A. Appendix

A.1. Experiment setting

Training details. We provide the details of the training process for different tasks using different optimizers, learning rates, decay rates, and data augmentation techniques. For the classification and part segmentation, we used the stochastic gradient descent (SGD) optimizer with momentum, weight decay, and initial learning rate set to 0.9, 0.0001, and 0.001, respectively. We trained the model for 260 epochs with a learning rate decay rate of 0.7 for every 20 iterations. For semantic segmentation on S3DIS, we utilized the AdamW optimizer with a β value of 0.9, weight decay of 0.001, and initial learning rate of 0.005. We trained the model for around 600 epochs with a learning rate decay rate of 0.1 at the 360th and 480th epochs. Similarly, for semantic segmentation on ScanNetV2, we used the AdamW optimizer with the same hyperparameters as above, but we trained the model for around 600 epochs with a cosine annealing learning rate decay schedule. For object detection, we utilized 50,000 points, which were augmented using random flip, rotation, and scaling techniques. We used the AdamW optimizer for training with a β value of 0.9, weight decay of $5e^{-4}$, and an initial learning rate of 0.006. We trained the model for 400 epochs with a learning rate decay rate of 0.1 at the 280th to 340th epoch. Additionally, the learning rate of the attention modules in the Group-Free method [35] was set to $\frac{1}{10}$ of the backbone setting. Table 9 provides the data augmentation setting on five datasets.

Dataset	Rotation	Scaling	Fliping	Jittering	Color
ModelNet40		✔			
ScanObecjectNN	✔	✔			
ShapeNetPart	✔	✔	✔		
S3DIS		✔	✔	✔	✔
ScanNetV2	✔	✔	✔	✔	✔

Table 9. Data augmentation on tasks

A.2. Computation FLOPs and parameter sizes

Table 10 provides the parameter sizes and flops for ADS and other approaches.

A.3. Representative feature of the cluster

We carry out experiments to assess the quality of different features. Our results, presented in Table 11, indicate that the global feature of clusters performs better than the feature of the cluster center. This suggests that the global feature provides a more comprehensive representation of the cluster than the cluster center feature.

A.4. Better augmentation on ScanObjectNN

In the main text, we have pointed out that the optimization settings of PointNeXt-S [33] could vary with respect to different tasks. However, we apply the same training setting (blue row of the table below) of ADS for all the tasks. As illustrated in Table 12, if we use the optimal augmentation strategy of [33] tuned for ScanObjectNN, the resulting ADS (denoted as ADS*) could also achieve comparable results.

Method	#Params	FLOPs
Set abstraction based		
PointNet [30]	1.700M	4.10G
PointNet++ [32]	0.969M	1.00G
RepSurf-U [35]	0.976M	1.04G
Attention based		
PointTransformer [59]	4.90M	2.80G
PointASNL [51]	22.40M	19.10G
RPNet-W9 [36]	2.40M	5.10G
ADS	4.86M	2.93G

Table 10. Ablation of the computation.

Feature type	ModelNet40 OA (%)
Center feature of Cluster	94.7
Global feature of Cluster	**95.1**

Table 11. Ablation of the Cluster Representative Feature.

A.5. Part segmentation on ShapeNetPart

We further provide the part segmentation results of all categories on ShapeNetPart. As shown in Table 13, the proposed ADS achieves top mIoU results for 12 categories without adopting any class-balancing loss.

A.6. The influence of variation of K

It is worthy to note that the variation of K among objects within the same category is not sensitive in our approach. This is because we purposedly choose a relatively small bandwidth in the mean-shift clustering process to ensure computational efficiency. The primary purpose of the mean-shift clustering is to provide a rough segmentation, and any redundancy in the clusters is subsequently reduced through the self-attention process.

A.7. Qualitative visual comparison

We present the visual comparison results of the proposed ADS method with the standard PointNet++ model on two different datasets: ShapeNetPart and S3DIS. Figure 8 shows the visualization of the sample quality comparison between ADS and PointNet++ with FPS sampling strategy on ShapeNetPart dataset. On the other hand, Figure 9 exhibits the results of dense semantic segmentation on 6-fold S3DIS. Figure 9 displays comprehensive results of our proposed model on

Method	Resampling	Rotation	Scaling	Jittering	Smoothing	Appending	OA (%)	mAcc (%)
PointNetXt-S [33]	✔	✔	✔		✔	✔	**87.7**	85.8
ADS*	✔	✔	✔		✔	✔	**87.7**	**85.9**
ADS		✔	✔	✔			87.5	85.1

Table 12. Better augmentation strategy [33] on ScanObjectNN.

