

PartNet: A Recursive Part Decomposition Network for Fine-grained and Hierarchical Shape Segmentation – Supplemental Material

Fenggen Yu^{1*} Kun Liu^{1*} Yan Zhang¹ Chenyang Zhu² Kai Xu^{2†}
¹Nanjing University ²National University of Defense Technology

1. Introduction

This supplemental material contains eight parts:

- Section 2 provides an overview and more statistics of our FineSeg benchmark dataset.
- Section 3 gives extended results of the ablation study (Figure 7 in the main paper).
- Section 4 reports the details of the subjective study of our fine-grained segmentation in the ShapeNet challenge (see Section 4.2 in the main paper).
- Section 5 gives more details of the network in the application of fine-grained structure driven image-to-shape reconstruction (Section 5 in the main paper). Meanwhile, we also demonstrate more results of this application.
- Section 6 shows the effect of hierarchy construction on the performance of fine-grained segmentation.
- Section 7 discusses complexity and timing of PartNet.
- Section 8 shows comparison of semantic segmentation on the Princeton Segmentation Benchmark [2].
- Section 9 provides a visual comparison of semantic segmentation against two non-learned hierarchical segmentation methods.
- Section 10 shows more results of fine-grained hierarchical segmentation on the FineSeg benchmark dataset.
- Section 11 shows some sample results of semantic segmentation on the ShapeNet part dataset.

	aero	bike	chair	heli.	sofa	table
# shapes	600	140	1000	100	600	500
# avg. parts	8	7	10	15	9	6

Table 1. The detailed statistics of our FineSeg benchmark dataset.

2. FineSeg benchmark dataset

Table 1 reports the detailed statistics of our FineSeg benchmark dataset. For each shape category, we list the total number of shapes and the average per-shape part count.

FineSeg dataset contains two versions, one *clean set* and one *noisy set*. The clean set is obtained simply by point sampling the 3D surface models. The noisy set is created by adding a moderate level of random noise to the clean point clouds, which can be used to test the robustness of segmentation against noisy input. Specifically, we use Gaussian noise with the bandwidth being 3% of the diagonal length of shape bounding box. In Section 10, we show extended results of fine-grained segmentation on both the clean and noisy sets.

3. More results of ablation study

Figure 1 gives extended results of Figure 7 in the main paper, on the remaining three categories, i.e., Chair, Helicopter and Table.

4. User study for ShapeNet challenge

We conducted a preliminary user study for the ShapeNet Segmentation Challenge. Specifically, we evaluate the fine-grained segmentation obtained PartNet on randomly picked shapes from ShapeNet. In the study, each user is presented

	aero	bike	chair	table
Ours	4.24	4.66	4.15	4.00
Baseline	4.07	4.03	3.80	3.54

Table 2. Average user ratings for the segmentations obtained by our and the baseline (w/o recursive context feature), on four shape categories.

*Joint first authors

†Corresponding author: kevin.kai.xu@gmail.com

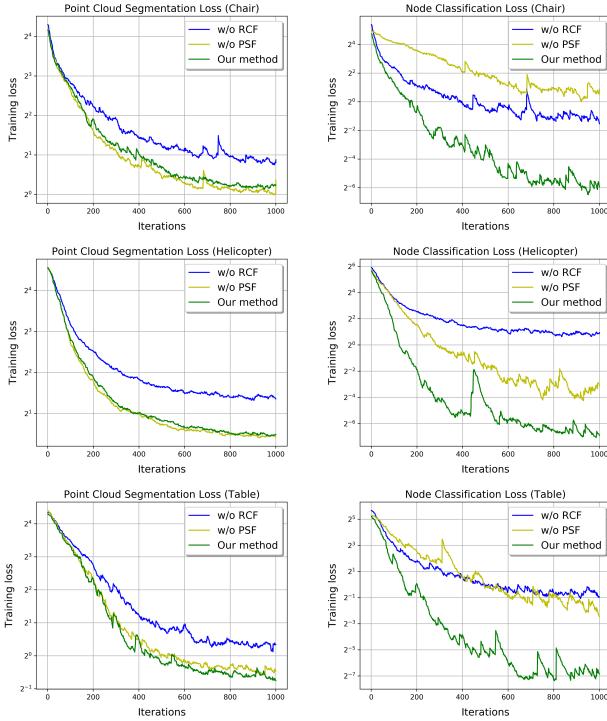


Figure 1. Plots of training loss over iterations for the ablation study of the two key node features (RCF and PSF) on three categories (Chair, Helicopter and Table). For each category, we plot both node segmentation loss (left column) and node classification loss (right column).

with 30 queries. Each query first shows a 3D shape (in surface rendering), and then a fine-grained segmented point cloud of the shape. The user is asked to rate the quality of the segmentation in the range from 1 (the poorest) to 5 (the best).

A total of 12 subjects participated in the study; these subjects are all graduate students from various research directions in computer science. Results are demonstrated in Table 2, where the average ratings of segmentations are reported. To contrast, we also report the average rating of the segmentations obtained by a baseline method, i.e., our method with the recursive context feature (RCF) disabled. The segmentation results by the full and baseline methods are shown in a random order. The results show that the segmentations by our full method were generally more favored by the human subjects.

5. The application network

The part refiner network used at each leaf node is composed of two channels of PointNet, to encode the point clouds of the part and the full shape, respectively. The resulting two features are concatenated and fed into a four layer fully-connected networks to generate a refined part

point cloud. To train this refiner network, we use reconstruction loss computed as the Chamfer distance and the earth mover’s distance between point clouds [3]. To gain more training signals, we opt to train the refiner with a hierarchical reconstruction loss, through a bottom-up composition of the refined part point clouds, following the hierarchy obtained by PartNet segmentation. This way, we can compute a reconstruction loss at each node of the hierarchy, with the corresponding point cloud composed from the part point clouds within its subtree.

Figure 2 shows the architecture of our application network for fine-grained structure driven image-to-shape reconstruction. Figure 4~6 gives more results of reconstruction refinement by this method.

6. Effect of hierarchy construction

In the main paper, we mentioned that we build for each training shape a part hierarchy using the heuristic method proposed in [5]. To evaluate the effect of hierarchy construction on the performance of fine-grained segmentation, we compare in Table 3 the segmentation accuracy of PartNet trained with pre-built and random part hierarchies, respectively. The results show that our method works better with pre-built hierarchies.

		mean	aero	bike	chair	heli.	sofa	table
IoU > 0.25	Random	58.7	80.4	63.3	56.3	53.5	36.4	62.4
	Pre-build	84.8	95.2	97.0	91.1	83.0	65.4	77.2
IoU > 0.5	Random	30.2	62.4	21.9	24.7	29.0	7.2	35.7
	Pre-build	72.8	88.0	89.4	80.5	69.4	46.7	62.6

Table 3. Comparing the average precision (AP) of fine-grained segmentation for our method with pre-built hierarchies (using the method in [5]) and with random hierarchies. AP (%) is measured with IOU threshold being 0.25 and 0.5 respectively.

7. Timing and complexity

The training time of PartNet for fine-grained segmentation is given in Table 4. The training is conducted on a NVIDIA GTX 1080Ti GPU. For fine-grained segmentation, PartNet test takes less than 1 second to process one 3D shape; the computational time is proportional to the depth of the hierarchy, thus the complexity is $O(N \log M)$ with N being the total number of points and M the target number of fine-grained parts. When doing semantic segmentation, the computational cost is low since the target number of semantic parts is usually small (< 10); such time is comparable to the existing methods.

	aero	bike	chair	heli.	sofa	table
Times (Hour)	10	2.5	21.0	2.2	11.1	7.6

Table 4. Training time of PartNet on different shape categories of the FintSeg benchmark.

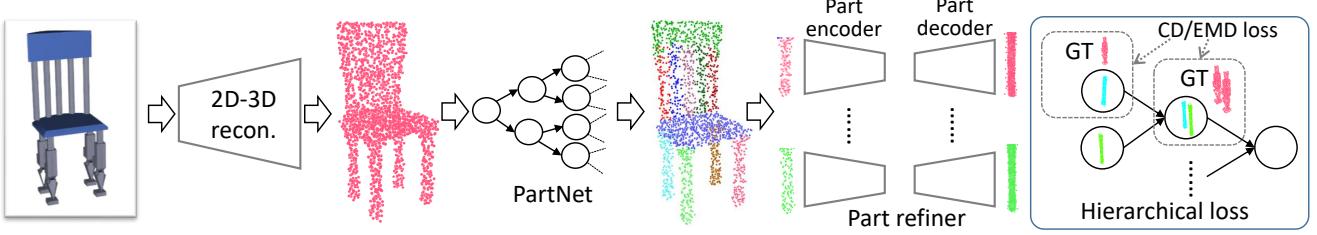


Figure 2. Architecture of fine-grained structure driven image-to-shape reconstruction. Given an image, we first reconstruct a 3D point cloud using the method in [3]. The point cloud is lack of detailed structures. To this end, we use PartNet to decompose the point cloud into fine-grained parts. Each of the decomposed parts is then passed through a part refiner which is an auto-encoder. The auto-encoder outputs a refined point cloud for the corresponding part. To train the part refiner, we use the ground-truth parts of the training 3D models, based on the CD/EMD loss between point clouds [3]. To gain more training signals, we opt to train the refiner with a hierarchical reconstruction loss, through a bottom-up composition of the refined part point clouds, following the hierarchy obtained by PartNet segmentation. This way, we can compute a reconstruction loss at each node of the hierarchy, where the corresponding point cloud is composed from the part point clouds within its subtree.

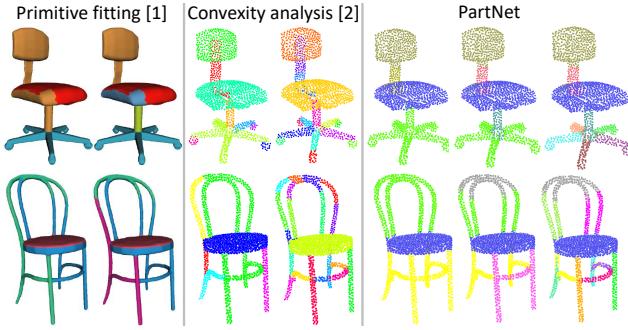


Figure 3. A qualitative comparison of segmentation granularity of different hierarchical methods. PartNet achieves more meaningful segmentation for varying number of parts, compared to [1, 4].

