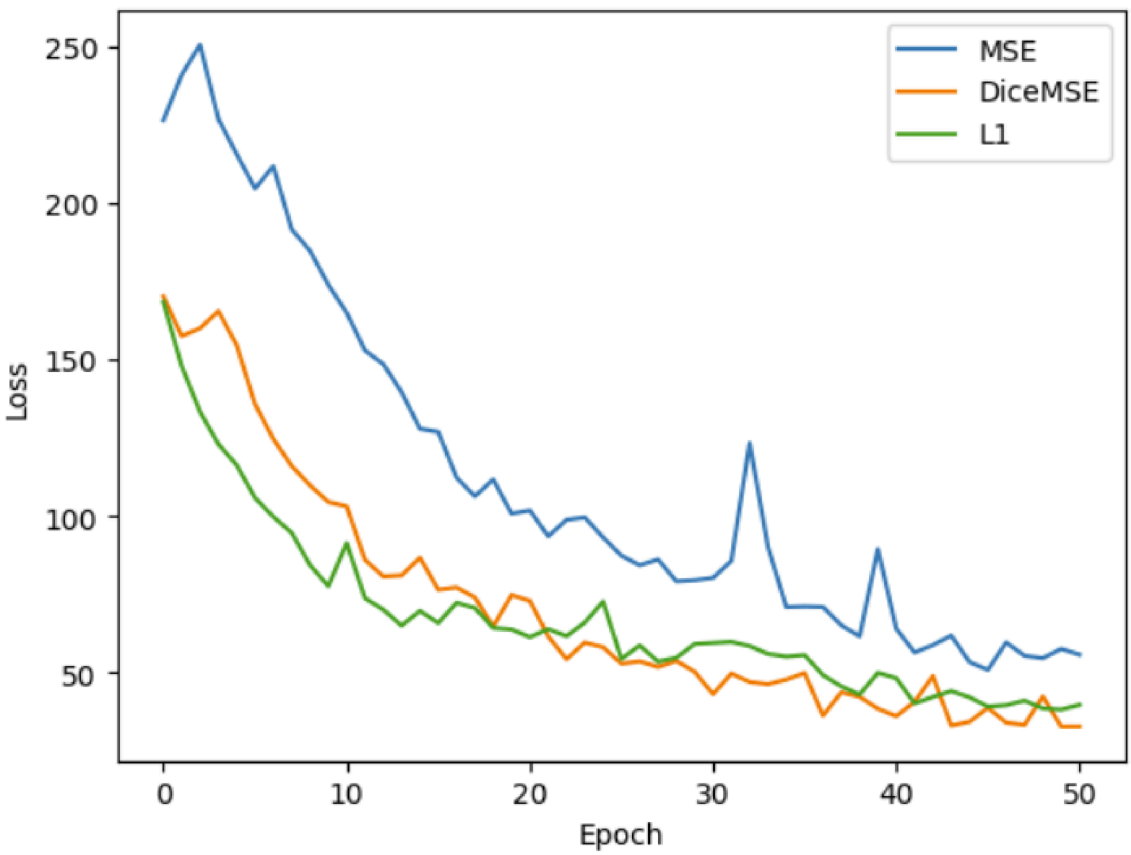
2023 IEEE GRSS DATA FUSION CONTEST: LARGE-SCALE FINE-GRAINED BUILDING CLASSIFICATION FOR SEMANTIC URBAN RECONSTRUCTION
Track 2: Multi-Task Learning of Joint Building Extraction and Height Estimation

For this project, we fixed the model to be U-net, the optimizer to be ADAM(lr=0.0001)
Building extraction and height estimation using multimodal optical and SAR satellite imagery is a regression problem.

Experiment-1 Selecting the loss function

We compared three different loss functions
1. MSE Loss
2. L1 Loss
3. DiceMSE Loss

