1

Results

In this chapter, we will discuss about the experiments conducted for out-of-distribution (OOD) detection on RandLA-Net using uncertainty techniques such as deep ensembles and flipout. These experiments were conducted on Semantic3D dataset as training in-distribution (ID) dataset and two datasets particularly S3DIS and Toronto3D are used as OOD datasets. The detailed description about datasets can be found in Chapter cite here, about RandLA-Net in Section cite here, and about deep ensembles and flipout in Section cite here The list of experiments conducted to acheive OOD detection are as follows:

1. Train and evaluate the deep ensembles of RandLA-Net on Semantic3D dataset and discussed in Section 1.1.

1.1 Deep ensembles-Semantic3D

Aim: Train multiple models of RandLA-Net with random initializations on Semantic3D dataset and combine the results from the get combined prediction. These models are evlauated using mean Intersection-over-Union (mIOU) and Accuracy.

		IoU per class								
#Ensembles	**MeanIOU**	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**Accuracy**
1	68.19	94.55	81.19	84.67	29.43	81.37	18.85	64.74	90.74	88.78
5	69.51	94.73	81.92	84.42	28.05	**86.41**	28.50	61.03	91.03	90.04
10	69.97	95.25	83.73	86.63	30.36	84.13	18.60	**66.01**	92.61	89.94
15	70.32	95.27	83.54	**88.22**	**32.19**	84.82	26.17	61.67	90.75	**90.57**
20	**70.80**	**95.55**	**84.11**	86.65	29.60	85.41	**29.58**	62.47	**93.06**	90.56

Table 1.1: Illustration of performance of RandLA-Net on Semantic3D over number of ensembles. meanIOU and IOU per class and overall accuracy are represented here. C1 to C8 are the classes of Semantic3D which are Manmadeterrain, Naturalterrain, Highvegetation, Lowvegetation, Buildings, Hardscapes, Scanningartifacts, and Cars.

We trained 20 models of RandLA-Net over Semantic3D dataset with training procedure described in Section cite here. Table 1.1 represents the meanIoU, Accuracy and per class IoU with each row representing the performance value with ensemble size being multiple of 5. Figure 1.1a and 1.1b representing the performance with mIoU and Accuracy with all ensmebles respectively. Some of the predicted points clouds in comparison with ground truth are represented in Figure cite here.

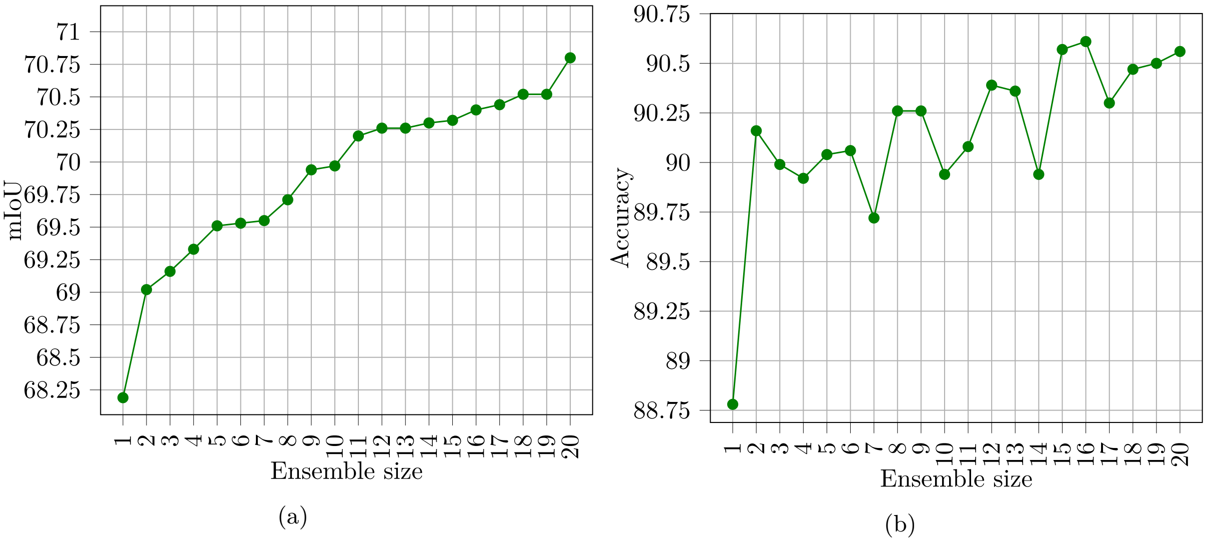


Figure 1.1: Evaluation results of the deep ensemble of RandLA-Net on Semantic3D dataset with (a) representing the meanIoU over ensemble size and (b) representing accuracy over ensemble size.

Conclusions: From this experiment, we draw the following conclusions

- Deep ensembles improved the overall performance of the model significantly in terms of mIoU and Accuracy.
- Figure **??** depicts low training points size for classes low vegetation, hardscapes, scanning artifacts and cars. Because of this classes low vegetation and hardscapes show less mIoU but use of deep ensembles improved their performance significantly.
- Eventhough cars are underrepresented in number of training points, efficient feature extraction of RandLA-Net helps in better segmentation of cars. As this is the same case in SemanticKITTI evaluation proposed in [1].
- From Figure 1.1a, the performance gains in terms of mIOU after the ensemble size of 10 is minimal.