8. Comparison on the Princeton Segmentation Benchmark

Princeton Segmentation Benchmark [2] is used for evaluating traditional segmentation methods. It is much smaller (containing 380 models in 19 categories) and thus inadequate for training deep learning models. Nevertheless, it is interesting to see how deep models trained with ShapeNet perform on PSB. Table 5 provides a comparison over PSB, which again demonstrates the advantage of our method. Note that only four sets, i.e., aeroplane, chair, cup and table, are tested, because these are the only categories that are shared by ShapeNet and PSB. We would include this table in the revision.

9. Visual comparison of semantic segmentation

Figure 3 provides a qualitative comparison of hierarchical segmentation granularity between our method and two baselines [1, 4]. We would add such qualitative comparison in the revision if requested.

Method	mean	aero	chair	cup	table
PointNet	89.6	82.0	97.0	84.6	94.9
PointNet++	92.2	84.4	95.9	90.8	97.6
O-CNN	90.3	76.6	97.2	89.7	97.6
SSCN	90.2	73.8	93.0	98.1	96.0
PCNN	91.9	83.2	97.5	89.6	97.2
SPLATNet	92.0	83.3	97.4	91.1	96.1
PointCNN	92.0	84.1	96.5	89.4	97.9
Ours	93.1	85.5	96.7	92.4	97.9

Table 5. Comparison of semantic segmentation on four categories of Princeton Segmentation Benchmark (shapes are point sampled). Metric is part-wise IoU (%).

		mean	aero	bike	chair	hel. sofa	table
IoU	Clean	84.8	95.2	97.0	91.1	83.0	65.4
> 0.25	Noisy	83.8	95.1	96.7	90.2	81.6	62.4
IoU	Clean	72.8	88.0	89.4	80.5	69.4	46.7
> 0.5	Noisy	69.4	83.7	88.5	77.7	64.5	40.8

Table 6. Comparing the average precision (AP) of fine-grained segmentation on the clean and noisy datasets. AP (%) is measured with IOU threshold being 0.25 and 0.5, respectively.

10. More results of fine-grained segmentation

Figure 7~12 demonstrate the extended results of fine-grained segmentation on the clean set of FineSeg, while Figure 13 and 14 show the results of that on the noisy set. From the results in Figure 13 and 14, our method is robust against random noise in the point clouds. Table 6 further compares the average precision (AP) of segmentation on clean and noisy sets. It can be seen that the performance drop is insignificant under a moderate level of random noise.

11. Visual results of semantic segmentation

From Figure 15 to Figure 30, we show visual examples of semantic segmentation on shapes from the ShapeNet part

dataset [6]. The quantitative evaluation can be found in the main paper.

References

- [1] M. Attene, B. Falcidieno, and M. Spagnuolo. Hierarchical mesh segmentation based on fitting primitives. *The Visual Computer*, 22(3):181–193, 2006. 3
- [2] X. Chen, A. Golovinskiy, and T. Funkhouser. A benchmark for 3D mesh segmentation. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 28(3), Aug. 2009. 1, 3
- [3] H. Fan, H. Su, and L. Guibas. A point set generation network for 3d object reconstruction from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2463–2471, 2017. 2, 3
- [4] O. van Kaick, N. Fish, Y. Kleiman, S. Asafi, and D. Cohen-Or. Shape segmentation by approximate convexity analysis. *ACM Trans. on Graph.*, 34(1):4, 2014. 3
- [5] Y. Wang, K. Xu, J. Li, H. Zhang, A. Shamir, L. Liu, Z. Cheng, and Y. Xiong. Symmetry hierarchy of man-made objects. *Computer Graphics Forum*, 30(2):287–296, 2011. 2
- [6] L. Yi, V. G. Kim, D. Ceylan, I.-C. Shen, M. Yan, H. Su, C. Lu, Q. Huang, A. Sheffer, and L. Guibas. A scalable active framework for region annotation in 3d shape collections. *SIGGRAPH Asia*, 2016. 4



Figure 4. A few chair examples on refining point clouds reconstructed from single view images, guided by the fine-grained segmentation of PartNet. In each row, we show from left to right the input image, result of holistic reconstruction, fine-grained segmentation of the reconstruction, and the final refinement result by our method.

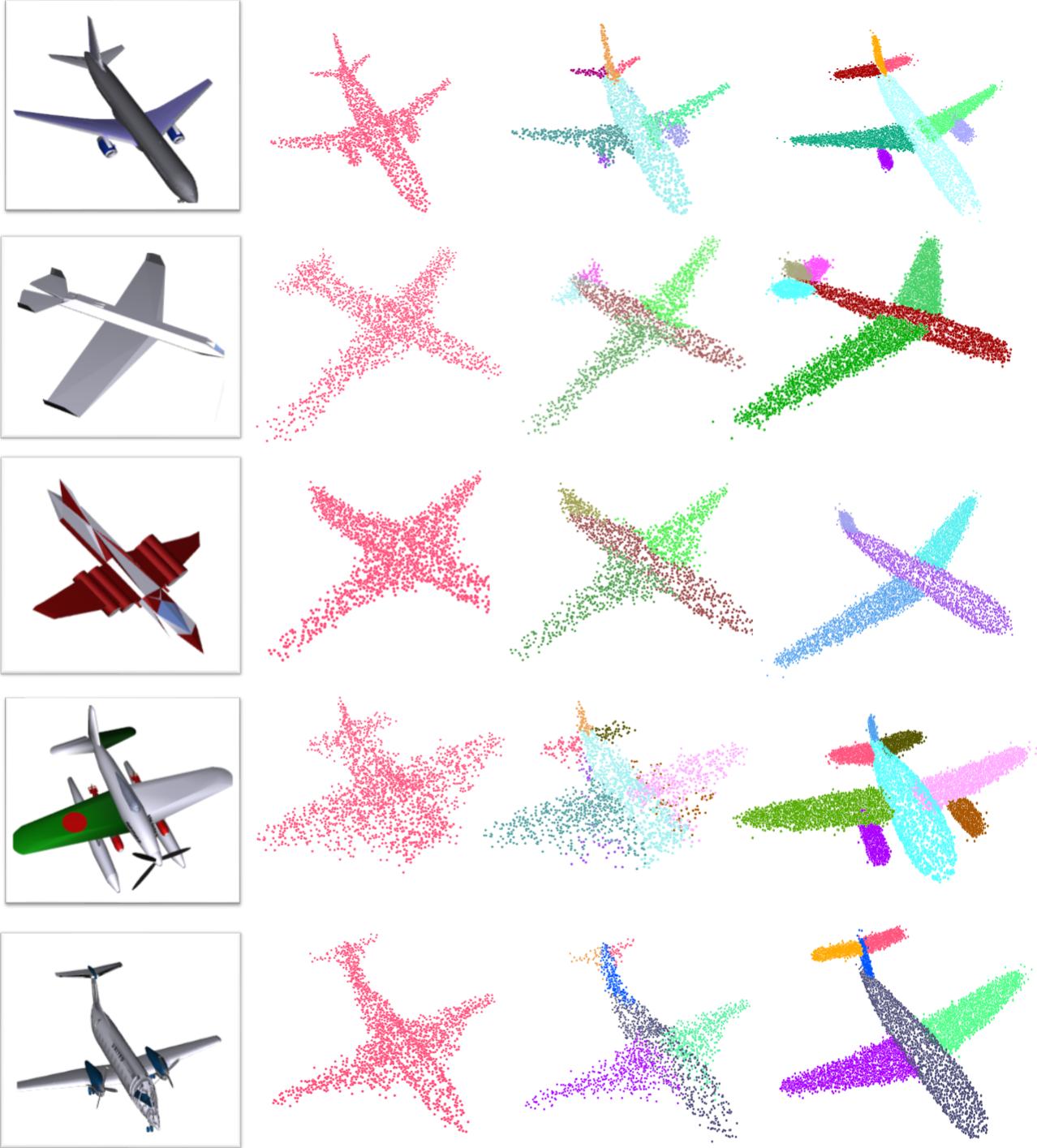


Figure 5. A few airplane examples on refining point clouds reconstructed from single view images, guided by the fine-grained segmentation of PartNet. In each row, we show from left to right the input image, result of holistic reconstruction, fine-grained segmentation of the reconstruction, and the final refinement result by our method.



Figure 6. A few bike examples on refining point clouds reconstructed from single view images, guided by the fine-grained segmentation of PartNet. In each row, we show from left to right the input image, result of holistic reconstruction, fine-grained segmentation of the reconstruction, and the final refinement result by our method.