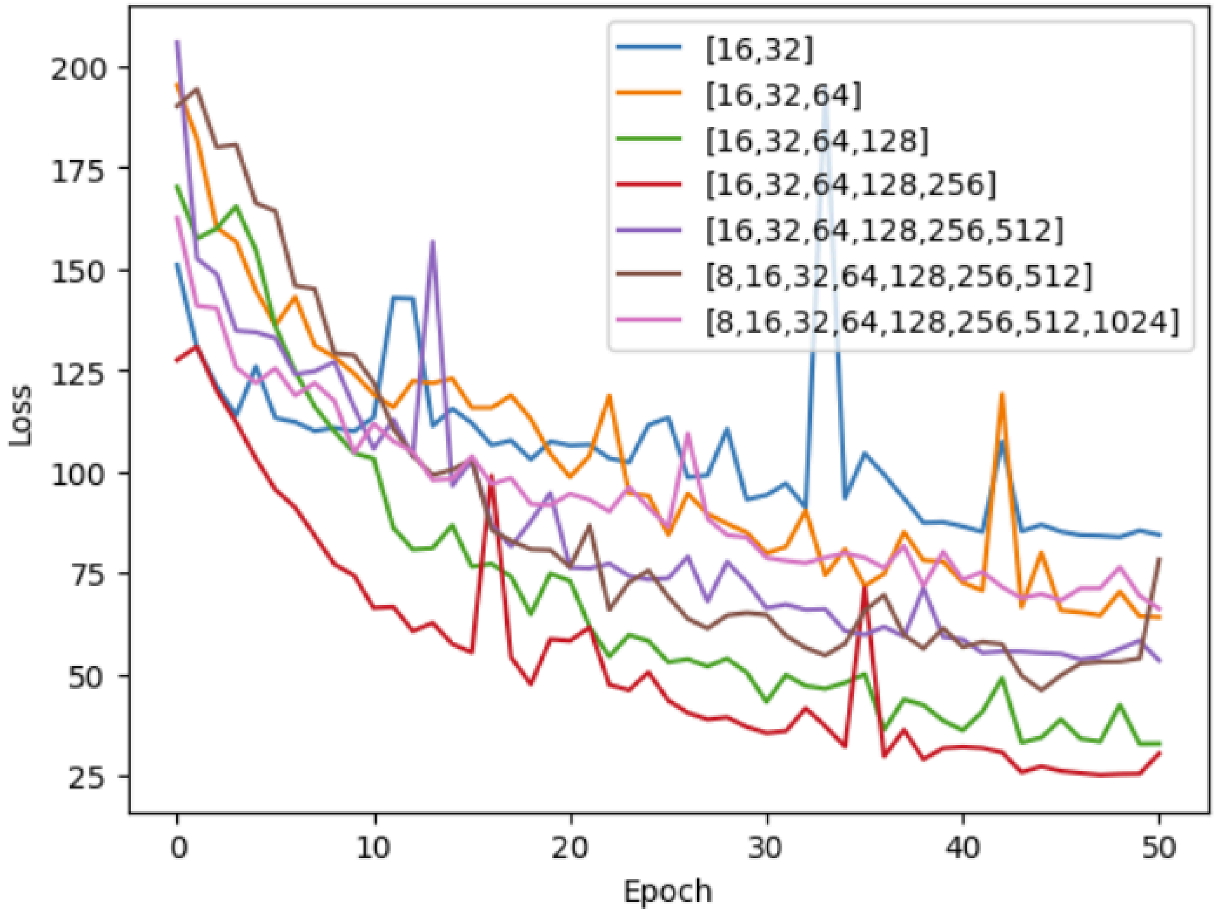


DiceMSE is designed by combining two matrices **DiceMSE Loss = MSELoss / Dice**
1. Dice score for calculating the overlapping between the target and output
2. MSE for building height regressing

Conclusion: both L1 loss and DIceMSE loss performed better than MSE loss with DiceMSE performing only slightly better than L1 loss

Experiment-2 Selecting the depth of the network

We fixed the number of filters in each layer at an increasing degree of 2 and ran each model for 50 epochs with DiceMSE as the loss function

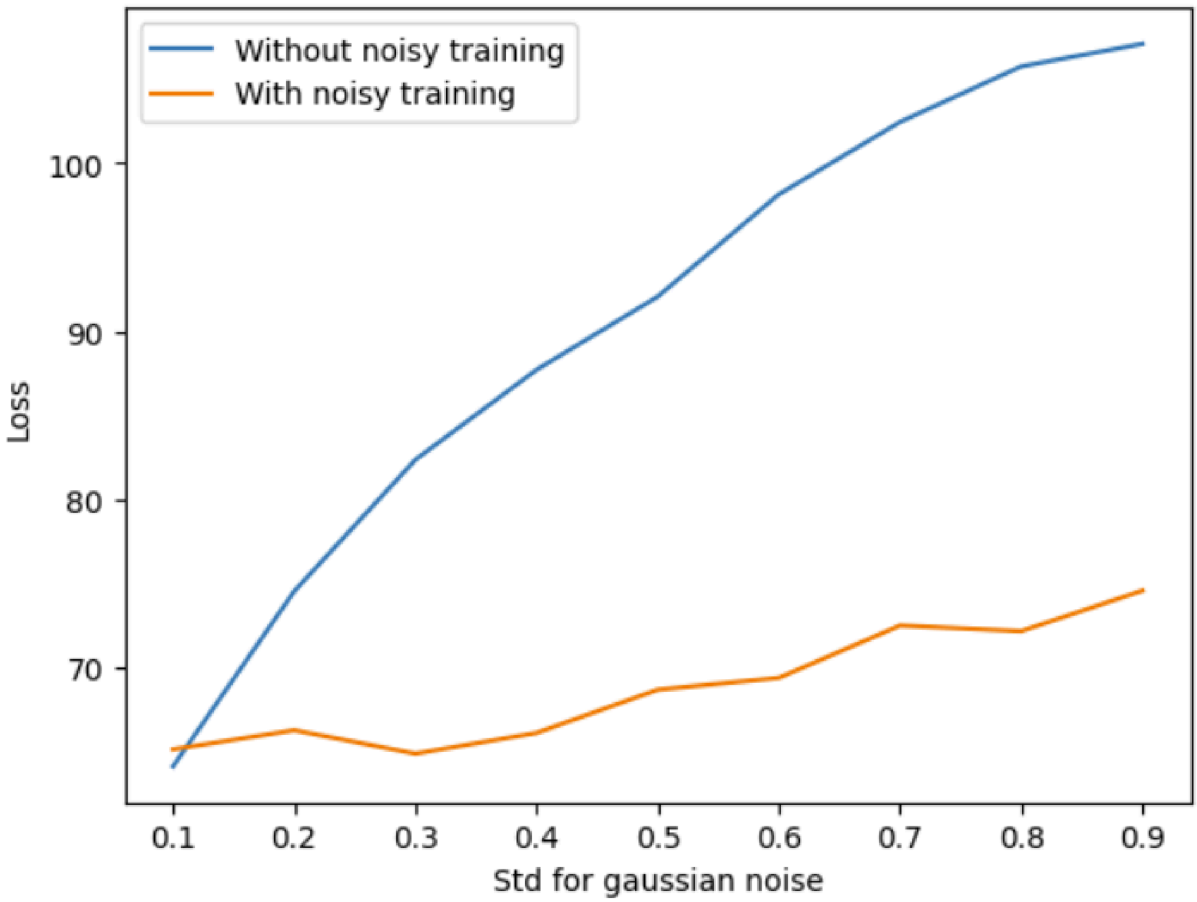


Conclusion: It is clear from the graph that networks with depths 4 and 5 performed the best.

Experiment-3 Effect of adding noise during training

We trained the 4-layer model with DiceMSE loss for 10 epochs

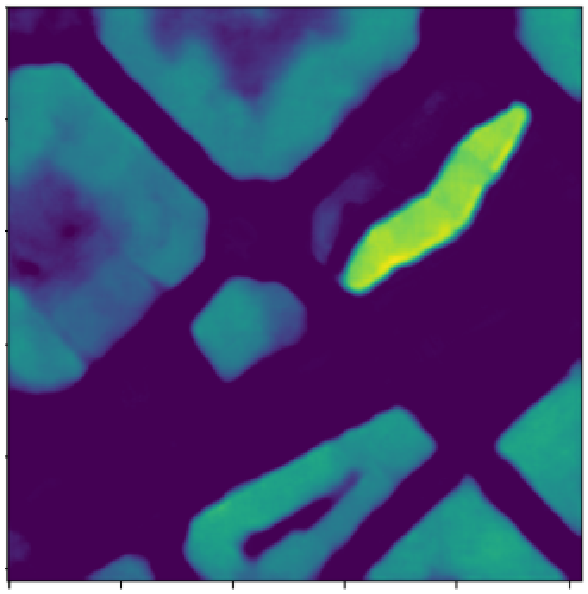
1. Without adding noise to the training data but adding noise to the testing data
2. Adding noise to the training data and testing data



Conclusion: This type of training not only increased the robustness of the model against noisy data but also increased accuracy when there was a partial cloud cover in the optical image.