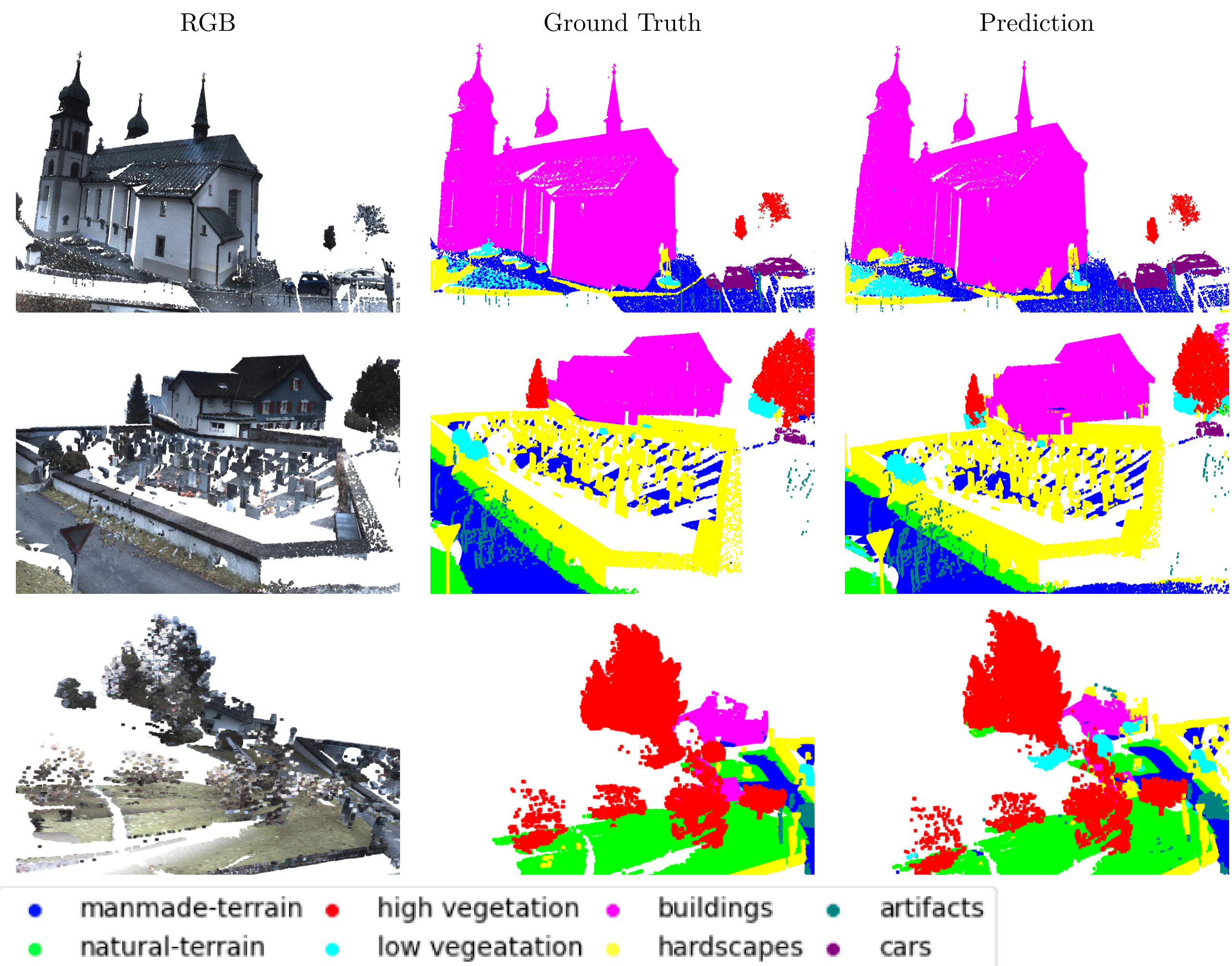


Figure 1.2: Output predictions of the RandLA-Net over the Semantic3D dataset (13 ensemble size) Legend spelling mistake.

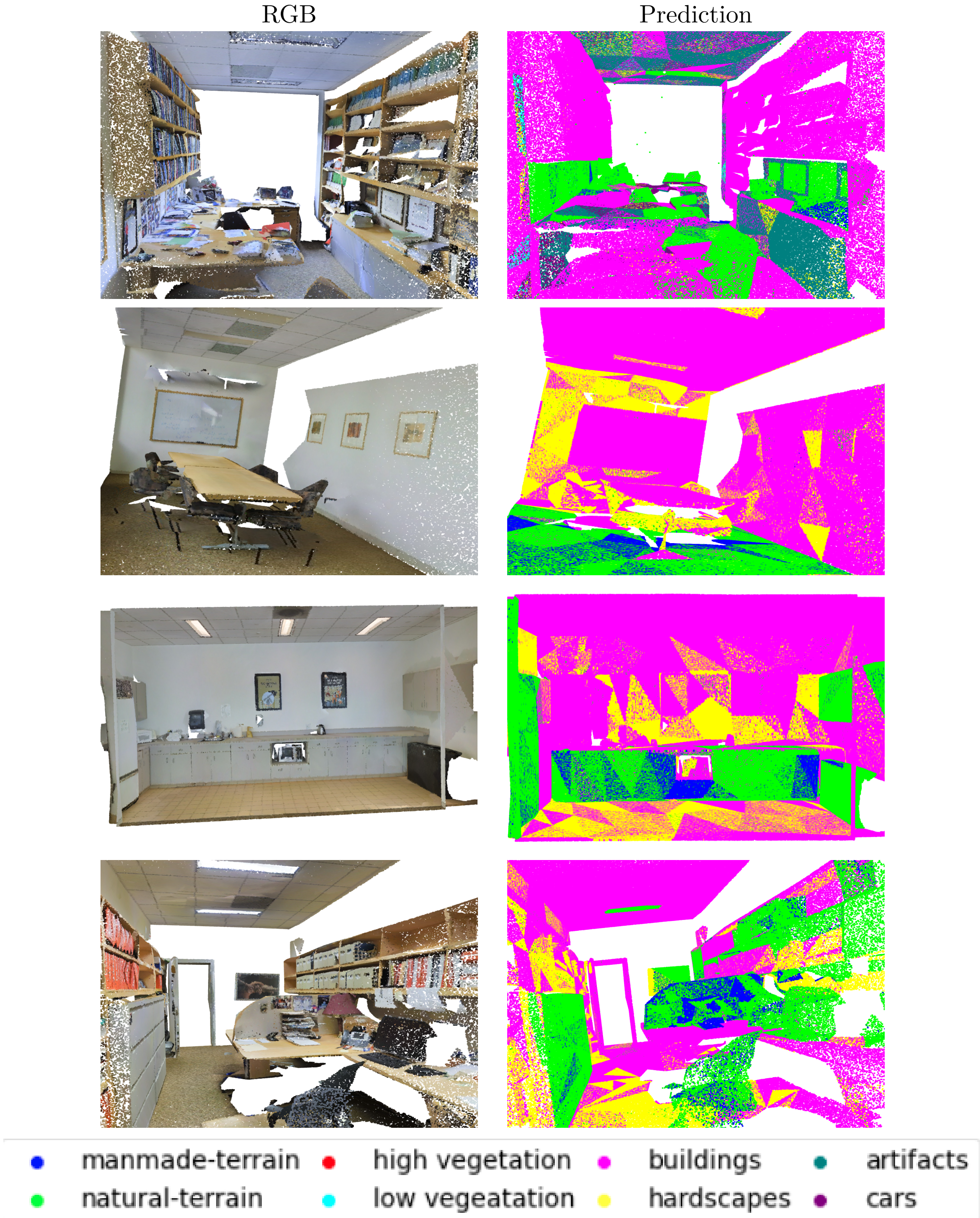


Figure 1.3: Output predictions of the RandLA-Net over the S3DIS dataset.