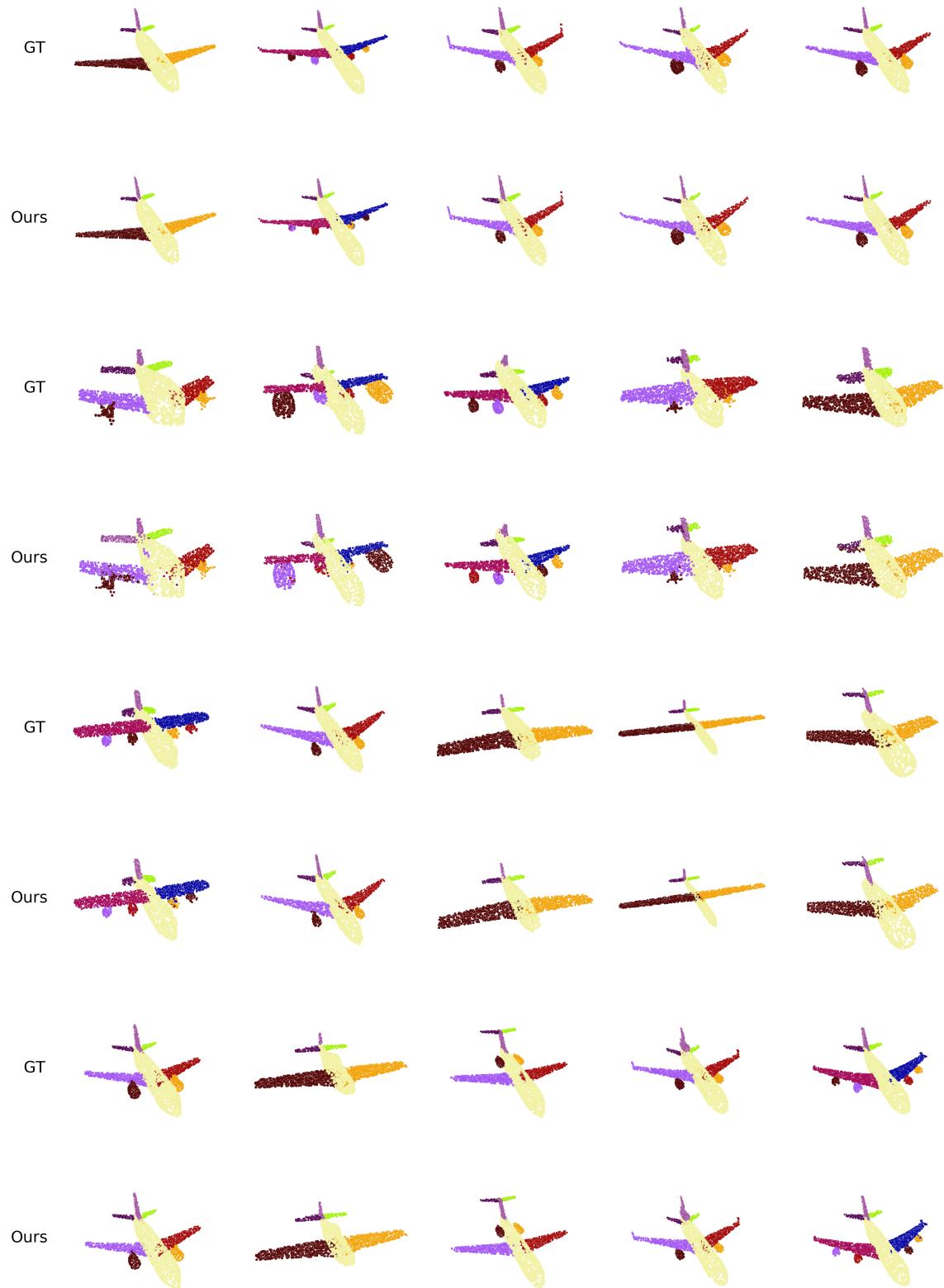


Figure 7. Fine-grained point cloud segmentation on clean dataset (Airplane). For comparison, we show for each shape the fine-grained segmentation result (bottom) and the corresponding ground-truth (top).



Figure 8. Fine-grained point cloud segmentation on clean dataset (Bike). For comparison, we show for each shape the fine-grained segmentation result (bottom) and the corresponding ground-truth (top).

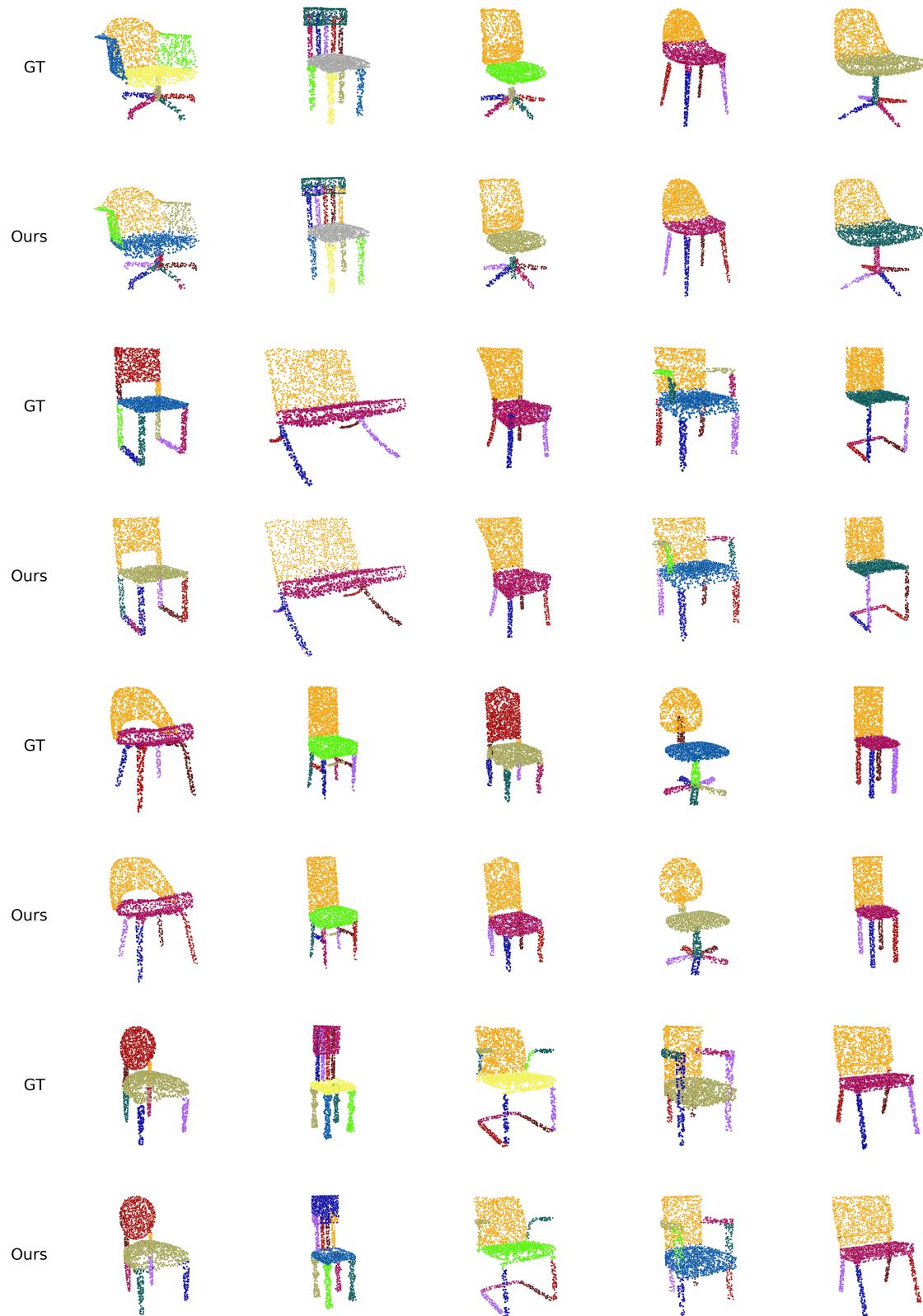


Figure 9. Fine-grained point cloud segmentation on clean dataset (Chair). For comparison, we show for each shape the fine-grained segmentation result (bottom) and the corresponding ground-truth (top).

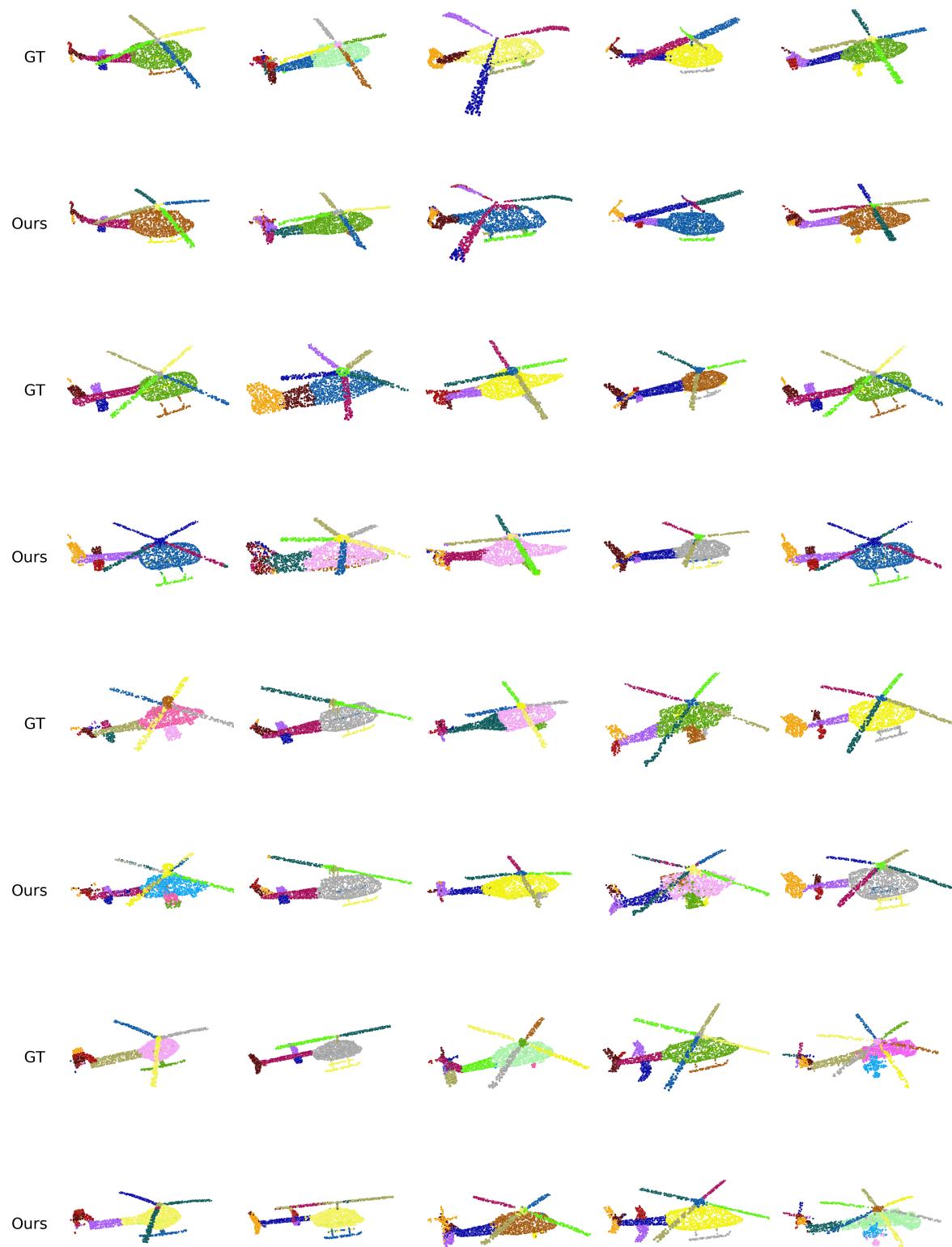


Figure 10. Fine-grained point cloud segmentation on clean dataset (Helicopter). For comparison, we show for each shape the fine-grained segmentation result (bottom) and the corresponding ground-truth (top).