Results till now
Best performing model { layers : **[16,32,64,128,256]** loss: **DiceMSE** optimizer: **Adam** }

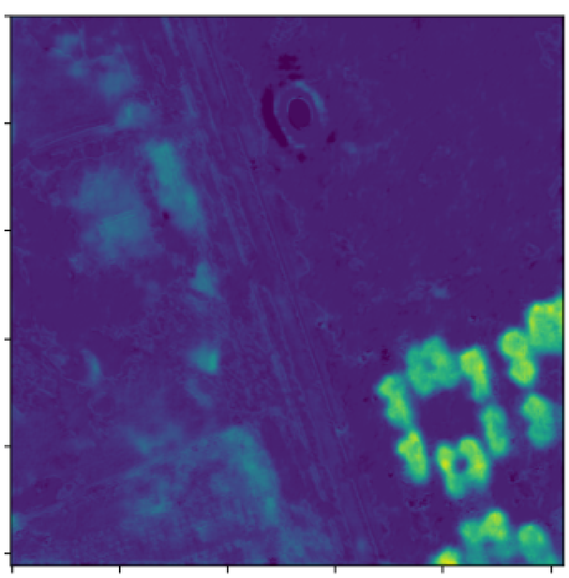
Training data

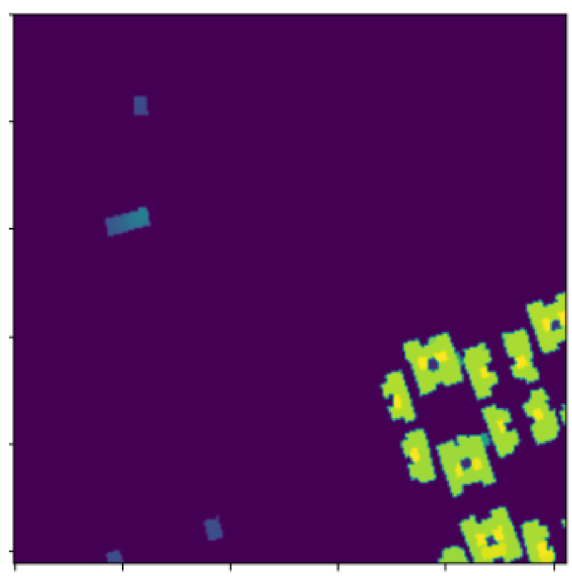


Prediction
MSE: **33.2418**
Dice score: **0.9070**

Ground Truth

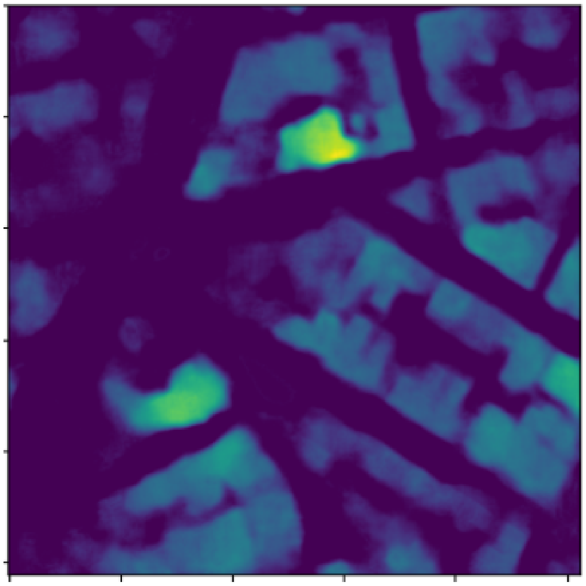
Testing data

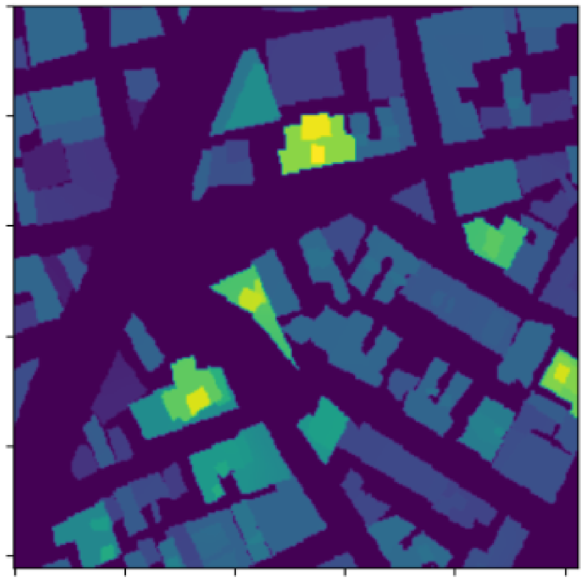




Prediction
MSE: **8.0506**
Dice score: **0.1324**

Ground Truth





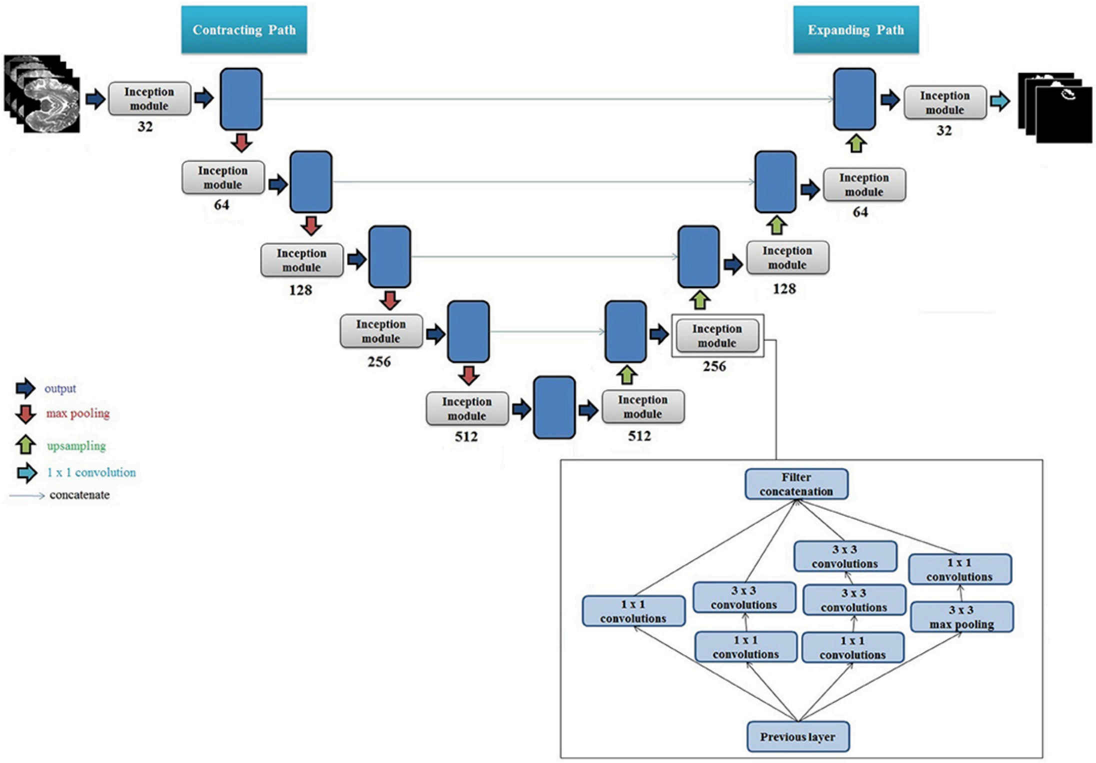
Prediction
MSE: **96.0360**