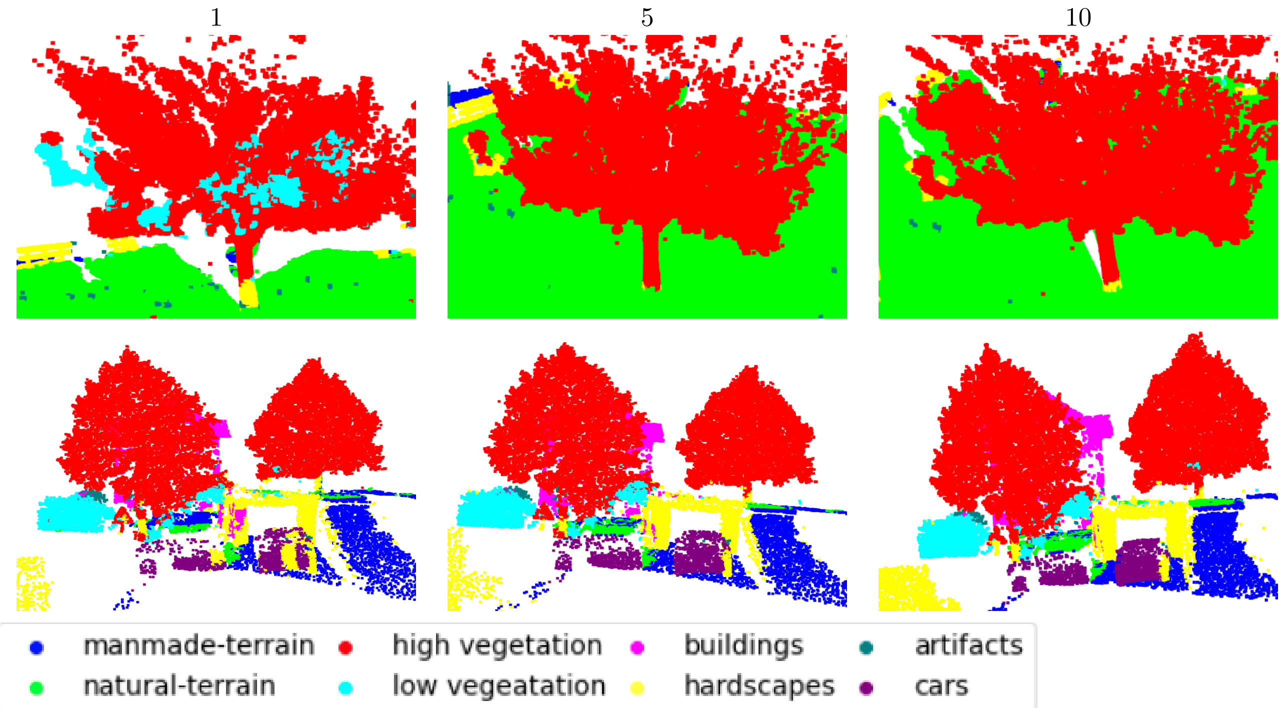
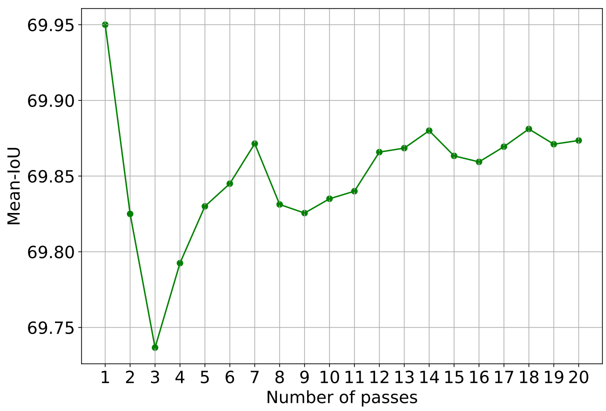


Figure 1.4: Deep ensembles performance on RandLA-Net over the Semantic3D dataset.

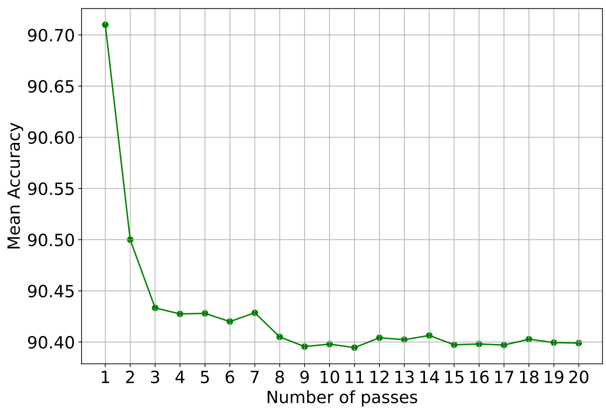
1.2 Flipout-Semantic3D

		IoU per class								
#Passes	MeanIOU	C1	C2	C3	C4	C5	C6	C7	C8	Accuracy
1	69.95	94.24	80.09	86.16	22.48	88.70	39.41	57.42	91.12	90.71
5	69.83	94.38	80.21	84.10	23.32	87.80	39.68	57.75	91.43	90.43
10	69.84	94.38	80.16	83.90	23.46	87.73	39.75	57.83	91.47	90.40
15	69.86	94.38	80.17	83.80	23.48	87.73	39.82	57.96	91.57	90.40
20	69.87	94.38	80.18	83.80	23.57	87.72	39.84	57.92	91.57	90.40

Table 1.2: Illustration of performance of RandLA-Net on Semantic3D over flipout initalized with variance 1. meanIOU and IOU per class and overall accuracy are represented here. C1 to C8 are the classes of Semantic3D which are Manmadeterrain, Naturalterrain, Highvegetation, Lowvegetation, Buildings, Hardscapes, Scanningartifacts, and Cars.



(a)



(b)

Figure 1.5: Flipout meanIOU and accuracy

Add the images of flipout performance here same as figures in deep ensembles

1.3 Maximum Softmax Probability(MSP)-Semantic3D vs S3DIS

Aim: In this experiment, we study how the probability scores are distributed in Semantic3D and S3DIS datasets which are ID and OOD datasets respectively.