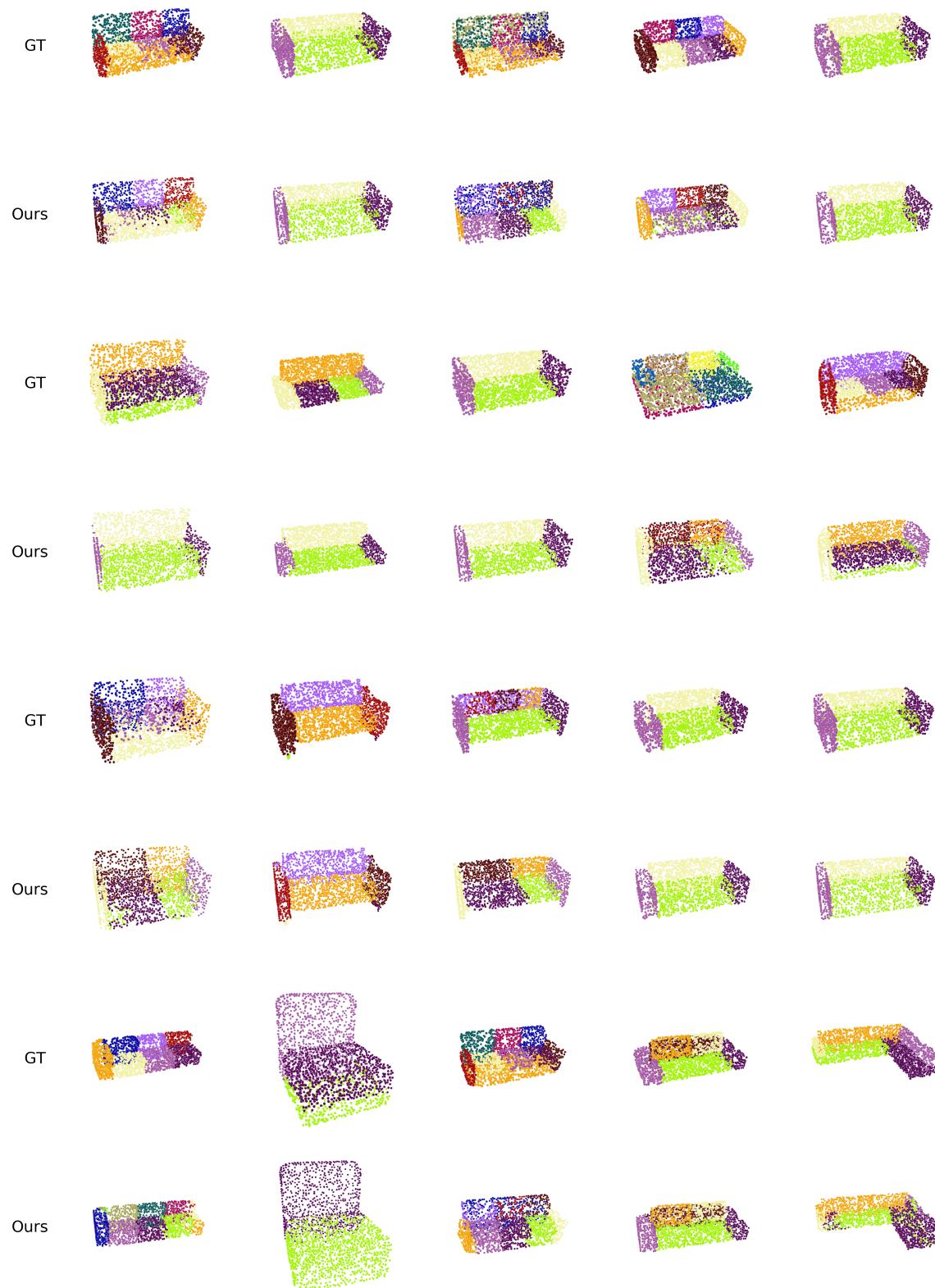


Figure 11. Fine-grained point cloud segmentation on clean dataset (Sofa). For comparison, we show for each shape the fine-grained segmentation result (bottom) and the corresponding ground-truth (top).

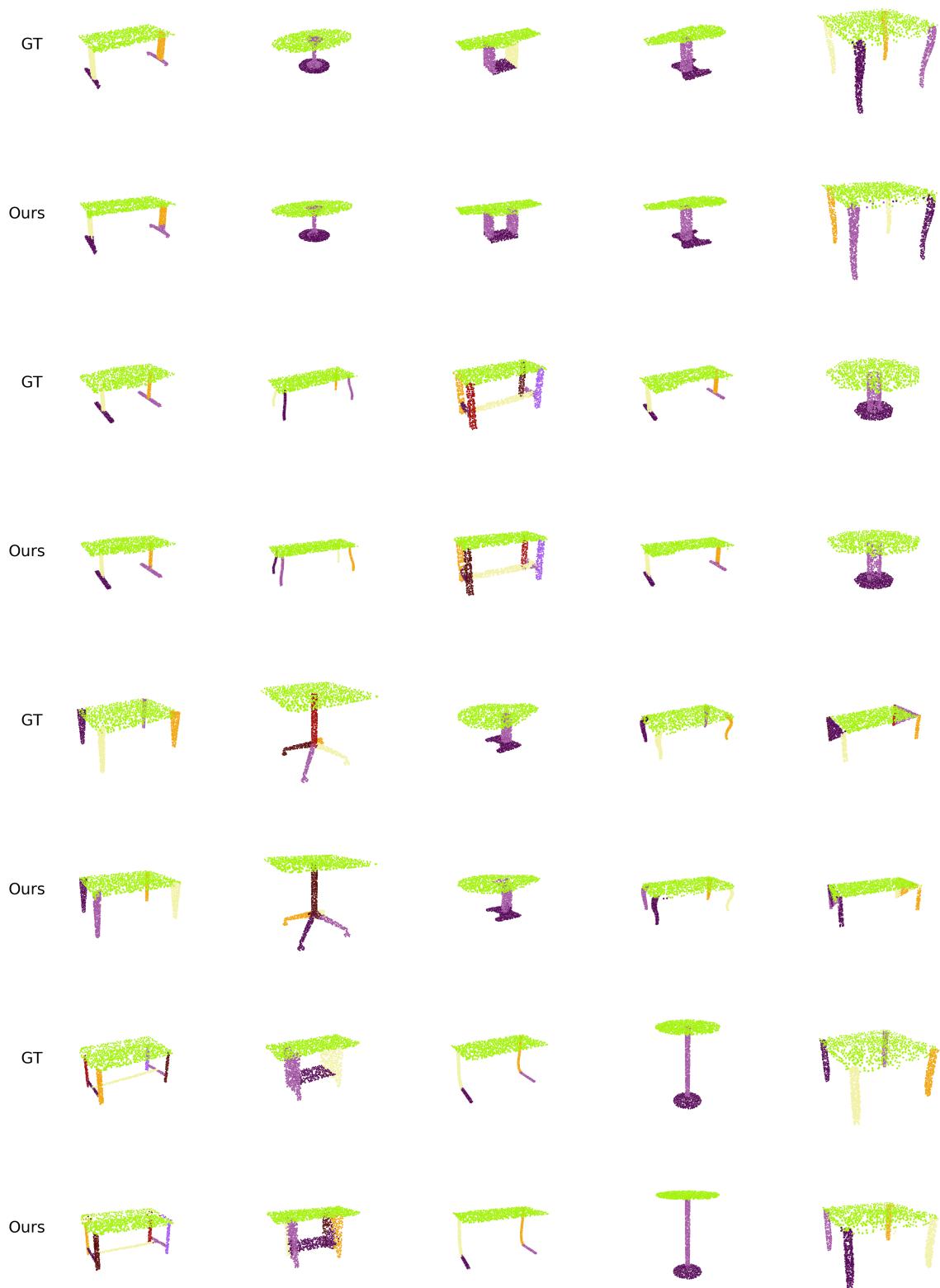


Figure 12. Fine-grained point cloud segmentation on clean dataset (Table). For comparison, we show for each shape the fine-grained segmentation result (bottom) and the corresponding ground-truth (top).



Figure 13. Fine-grained point cloud segmentation on noisy dataset (Airplane, Bike and Chair). For comparison, we show for each shape the fine-grained segmentation result (bottom) and the corresponding ground-truth (top).

