

Point Cloud Segmentation of Agricultural Vehicles using 3D Gaussian Splatting

Supplementary Material

Epochs	PointNet++	PTv3	OACNN
10	0.7608	0.8549	0.8708
20	0.7844	0.8867	0.8802
30	0.7894	0.8911	0.8917

Table 3. Table displays the mIoU of the three different models when trained for three different epoch amounts.

A. Model and Training Tuning

Given the objective of the paper has not been to fine-tune the individual models to obtain the best possible performance, model training efficiency was valued highly when weighing performance against training time for the hyperparameter selection, and for the most part, hyperparameters have been chosen based solely on the model’s original paper implementations [19, 21, 33]. Thus, as a last test it also seemed interesting to observe the performance differences when modifying the base hyperparameters. The tests focused on three key hyperparameters: the number of training epochs, the learning rate and the degree of point cloud downsampling.

A.1. Epoch Tuning

The default number of epoch used throughout the test has been 20, thus when varying the number of epochs it was chosen to test {10, 20, 30} epochs. The results can be found in Table 3. As the results suggest, increasing the number of training epochs helps performance quite notably, and from the training process it seems like the models haven’t quite converged, and thus could improve even further given additional epochs.

A.2. Learning Rate Tuning

The learning rate is usually the most impactful hyperparameter when changed, in this case a sweep was conducted where the base learning rate used for all other model’s tests was scaled by {0.1, 1.0, 10.0}. The results can be seen in Table 4. Interestingly it seems the PTv3 model would benefit from a lowered learning rate, and when scaling it by 10, it would outright crash because of exploding gradients, further suggesting that the PTv3 model should have its learning rate lowered. The results for the OACNN model encourages the opposite, that it should be trained with a higher learning rate.

A.3. Point Cloud Downsampling

To assess the trade-off between computational efficiency and segmentation performance, different point cloud den-

LR scaling	PointNet++	PTv3	OACNN
0.1	0.6876	0.8936	0.8914
1.0	0.7894	0.8867	0.8802
10.0	0.6932	N/A	0.8968

Table 4. Table displays the mIoU of the three different models when trained on three different learning rate scales. The original implementation learning rates were $2e-3$ with batch size 16 for both PointNet++ [21] and OACNN [19] and $5e-3$ for PTv3 with batch size 12 [33]. The conducted tests were run with batch size 32 for all models, consequently the learning rates were scaled proportionally to this as well.

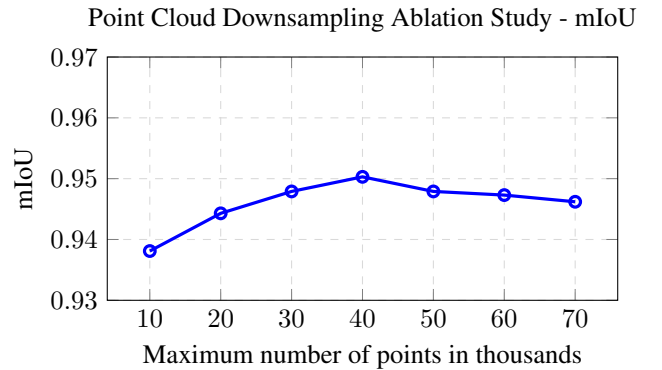


Figure 11. Figure shows the mIoU results for the OACNN model, trained on a real only dataset, with varying amounts of downsampling performed on the input point clouds.

sities were tested. A downsampled point cloud is obtained by performing a uniform random sampling without replacement from the original point cloud. The results can be seen in Figure 11. Mostly due to VRAM usage constraints, point clouds have been downsampled to 30,000 points for PointNet++ and 40,000 points for OACNN and PTv3 for all tests, and as the results display, this downsampling does not hinder the model from sufficiently learning the classwise point cloud representations, unless the point clouds are significantly downsampled.

A.4. Mixed Training Dataset

To establish the effects of synthetic data on the 3D semantic segmentation models, we trained the three different models on a combination of synthetic data and real data, and then compared this to models trained only on real data. This was inspired by Ma et al. [16], Wang et al. [30], and Yue et al. [37]. In total, the training dataset consists of 10,000 point clouds, with a 50/50 split of synthetic and real, as done by

Sample percentage	PointNet++		PTv3		OACNN		Mean increase
	Real only	Mixed	Real only	Mixed	Real only	Mixed	
20%	0.6982	0.7461	0.8369	0.8762	0.8465	0.9247	0.0551
40%	0.7215	0.7942	0.8983	0.9116	0.9191	0.9451	0.0373
60%	0.7406	0.7992	0.9096	0.9301	0.9349	0.9474	0.0305
80%	0.7505	0.7673	0.9065	0.9297	0.9378	0.9498	0.0173
100%	0.7824	0.7894	0.9232	0.9328	0.9403	0.9517	0.0093

Table 5. Table displays the mIoU performance comparison between models trained on real dataset comprised of different amounts of real point clouds, and mixed datasets. The real only column shows baseline models trained on varying amounts of real data. Mean increase displays the average increase seen when using a mixed dataset compared to using only real. The data is presented for each sample percentage and each individual model.