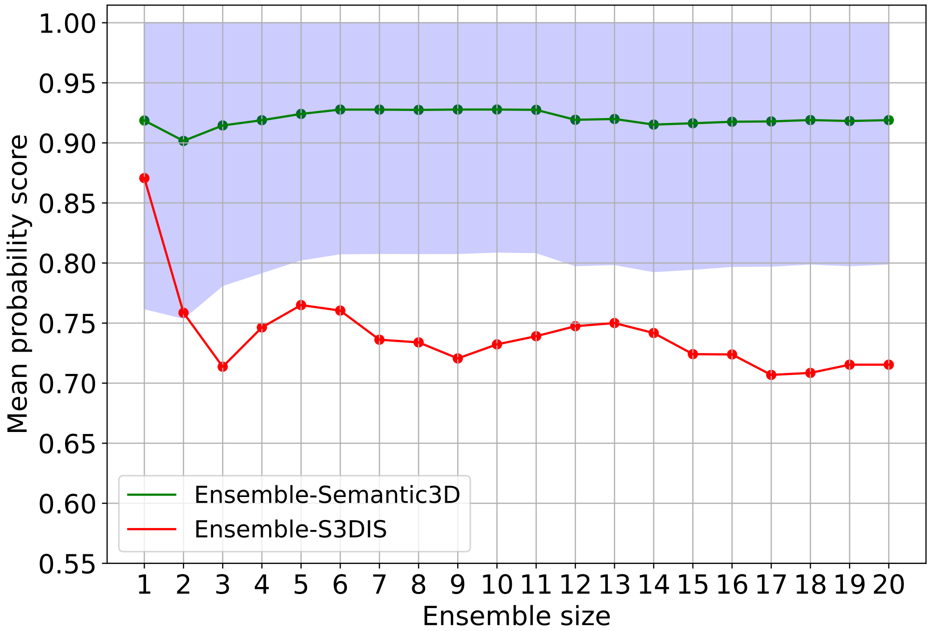


Figure 1.6

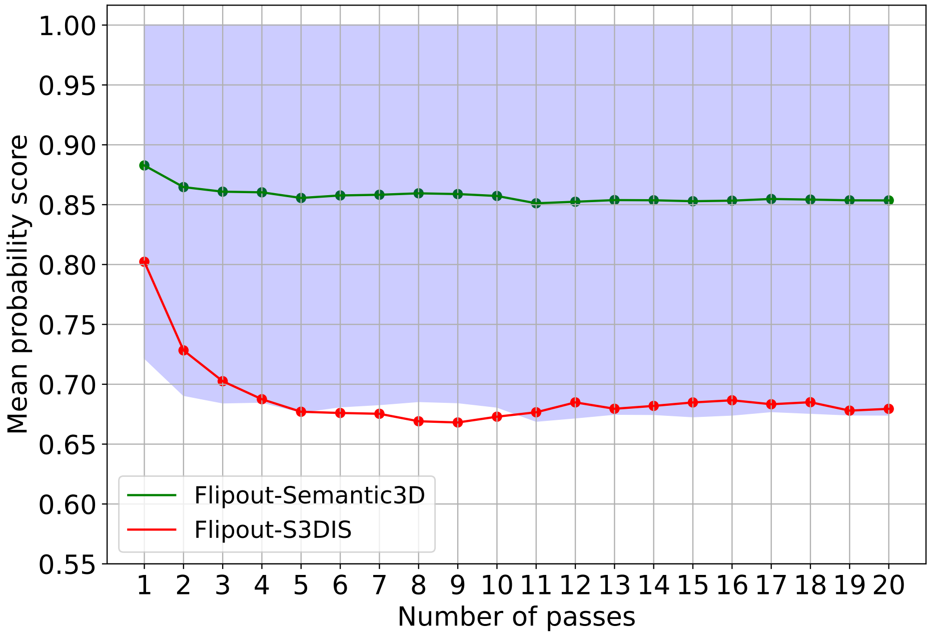


Figure 1.7

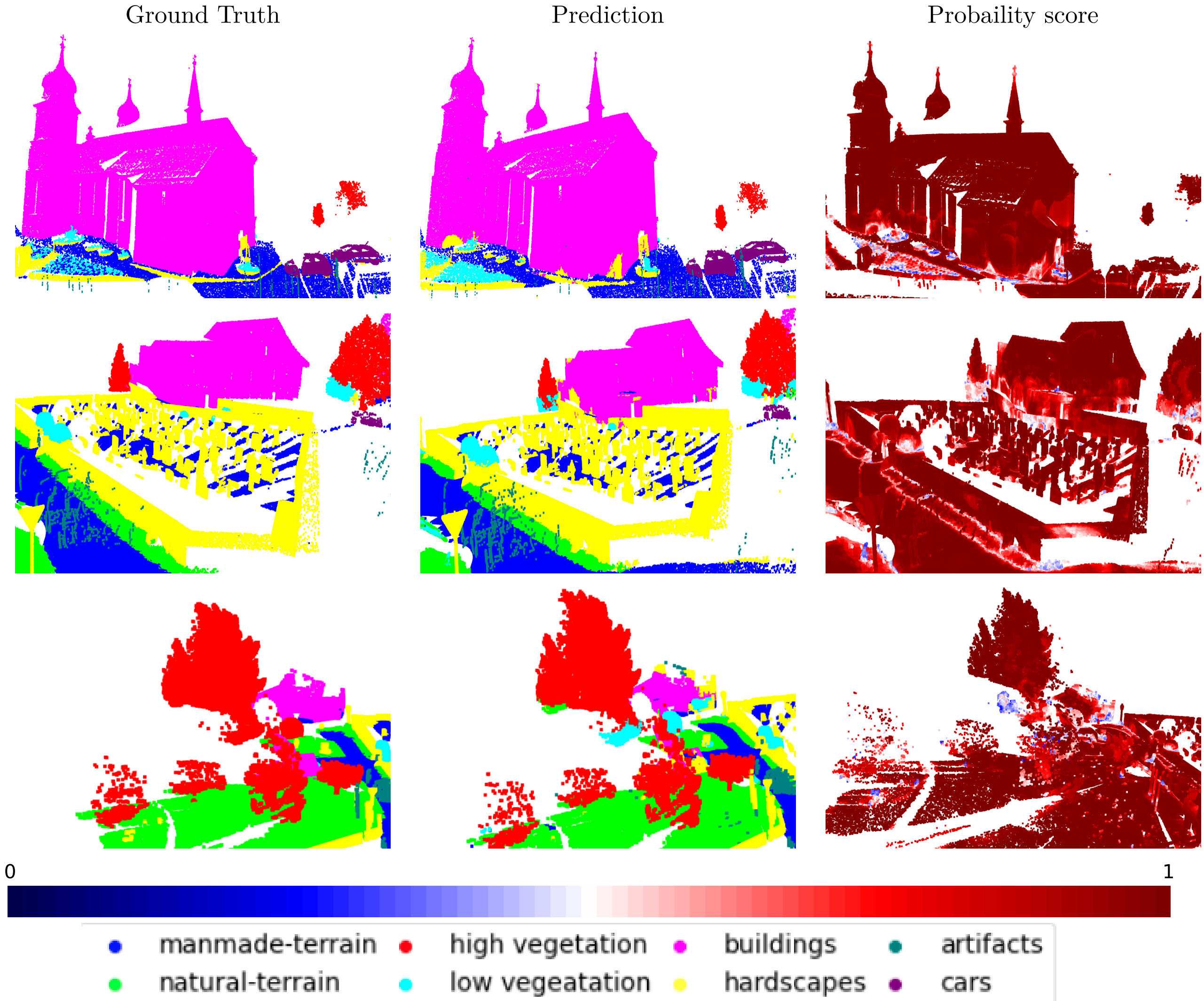


Figure 1.8: Perpoint probability visualization of the semantic3D dataset.