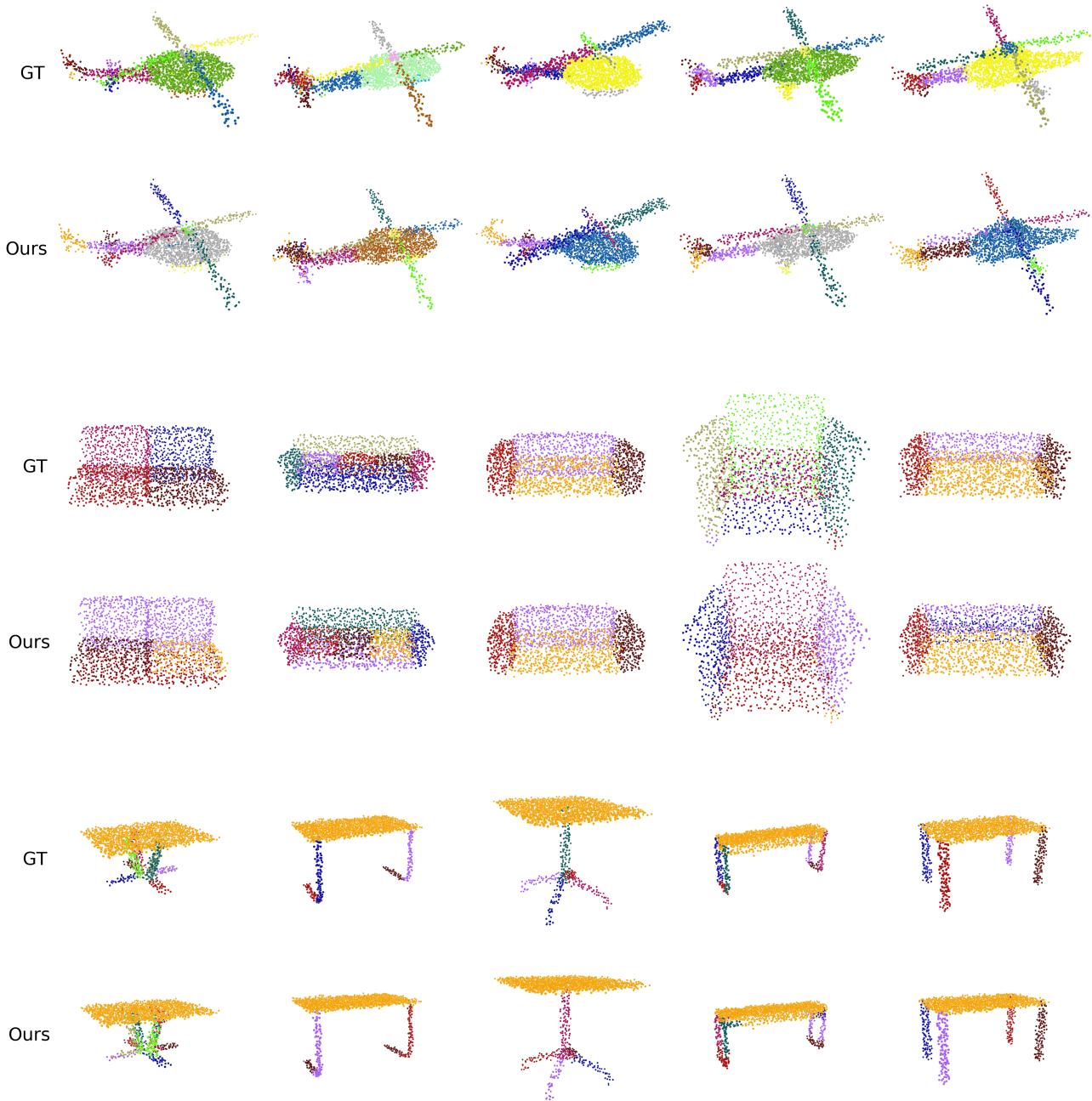


Figure 14. Fine-grained point cloud segmentation on noisy dataset (Helicopter, Sofa and Table). For comparison, we show for each shape the fine-grained segmentation result (bottom) and the corresponding ground-truth (top).

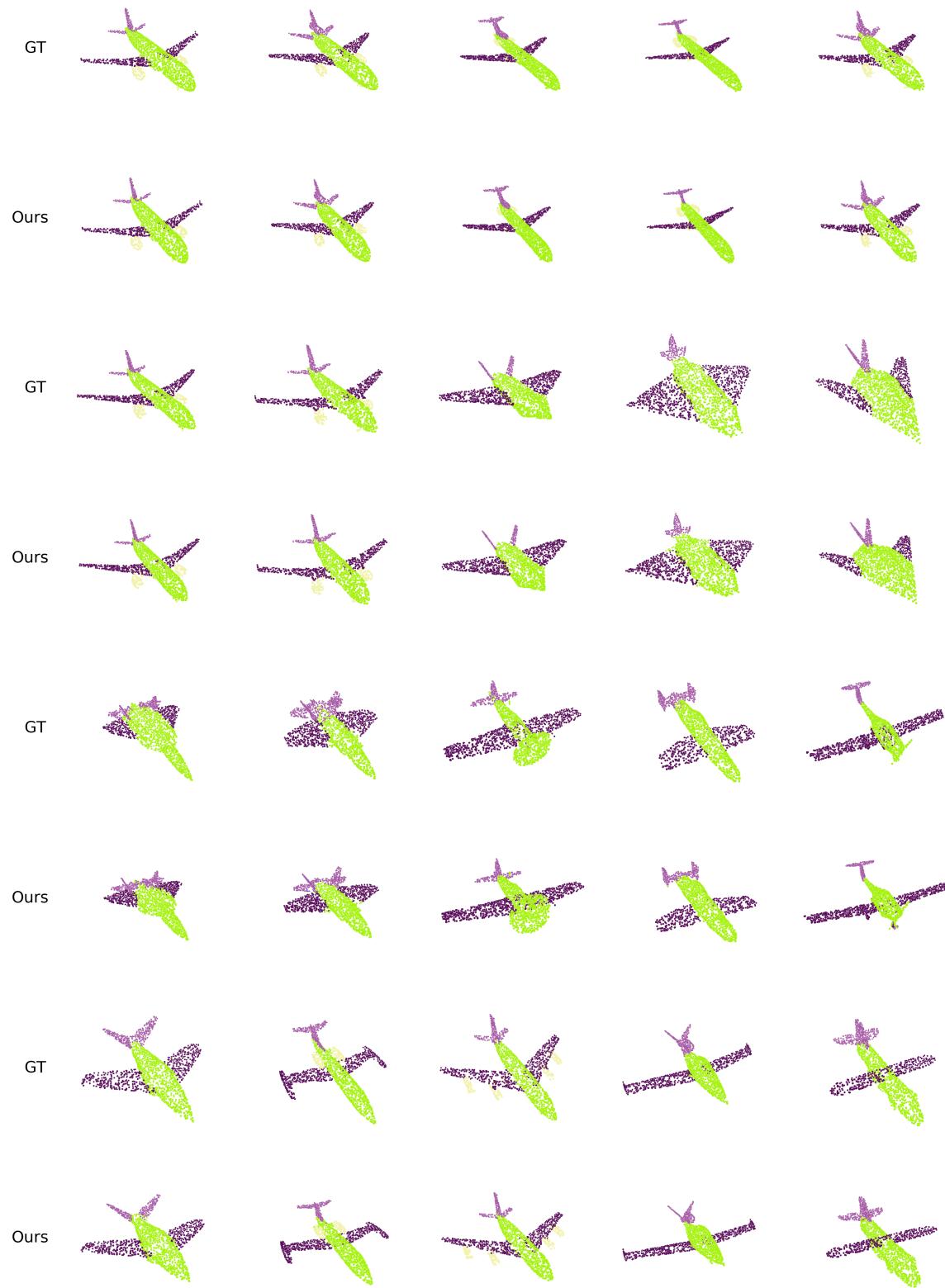


Figure 15. The visual result of semantic segmentation on ShapeNet part dataset (Airplane). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

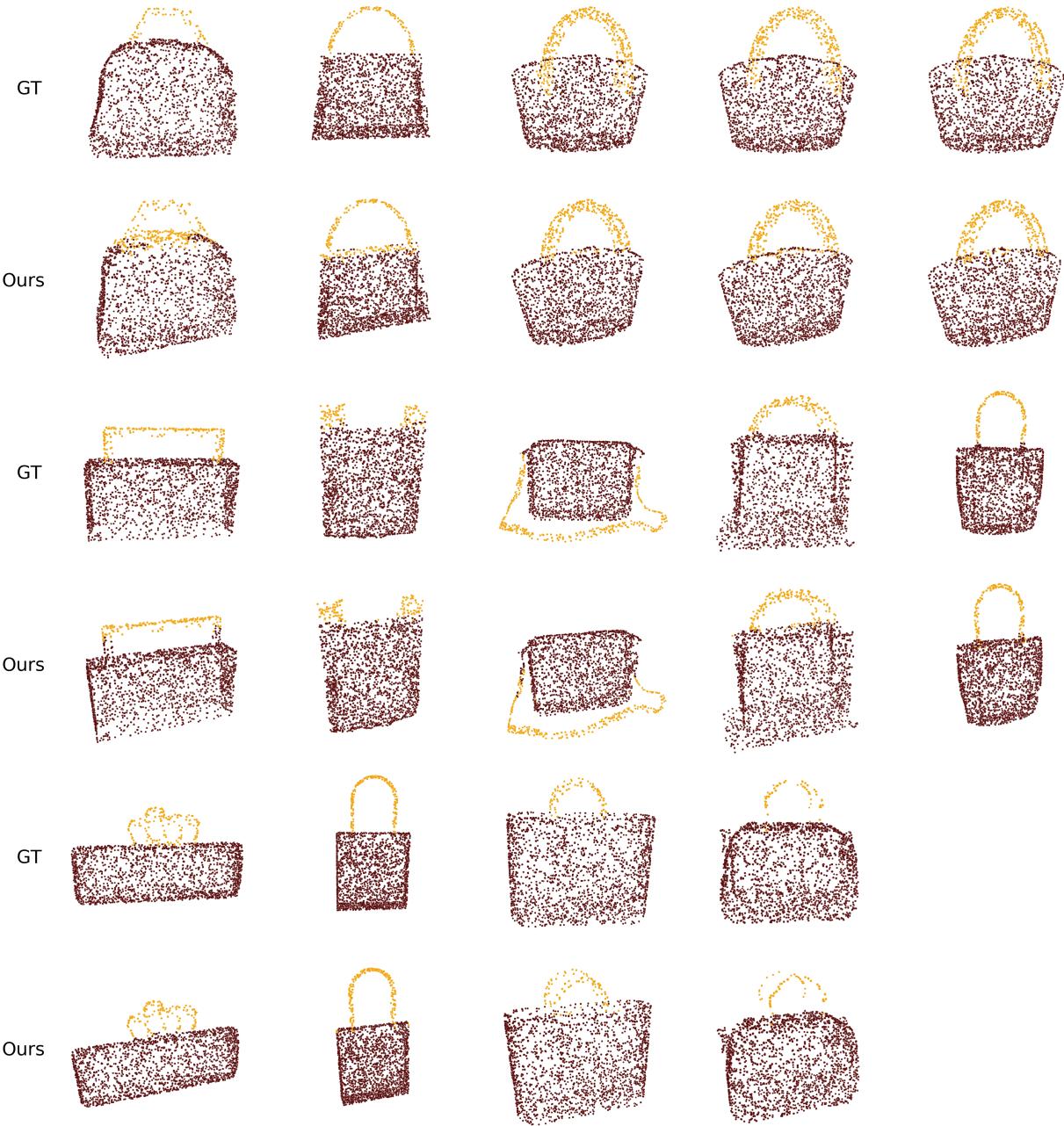


Figure 16. The visual result of semantic segmentation on ShapeNet part dataset (Bag). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