Dice score: **0.6235**

Ground truth

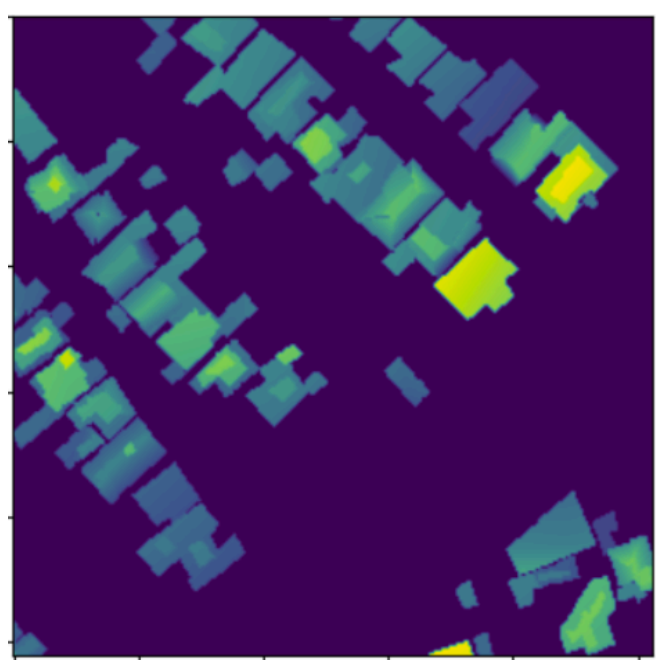
Conclusion: when buildings are closely clustered it becomes difficult for the model to properly mask buildings and estimate their height.
The model works well with large continuous buildings but struggles with smaller buildings.

Experiment-4: Combining the Inception module with Unet to learn smaller objects

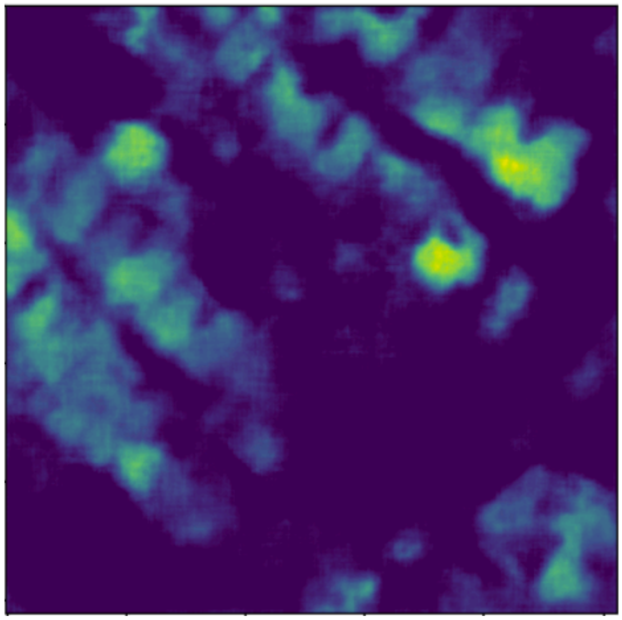
Architecture:



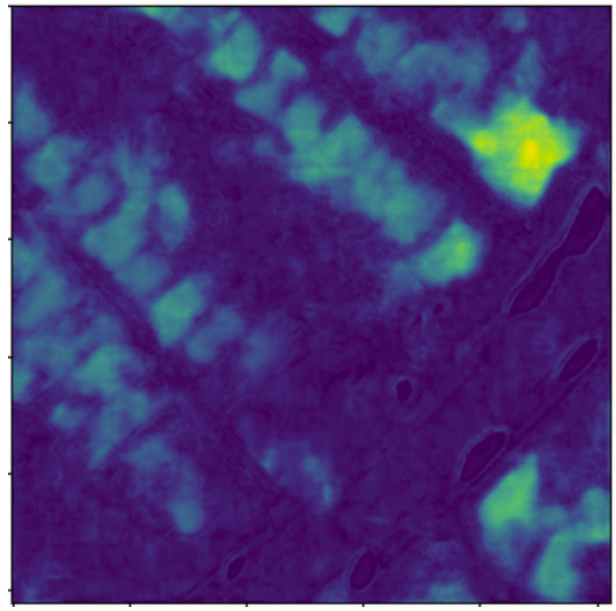
Results:



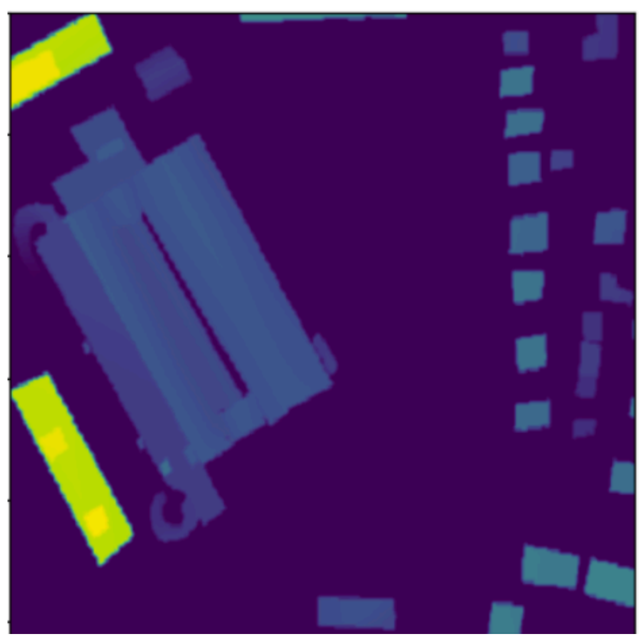
Ground Truth



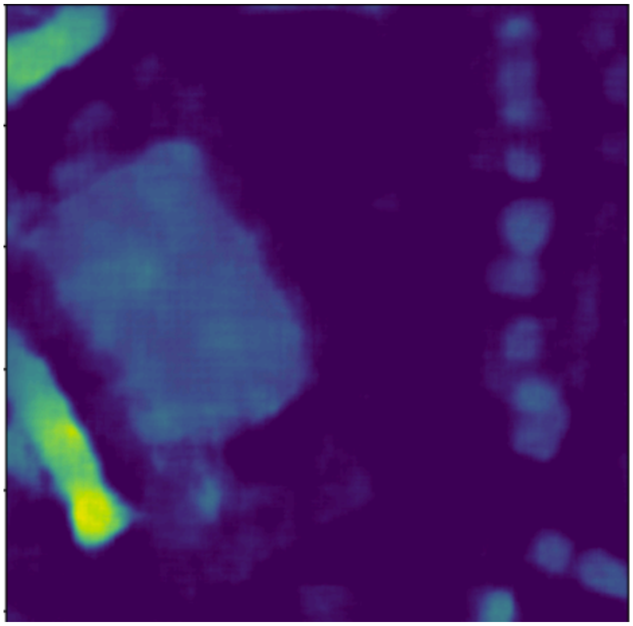
InceptionUnet
MSE: **15.6899**
Dice score: **0.6478**



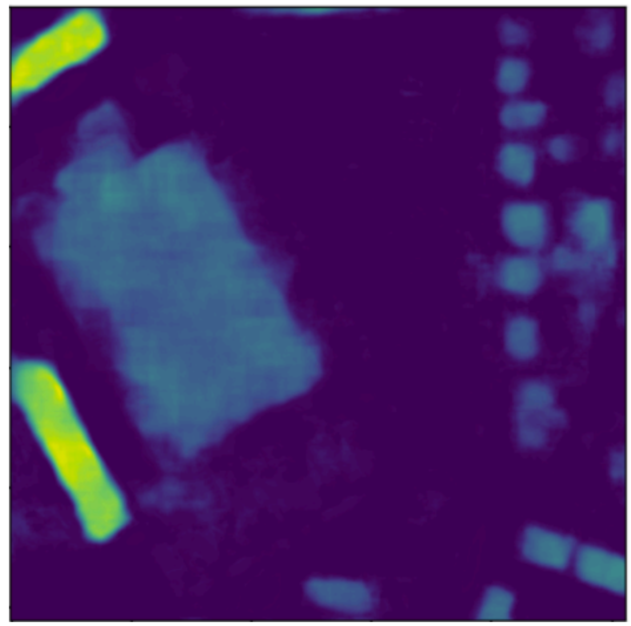
BaseUnet
MSE: **5.0163**
Dice score: **0.5896**