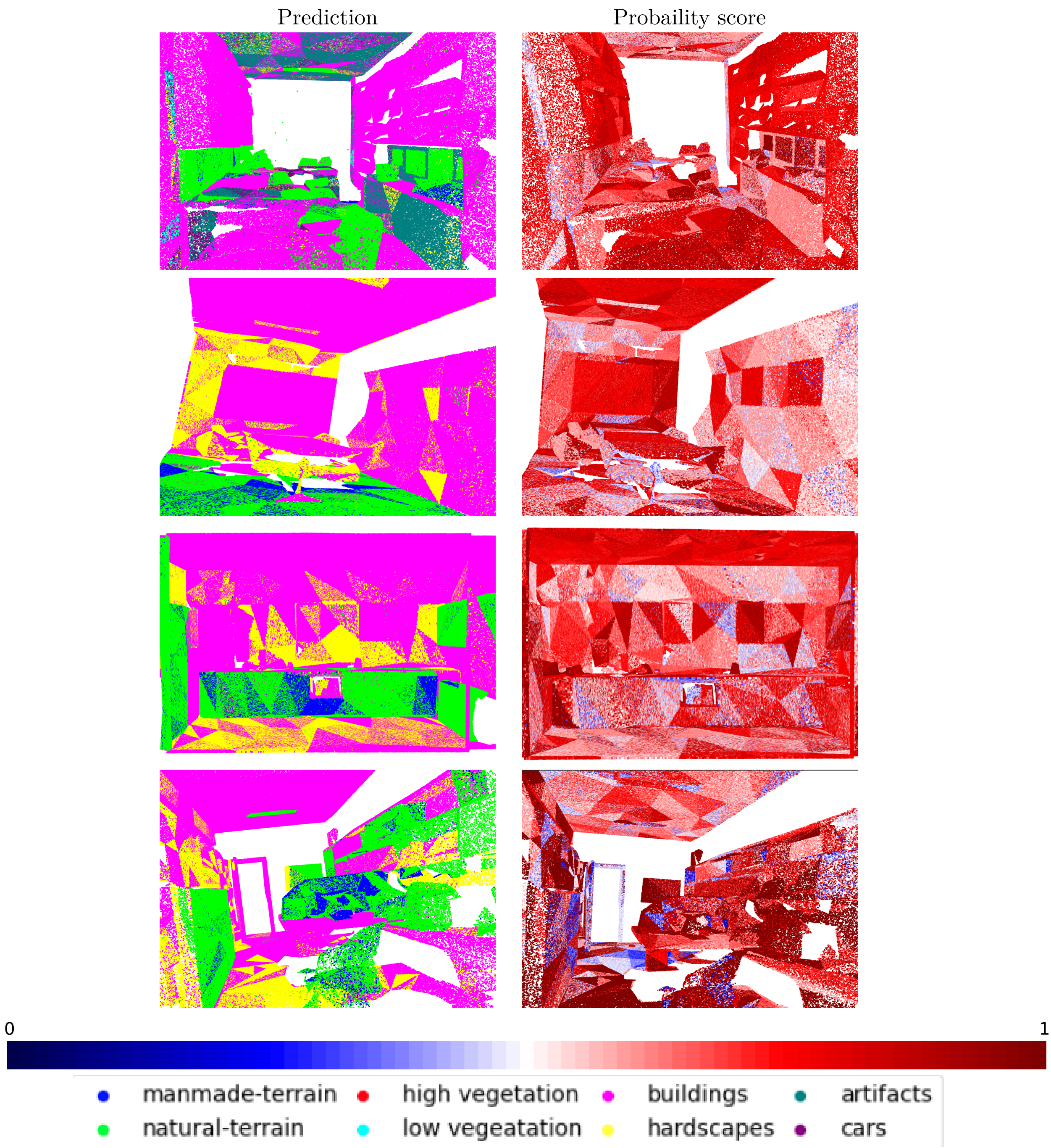


Figure 1.9: Perpoint probability visualization of the S3DIS dataset.

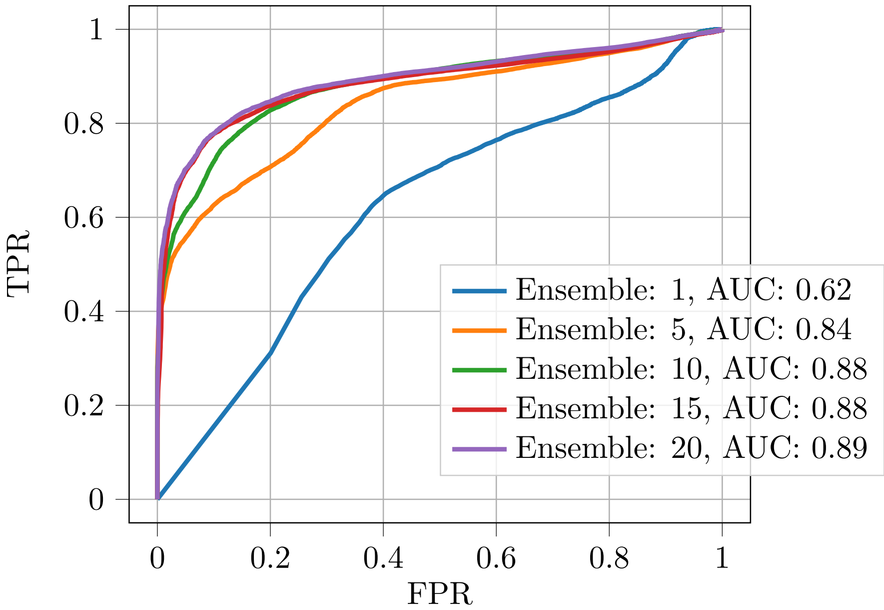


Figure 1.10: ROC plots and AUROC scores for maximum probability in semantic3D vs S3DIS