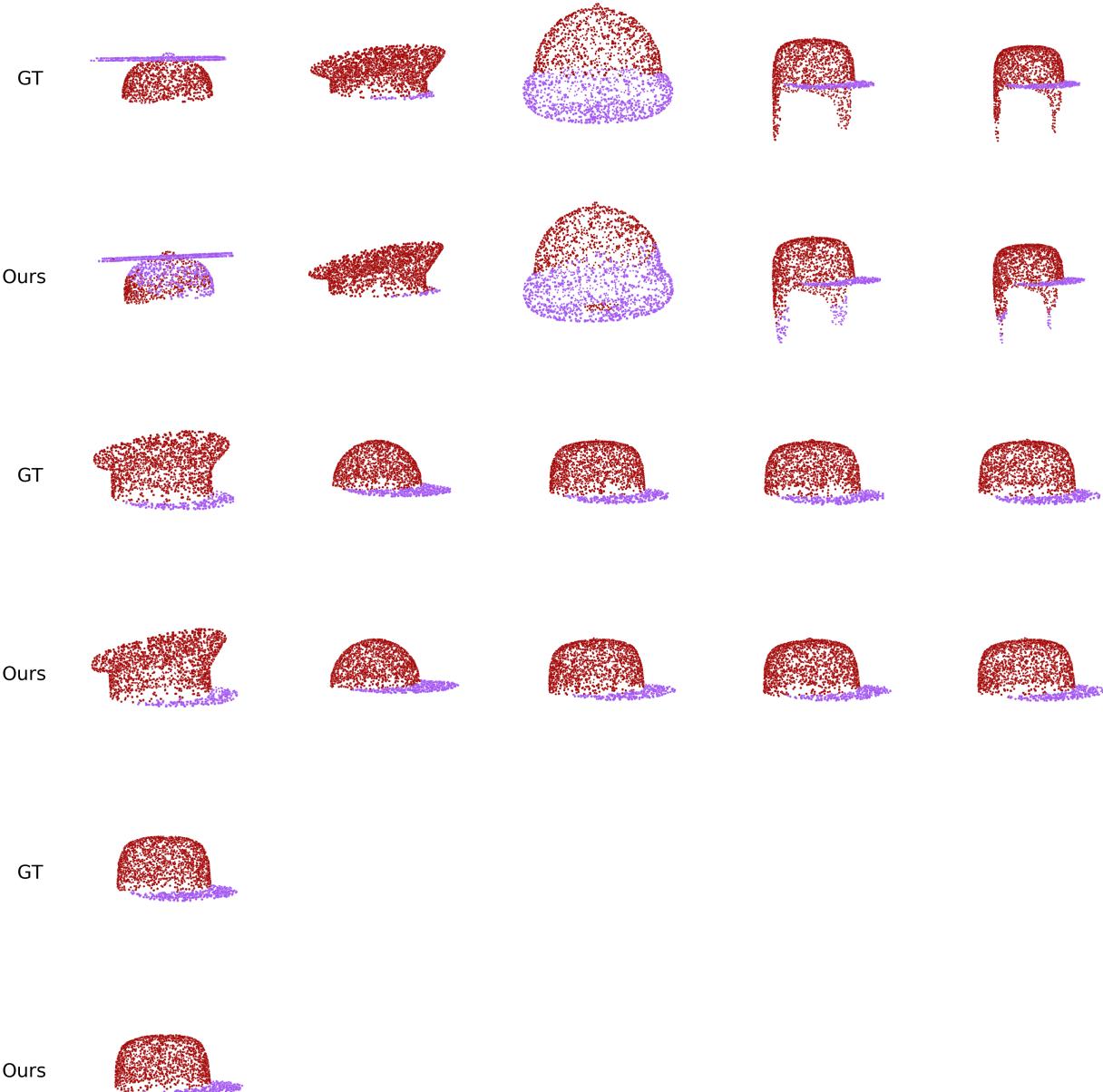


Figure 17. The visual result of semantic segmentation on ShapeNet part dataset (Cap). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

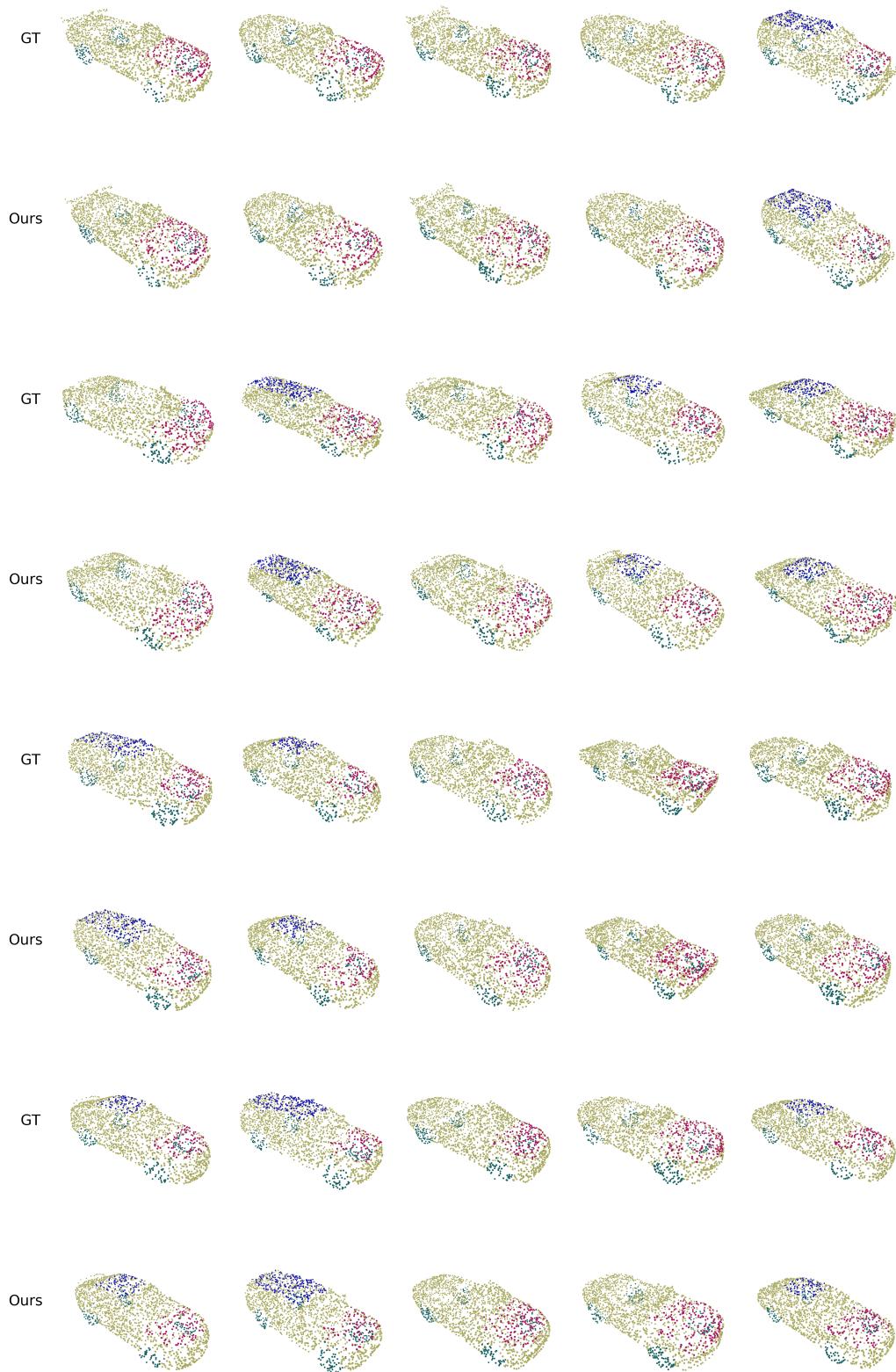


Figure 18. The visual result of semantic segmentation on ShapeNet part dataset (Car). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

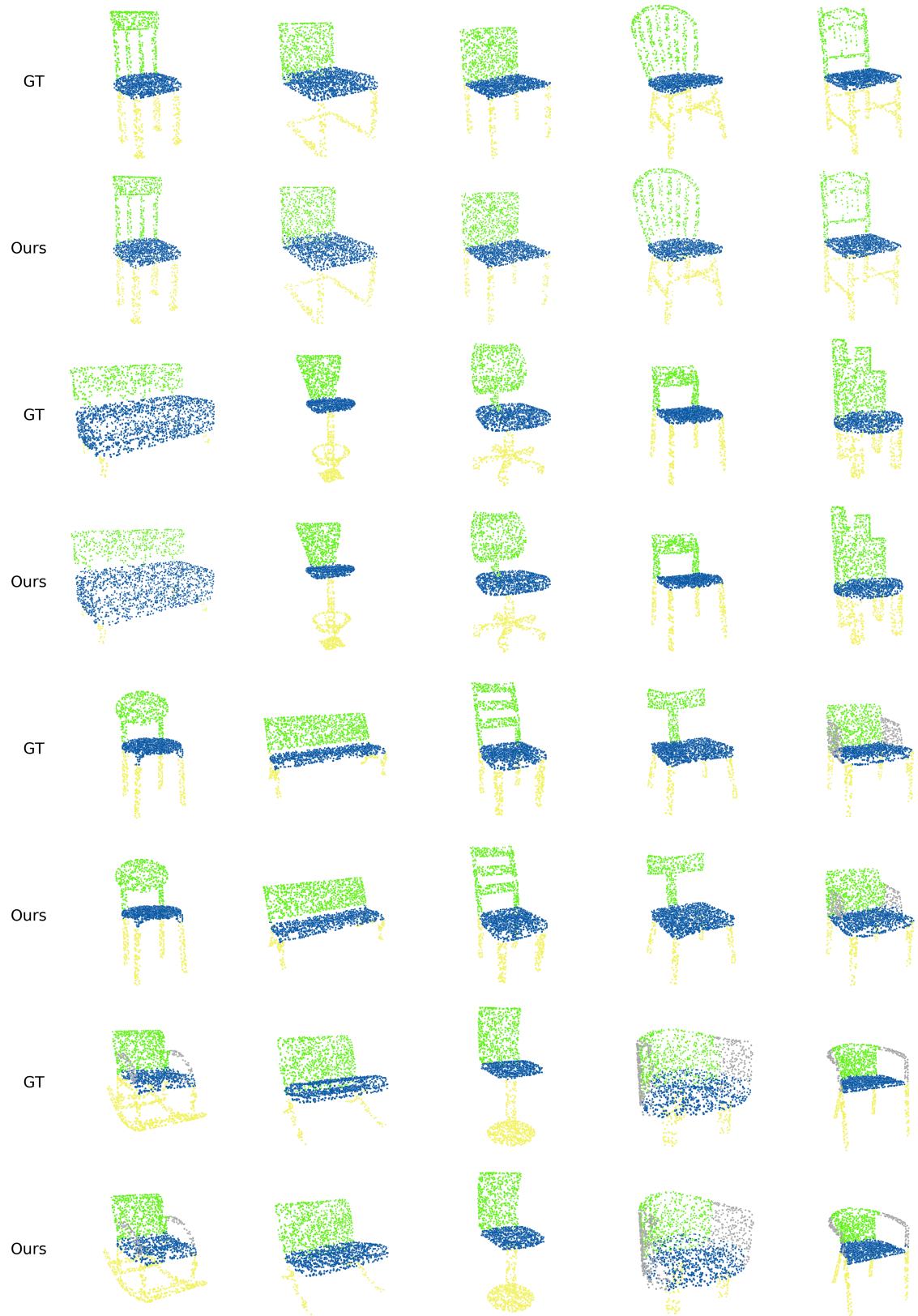


Figure 19. The visual result of semantic segmentation on ShapeNet part dataset (Chair). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