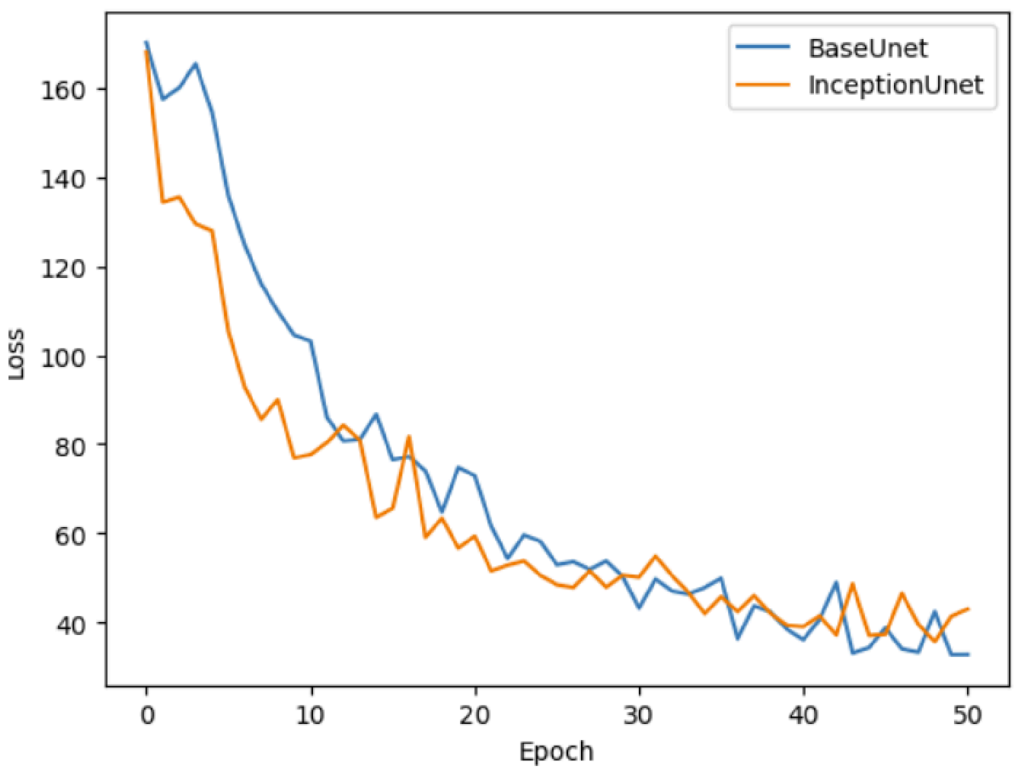
Ground Truth



InceptionUnet
MSE: **13.5244**
Dice score: **0.7320**

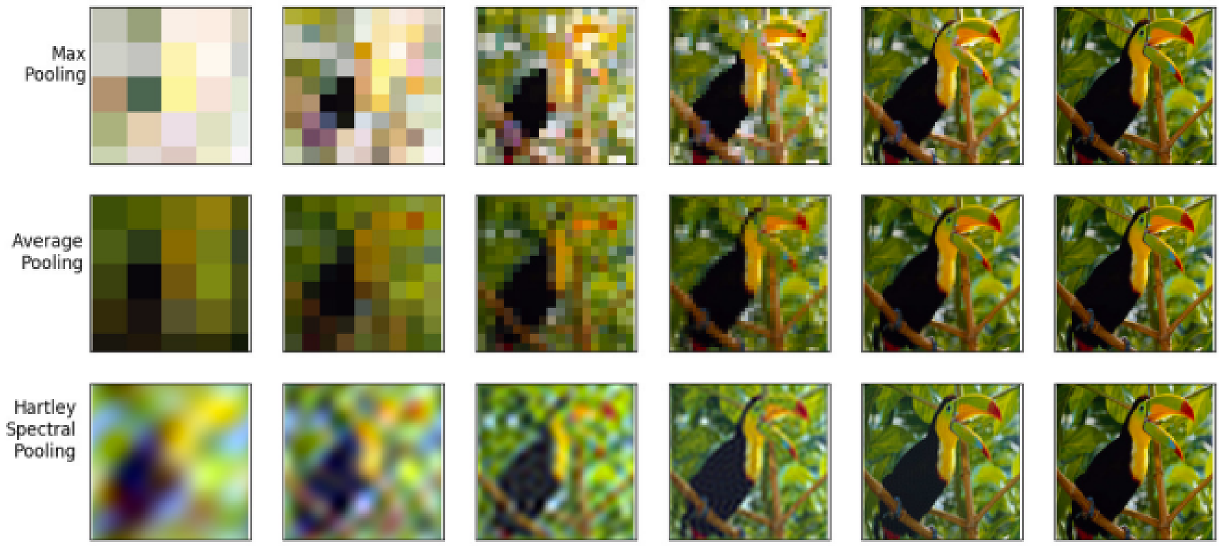


BaseUnet
MSE: **5.9268**
Dice score: **0.7021**

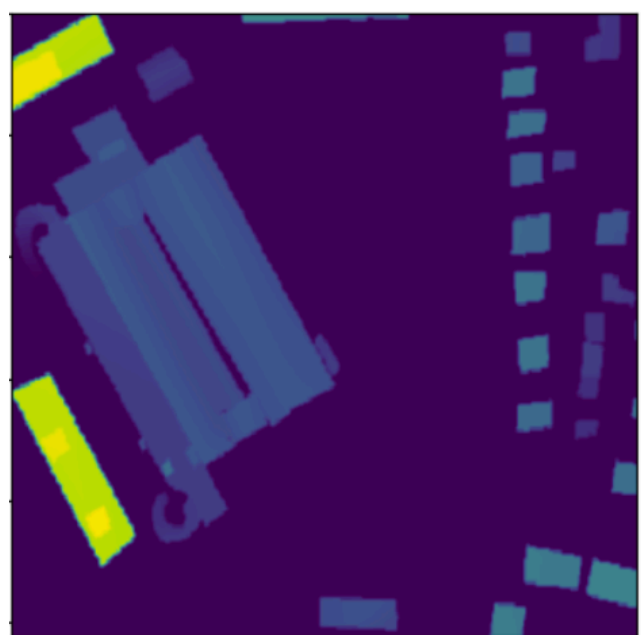


Conclusion: No visible advantage BaseUnet still performs better for most of the examples

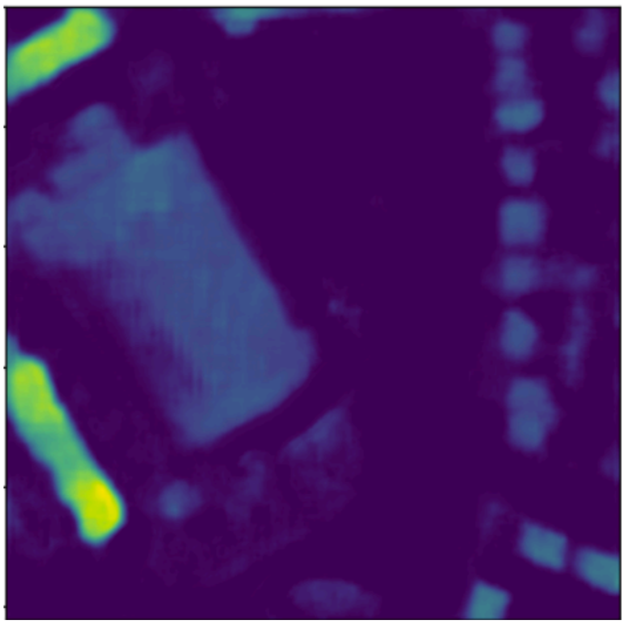
Experiment-5: Applying Spectral pooling instead of Max or Avg pooling since it retains more information after the pooling operation, in order to learn borders more precisely