1.4 Entropy-Semantic3D vs S3DIS

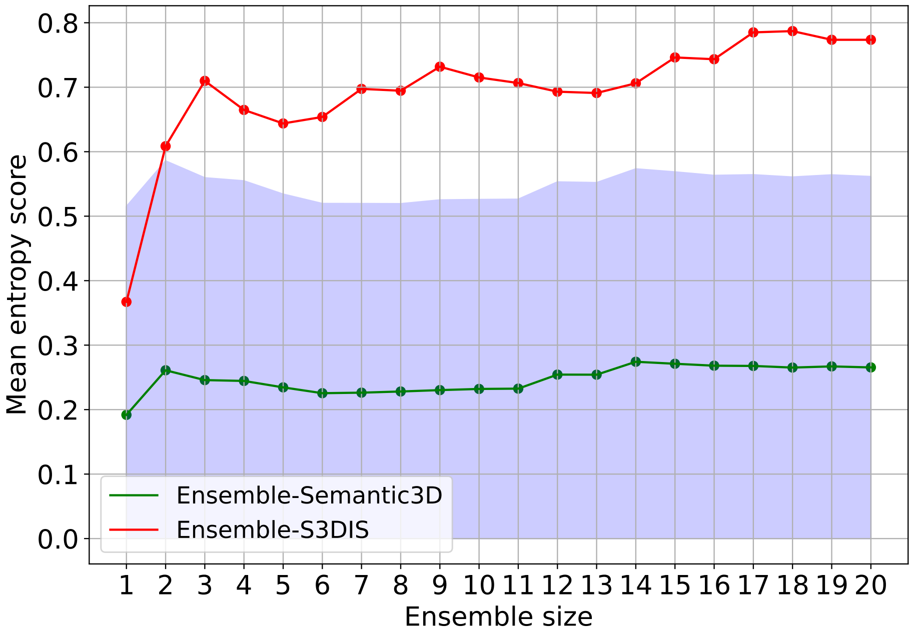


Figure 1.11

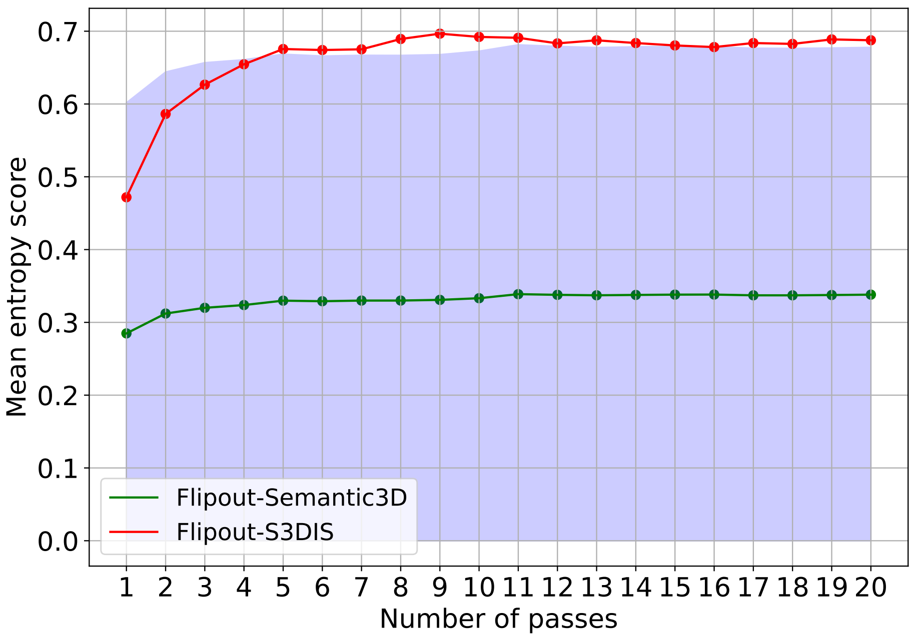


Figure 1.12

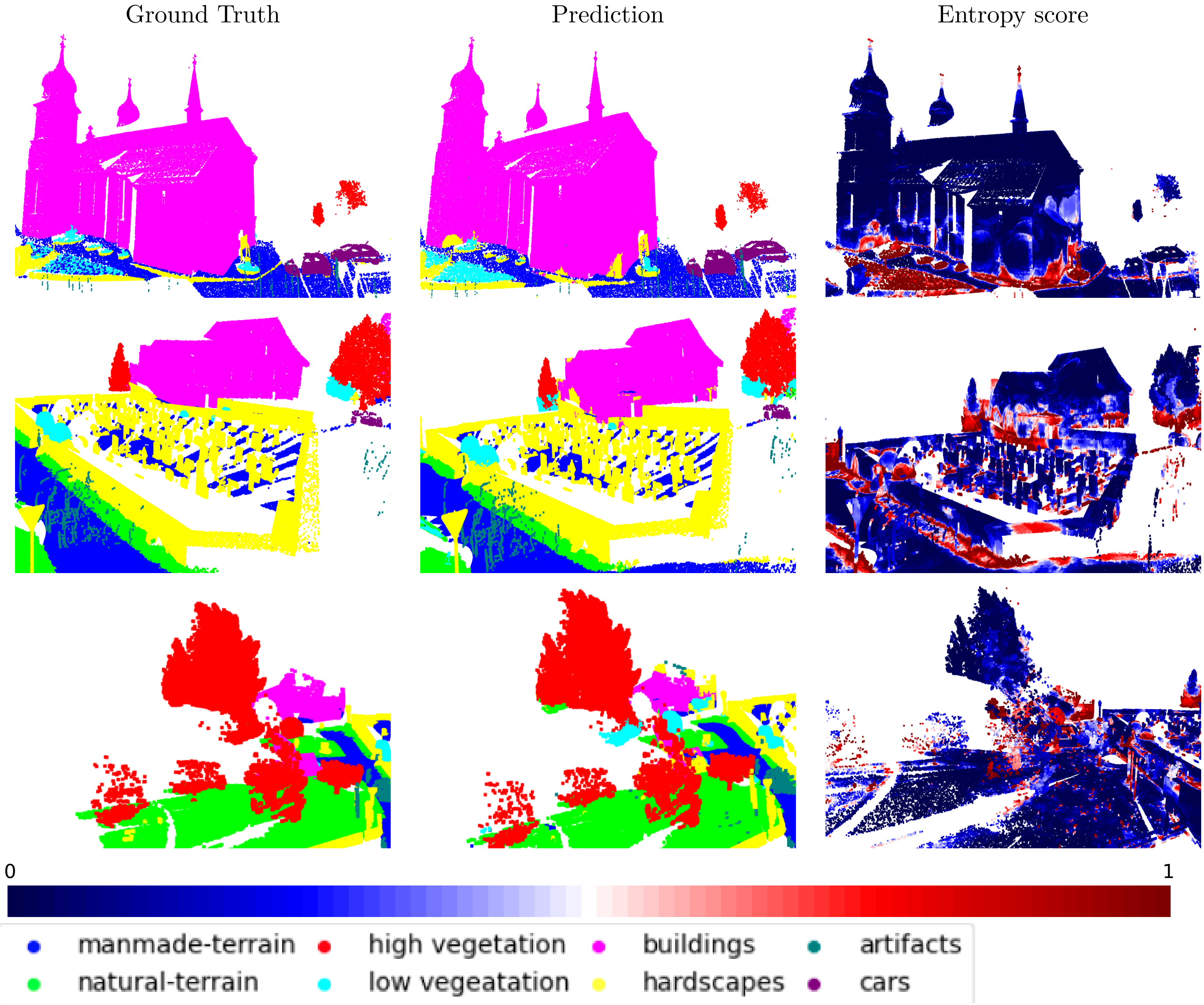


Figure 1.13: Perpoint entropy visualization of the semantic3D dataset.