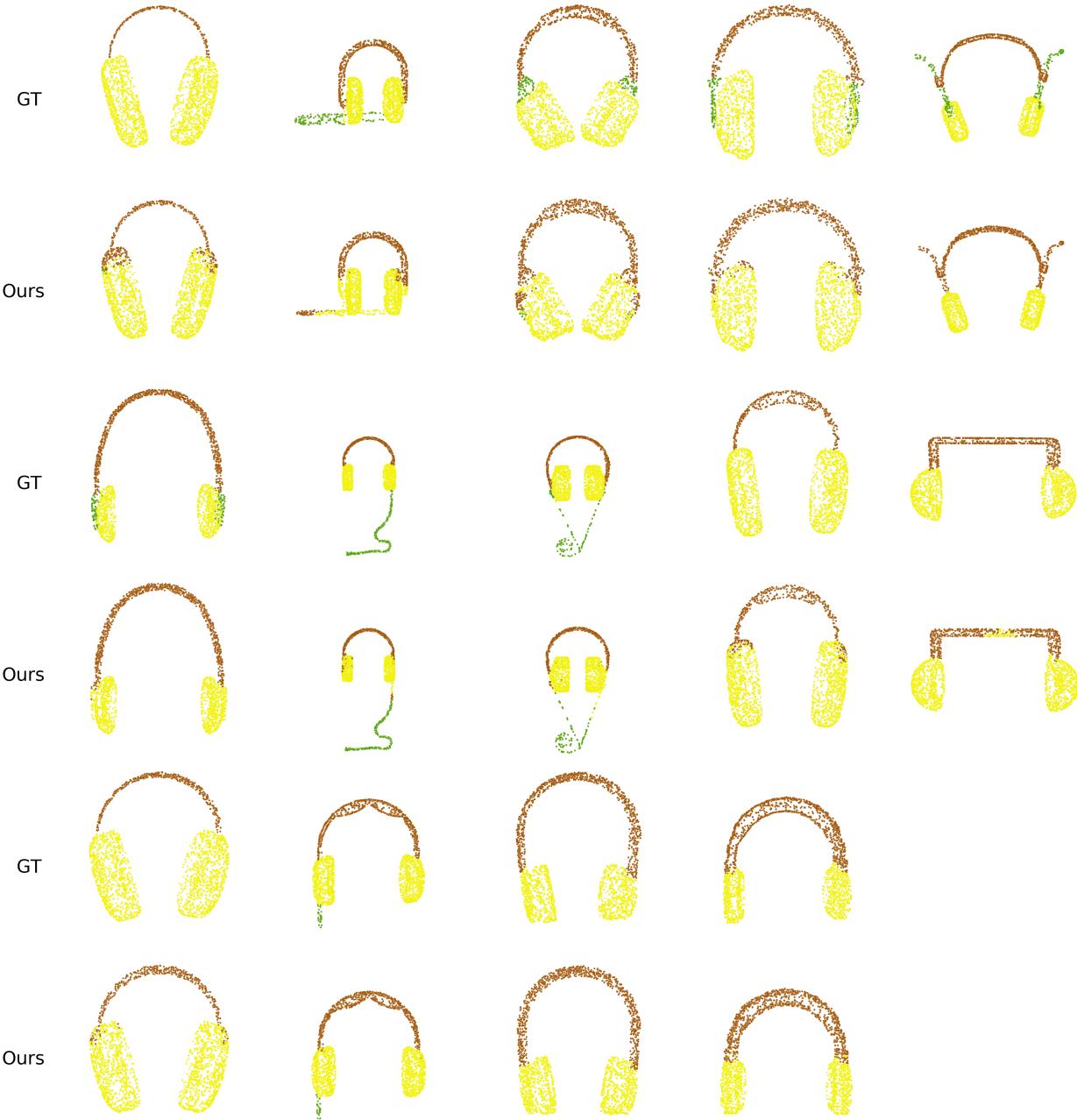


Figure 20. The visual result of semantic segmentation on ShapeNet part dataset (Earphone). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

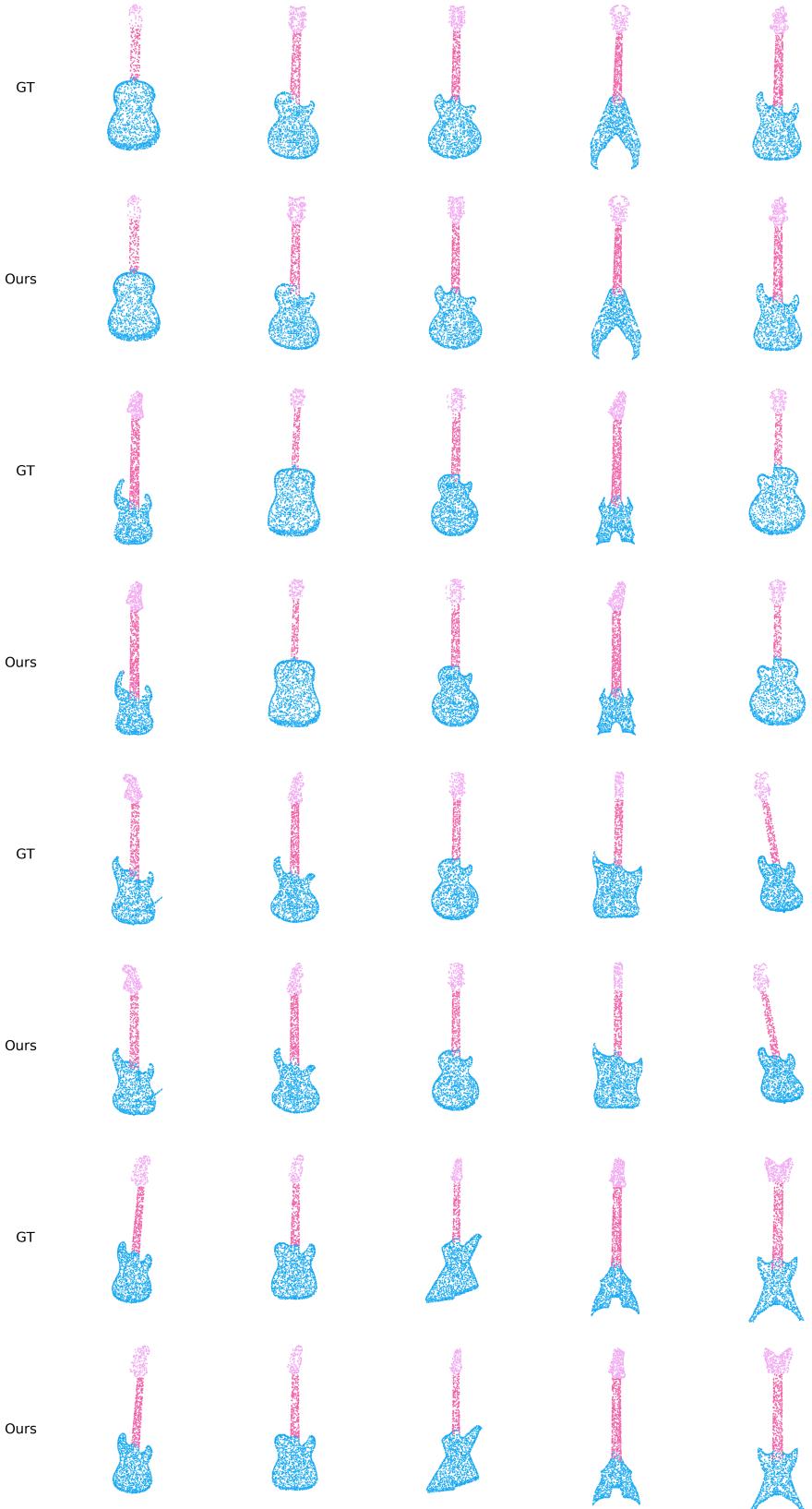


Figure 21. The visual result of semantic segmentation on ShapeNet part dataset (Guitar). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