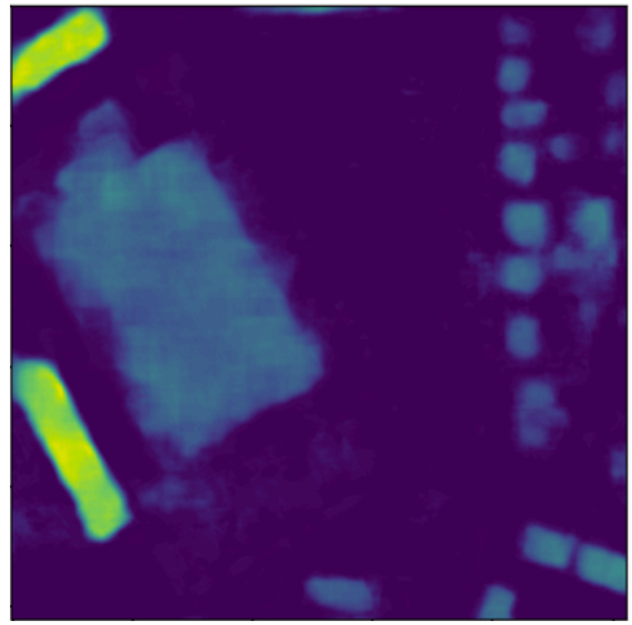
Results:



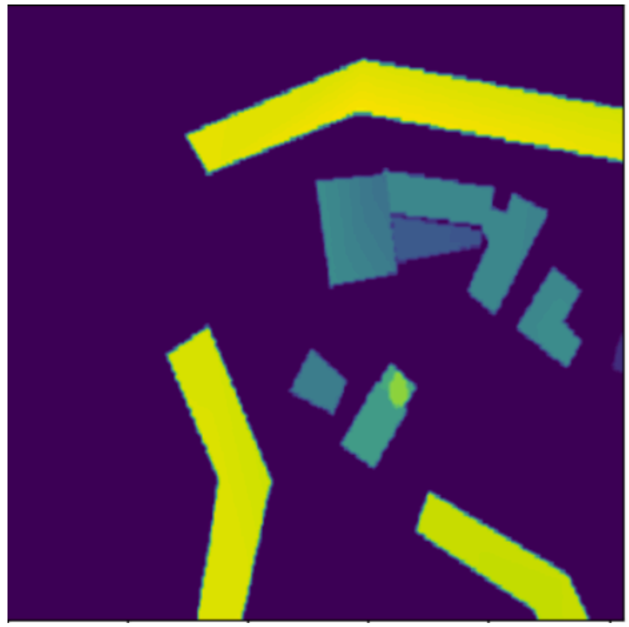
Ground Truth



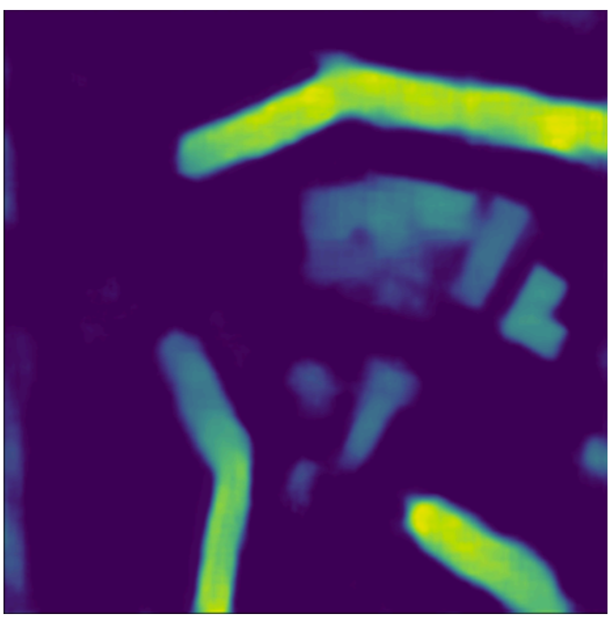
Spectral pool
MSE: **9.3442**
Dice score: **0.7049**



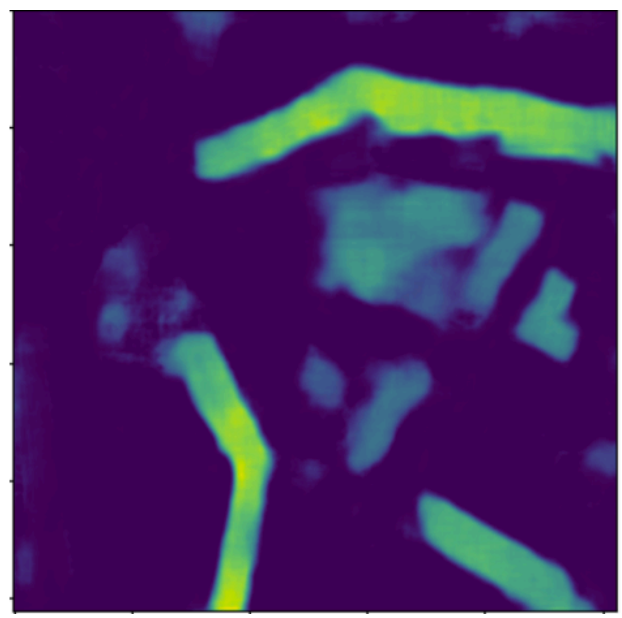
Max pool
MSE: **5.9268**
Dice score: **0.7021**