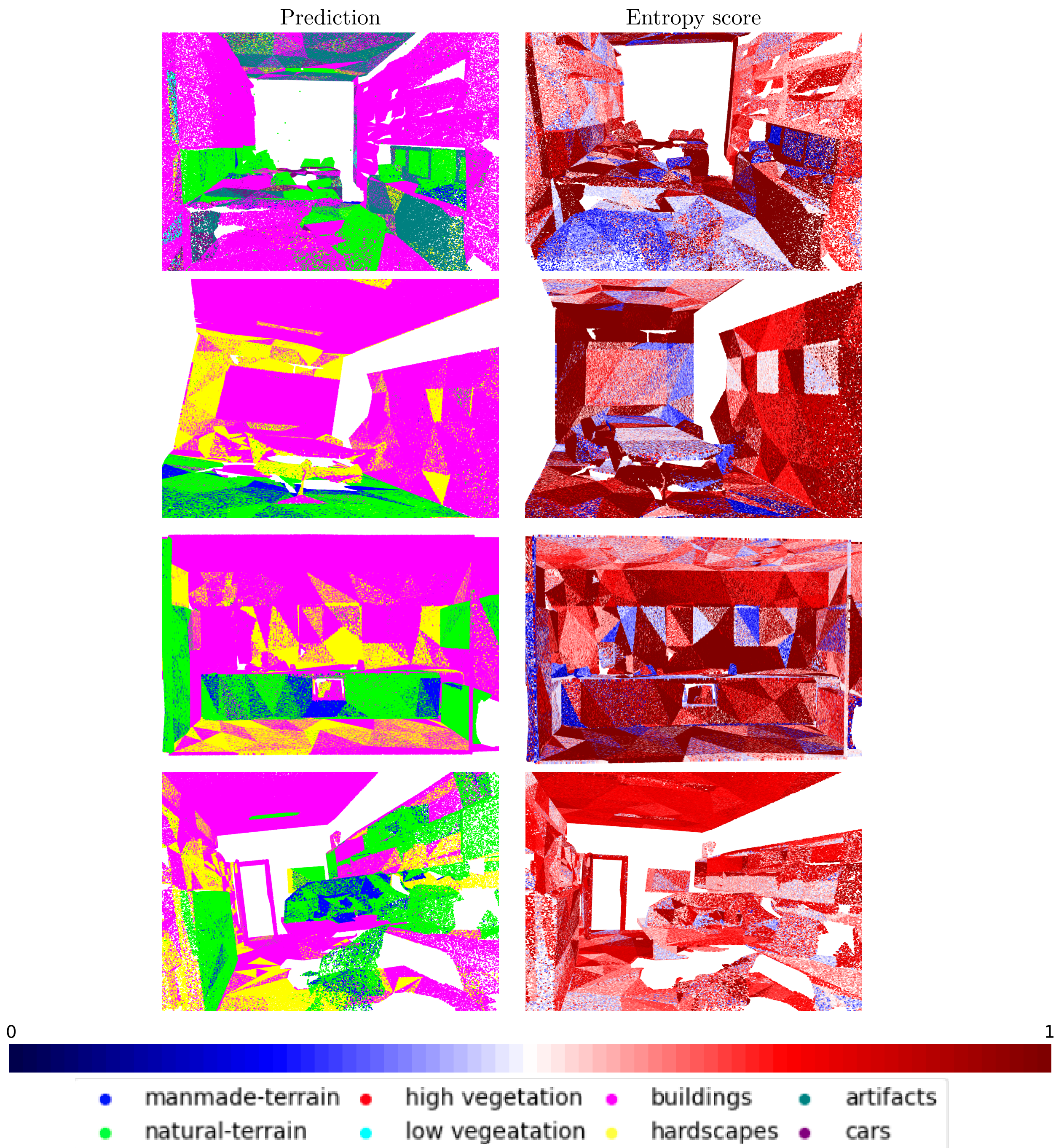


Figure 1.14: Perpoint entropy visualization of the S3DIS dataset.

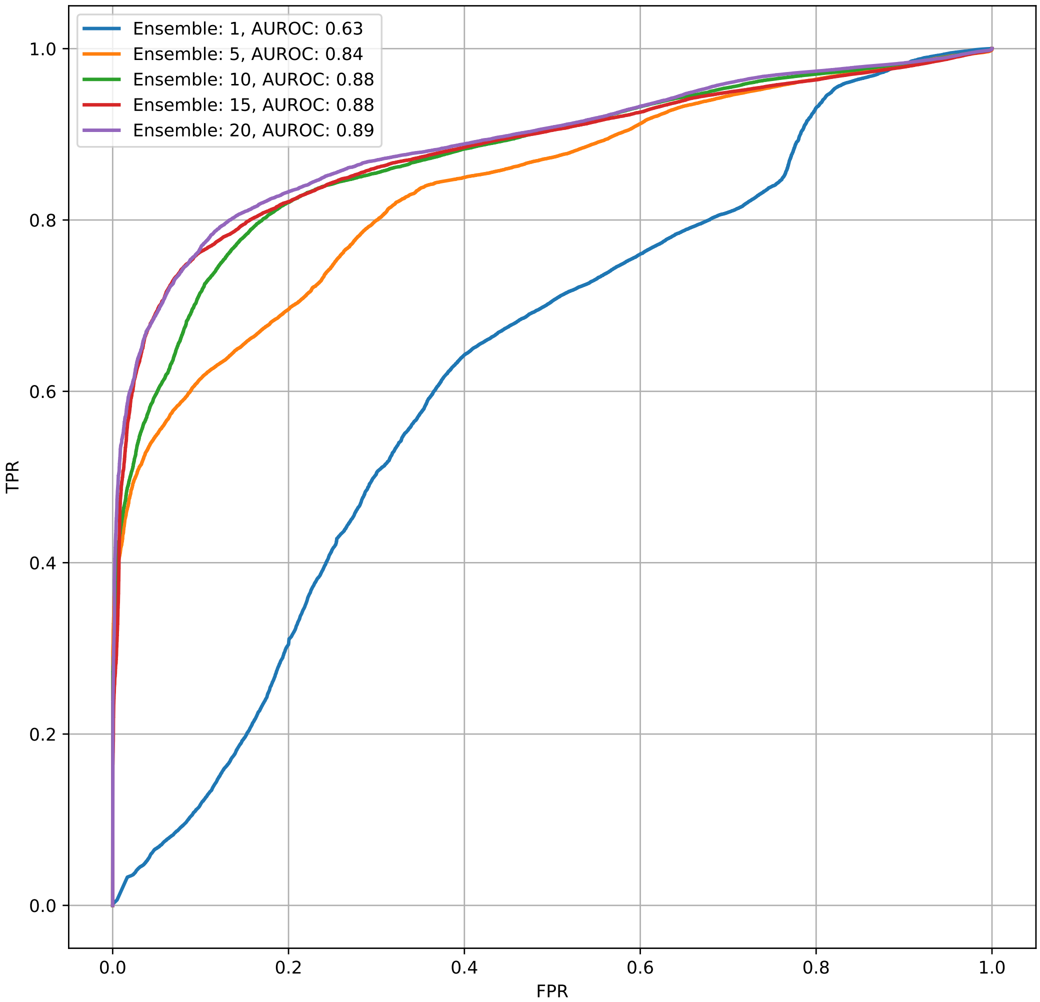
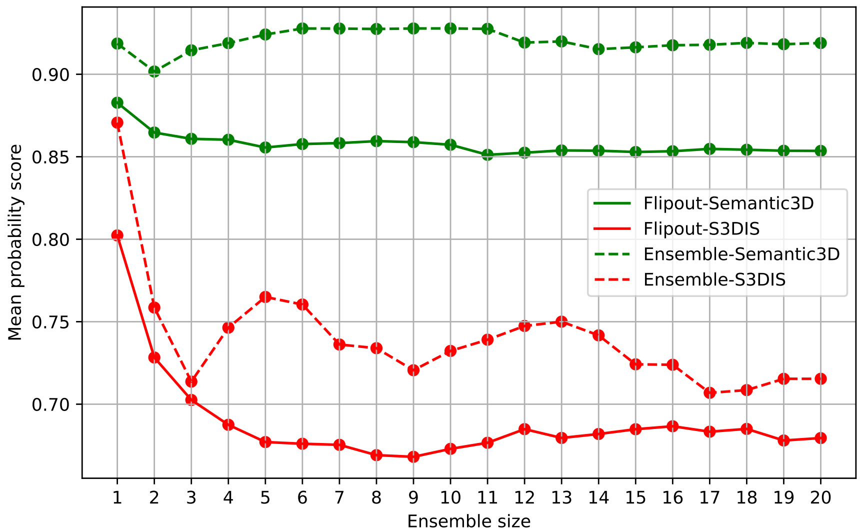
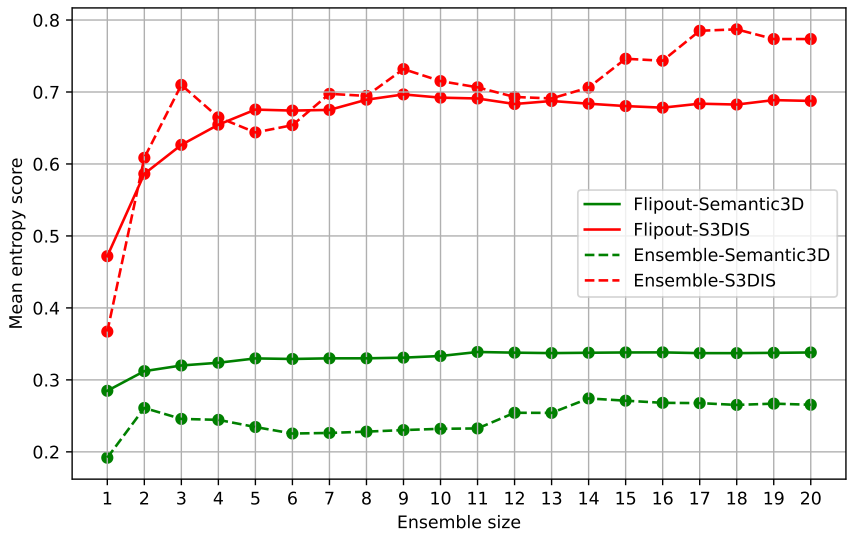