Yue et al. [37]. The synthetic part is composed solely from unique point clouds, while the real part is composed of 20%, 40%, 60%, 80% and 100% of all available unique real point clouds. The real point clouds are over-sampled to match the amount of synthetic point clouds, to avoid biasing the model towards the synthetic data. In total, five different datasets are produced, which individual models are trained upon, each of these datasets also contains the same validation split consisting of 1,200 real point clouds. For evaluation, all models are tested on the test set containing 3,000 real point clouds, outlined in Section 3.2.

The test results evaluating the impact of the mixed dataset training compared to real-only dataset training are presented in Table 5 and Figure 12. Table 5 compares the performance of the models, trained only on real data, to the models which were trained on a combination of real data and synthetic data. Additionally, the mean increase is shown, and it can be seen that the synthetic data improves performance, especially when only a small amount of real data is available. Figure 12 visually illustrates the mIoU performance of the OACNN model, comparing mixed dataset training with real-only dataset training as the number of different real point clouds increases.

B. Combined point clouds with predictions

Selection of combined point clouds with predictions from multiple models are shown in 13. Notice how the baseline model can produce better prediction in the presence of tall grass, due to tall grass not being included in the data simulation.

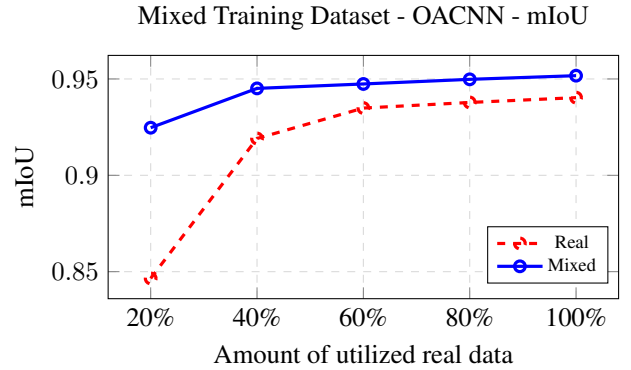


Figure 12. Figure shows the mIoU results for the OACNN model, trained on both a mixed dataset and a real only dataset, consisting of different amounts of different real point clouds.

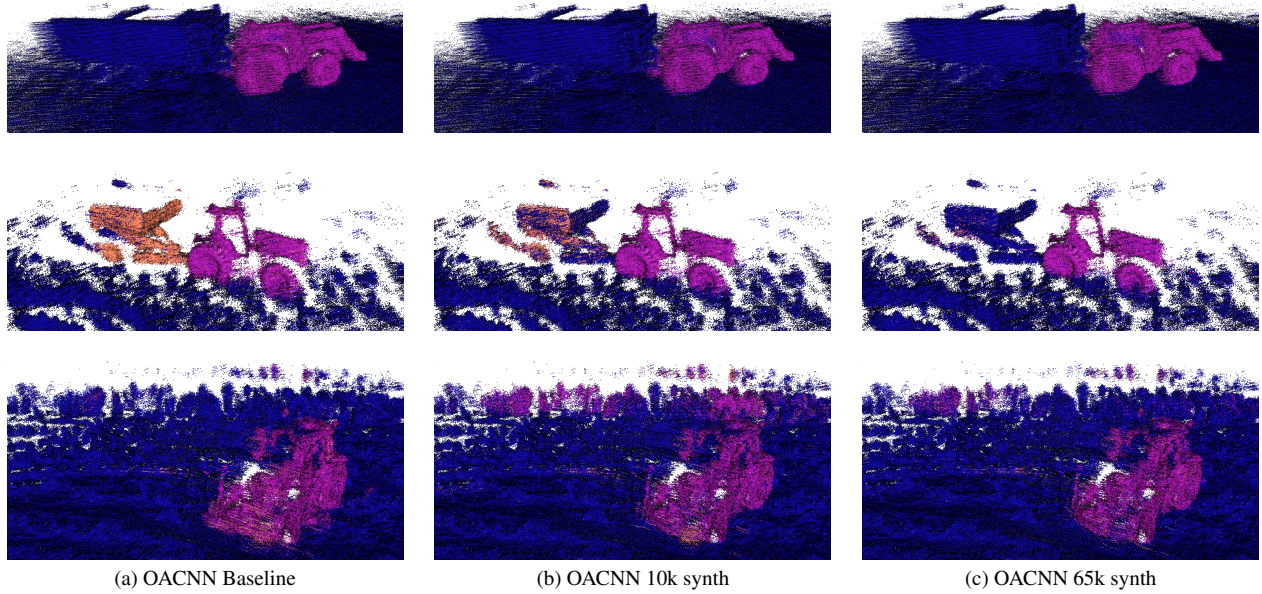


Figure 13. ■ other, ■ tractor, and ■ combine harvester Top row: tractor with large trailer, second row: tractor with small trailer, Third row tractor with tall grass backdrop.