# Shapes		2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271
Method	**mIoU** (%)	**Aero**	**Bag**	**Cap**	**Car**	**Chair**	**Phone**	**Guitar**	**Knife**	**Lamp**	**Laptop**	**Motor**	**Mug**	**Pistol**	**Rocket**	**Skate**	**Table**
SyncSpecCNN [55]	81.4	81.0	78.4	77.7	75.7	87.6	61.9	92.0	85.4	82.5	95.7	70.6	91.9	**85.9**	53.1	69.8	75.3
PointNet [30]	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6
PointCNN [20]	84.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RSNet [16]	84.9	82.7	86.4	84.1	78.2	90.4	69.3	91.4	87.0	83.5	95.4	66.0	92.6	81.8	56.1	75.8	82.2
PointNet++ [32]	85.1	84.0	83.4	86.7	77.8	90.6	74.7	91.2	87.5	82.8	95.7	66.3	94.9	81.1	63.5	74.5	82.6
DGCNN [44]	85.1	84.0	83.4	86.7	77.8	90.6	74.7	91.2	87.5	82.8	95.7	66.3	94.9	81.1	63.5	74.5	82.6
Tsai [40]	85.1	82.3	83.5	84.5	77.3	89.8	76.3	91.0	87.3	84.2	95.5	67.8	92.5	82.8	52.1	73.9	83.5
SpiderCNN [50]	85.3	83.5	81.0	87.2	77.5	90.7	76.8	91.1	87.3	83.3	95.8	70.2	93.5	82.7	59.7	75.8	82.8
Point2Sequence [40]	85.2	82.6	81.8	87.5	77.3	90.8	77.1	91.1	86.9	83.9	95.7	70.8	94.6	79.3	58.1	75.2	82.8
PointCNN [20]	86.1	84.1	86.5	86.0	**80.8**	90.6	79.7	92.3	88.4	85.3	96.1	77.2	95.2	84.2	64.2	80.0	83.0
PointASNL [51]	86.1	84.1	84.7	87.9	79.7	92.2	73.7	91.0	87.2	84.2	95.8	74.4	95.2	81.0	63.0	76.3	83.2
PointMLP [26]	86.1	83.5	83.4	87.5	80.5	90.3	78.2	92.2	88.1	82.6	**96.2**	**77.5**	95.8	85.4	64.6	**83.3**	84.3
RS-CNN [22]	86.2	83.5	84.8	**88.8**	79.6	91.2	81.1	91.6	88.4	86.0	96.0	73.7	94.1	83.4	60.5	77.7	83.6
PointTransformer [59]	86.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PointNeXt-S [33]	86.7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ADS	**86.9**	**84.6**	**86.9**	**88.8**	80.5	**92.5**	**82.2**	**92.8**	**90.1**	**86.3**	**96.2**	76.8	**95.8**	83.5	**66.3**	82.5	**84.4**

Table 13. Part segmentation performance with mean part-averaged IoU on ShapeNetPart.

the 6-fold S3DIS dataset, further demonstrating its accuracy and effectiveness in point label segmentation.

B. Limitation

While the proposed ADS is shown to effectively achieve good performances for comprehensive 3-D tasks, the various attention modules within its design could result in a more complex model. As highlighted in Table 10, the ADS model requires larger numbers of FLOPs and parameters compared to the current state-of-the-art approach based on set abstraction. However, despite this added complexity, the ADS model is still able to compete in terms of FLOPs and parameter size with other attention-based methods while offering improved performances.

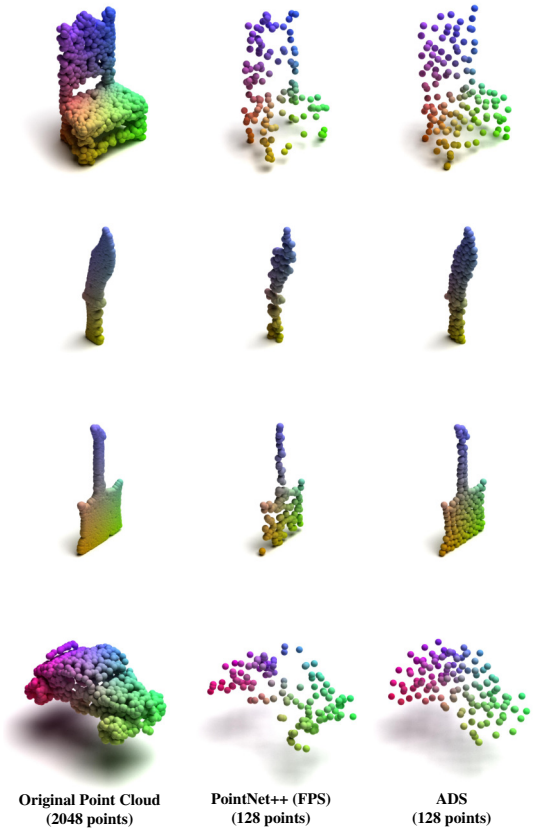

Figure 8. **Qualitative comparison on ShapeNetPart.** The visual results demonstrate the proposed method and basic PointNet++ on a dense Part segmentation example.