Figure 1.15: ROC plots and AUROC scores for entropy in semantic3D vs S3DIS

1.5 Flipout Vs Ensembles comparison





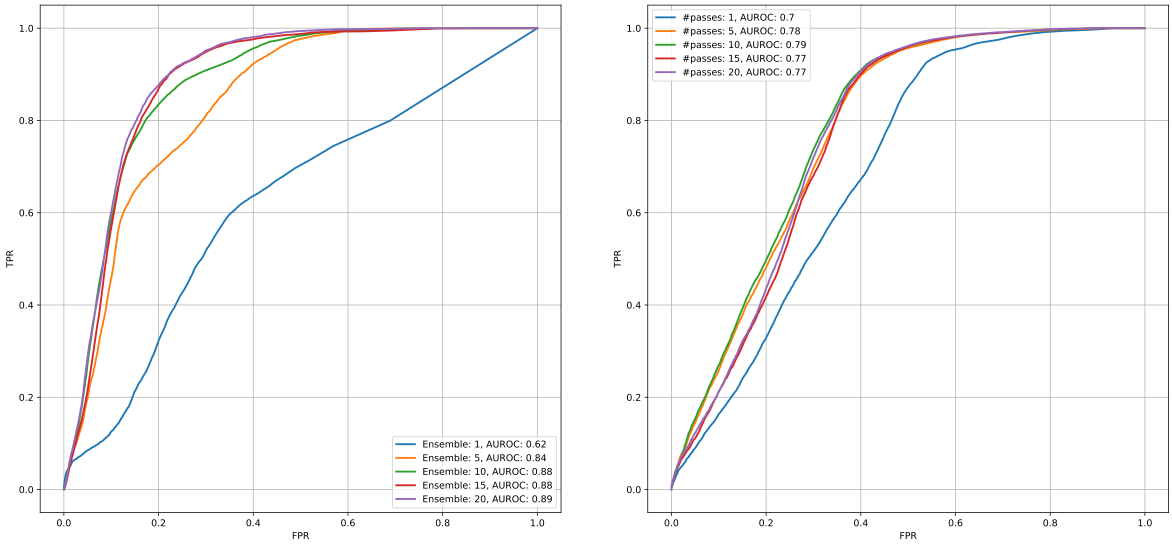


Figure 1.16: Ensmeble Vs Flipout - Probability scores

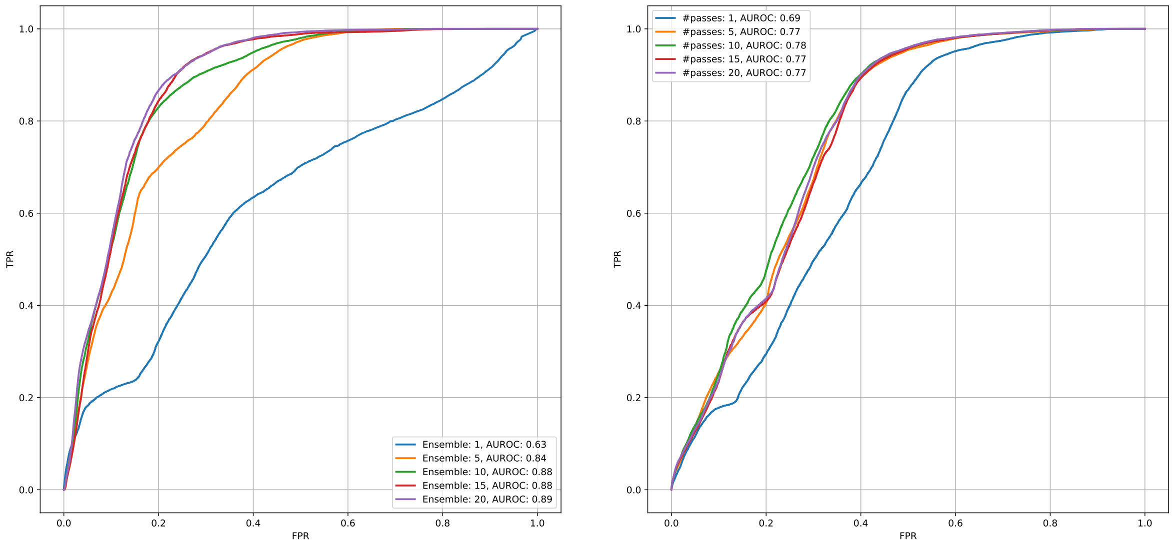


Figure 1.17: Ensemble Vs Flipout - Entropy scores

References

[1] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.