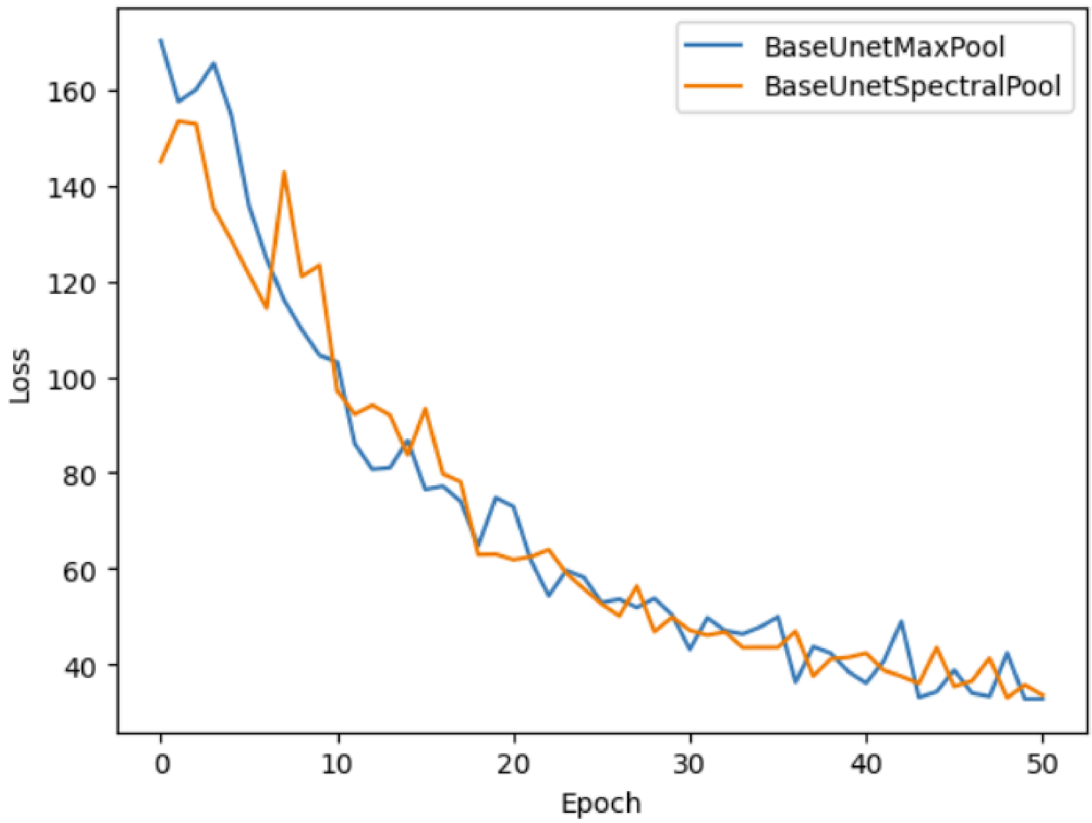
Ground Truth



Spectral pool
MSE: **25.8823**
Dice score: **0.7352**

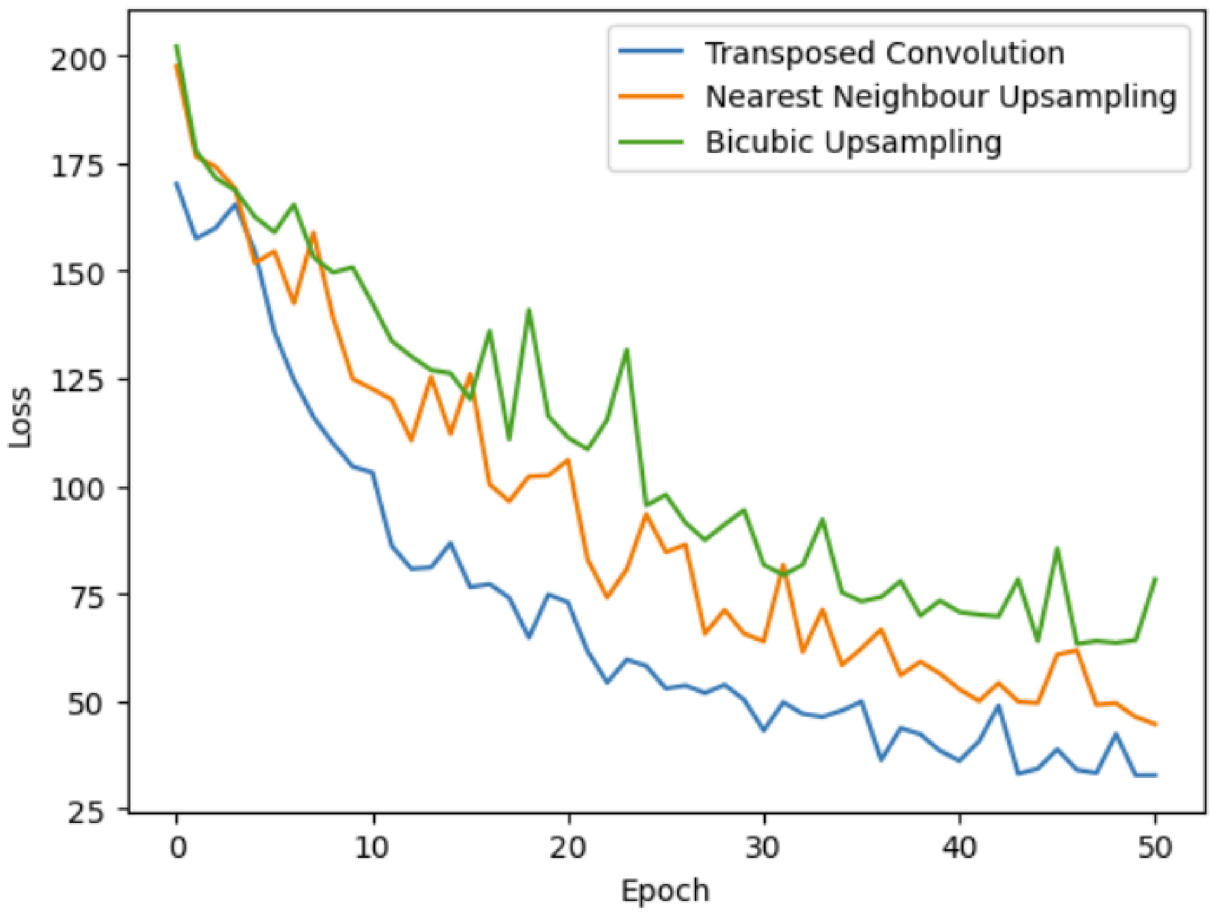


Max pool
MSE: **38.8715**
Dice score: **0.6599**



Conclusion: Spectral pooling improved the overall edge detection

Experiment-6: Trying different upsampling methods



Conclusion: Transposed convolution-based upsampling performed the best.

Future Scope:

Using anchor box based network for detecting buildings more precisely
Experimenting with label noises
Optimizing loss metric
Experimenting with OOD datasets