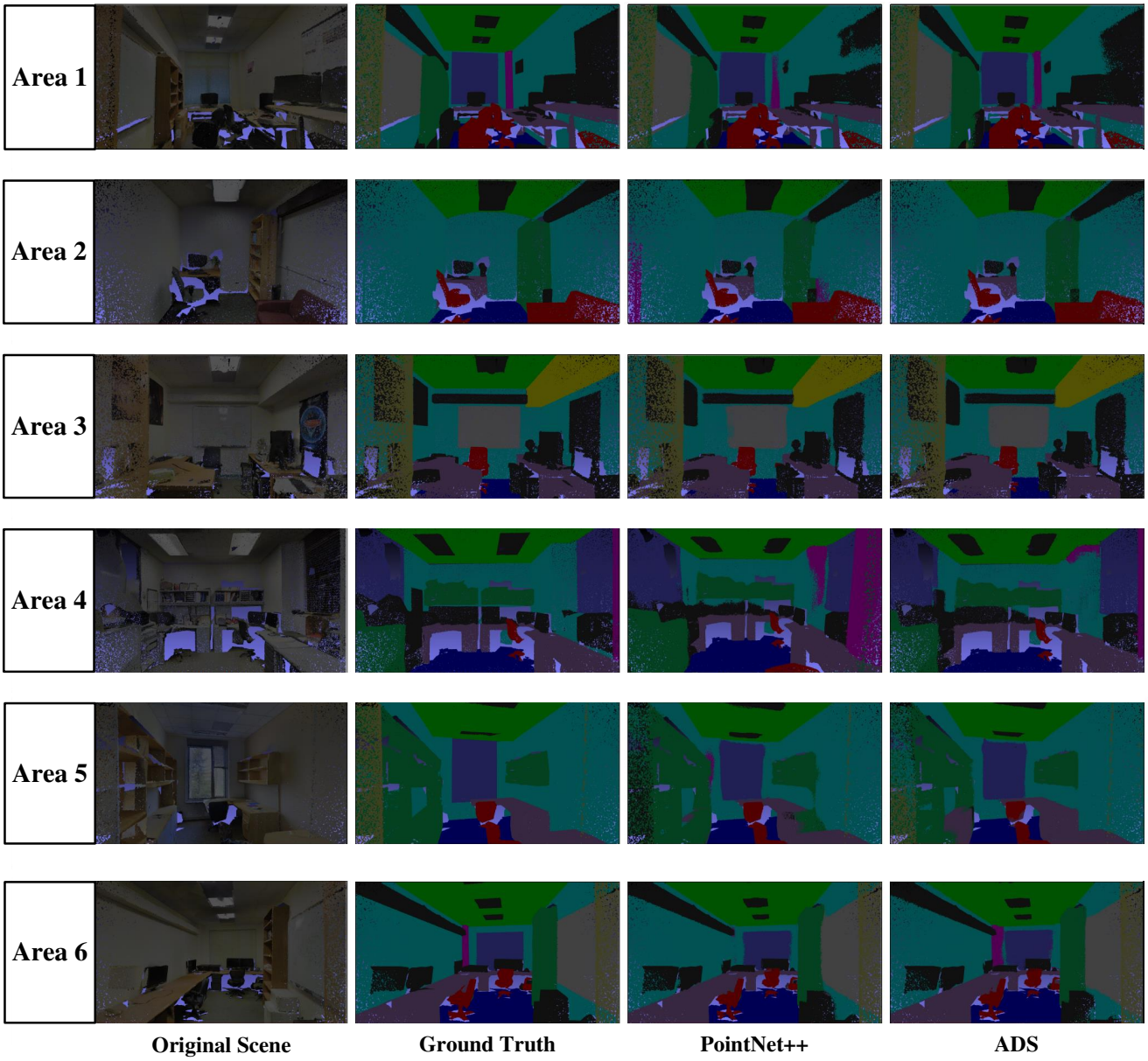

Figure 9. **Qualitative comparison on 6-fold of S3DIS.** The visual results demonstrate the proposed method and basic PointNet++ on a dense semantic segmentation example.