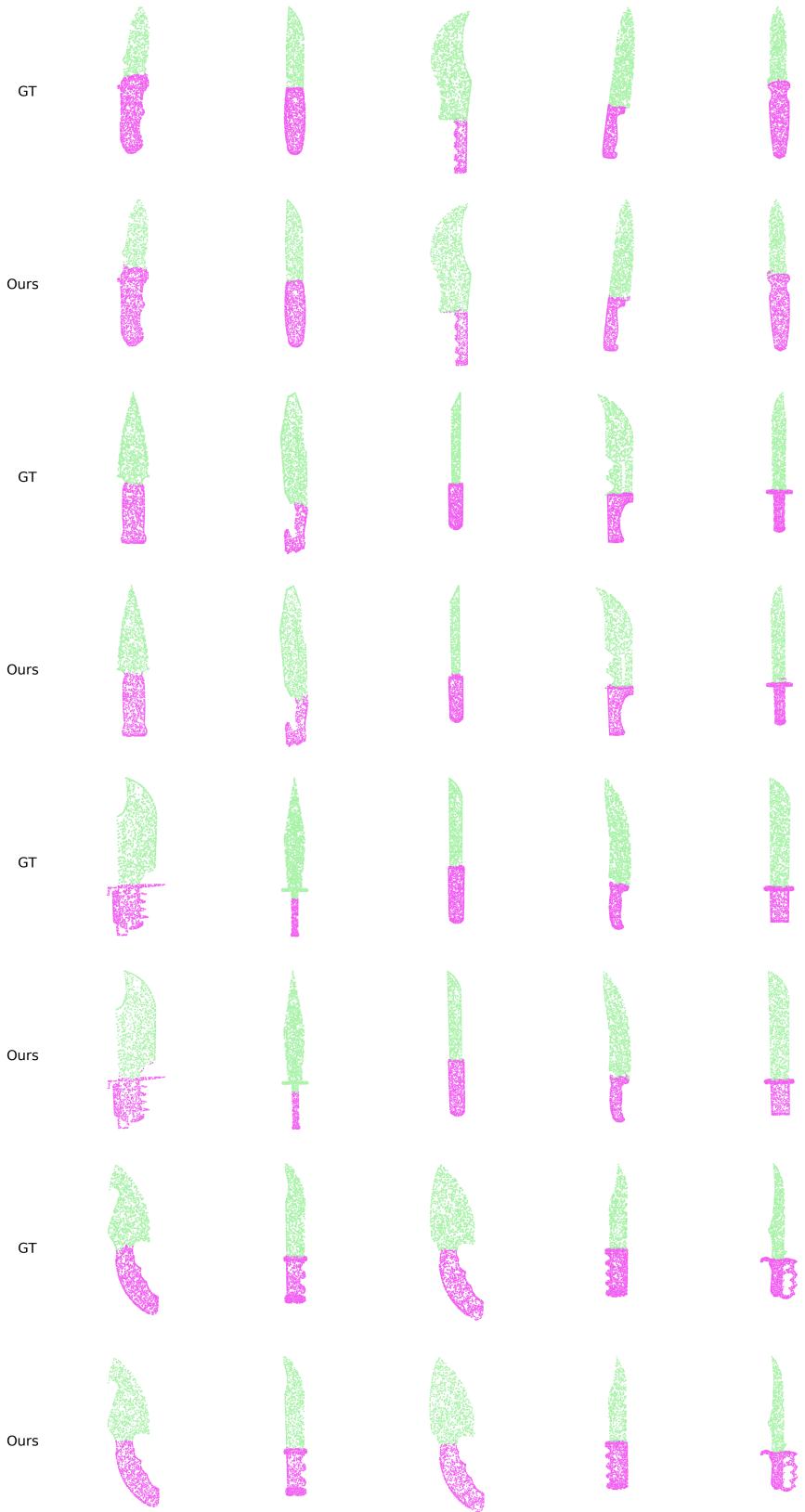


Figure 22. The visual result of semantic segmentation on ShapeNet part dataset (Knife). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

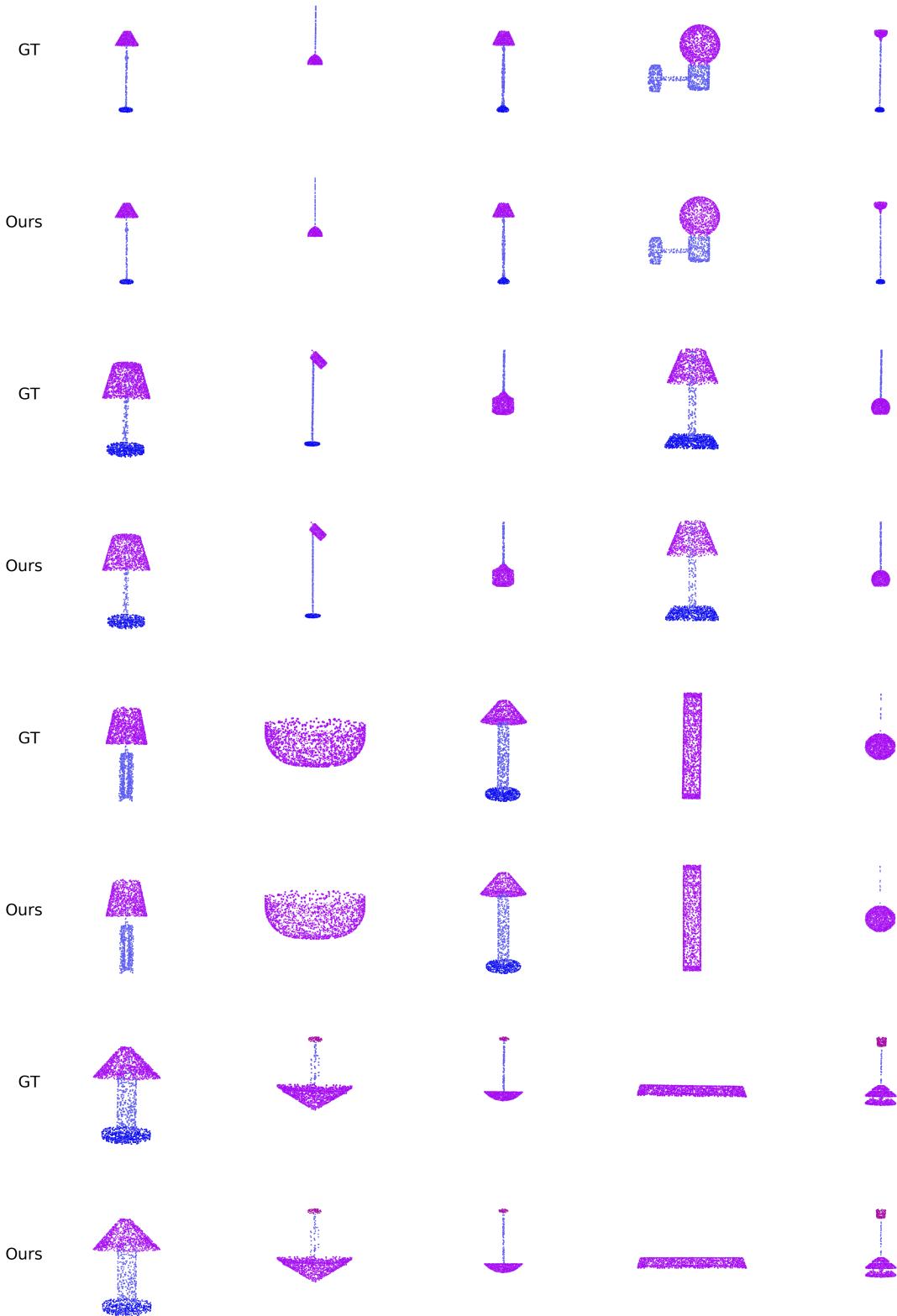


Figure 23. The visual result of semantic segmentation on ShapeNet part dataset (Lamp). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

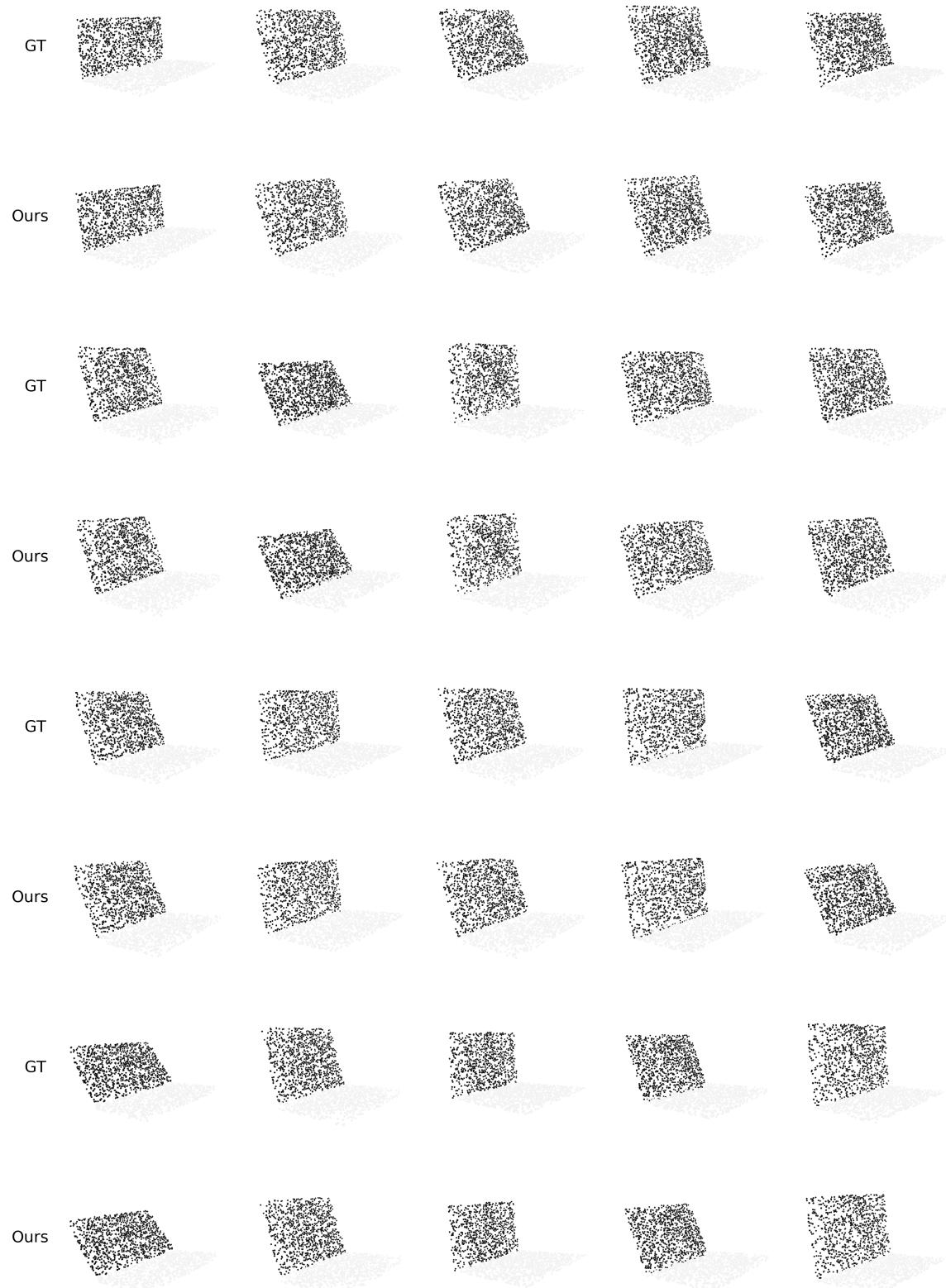


Figure 24. The visual result of semantic segmentation on ShapeNet part dataset (Laptop). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

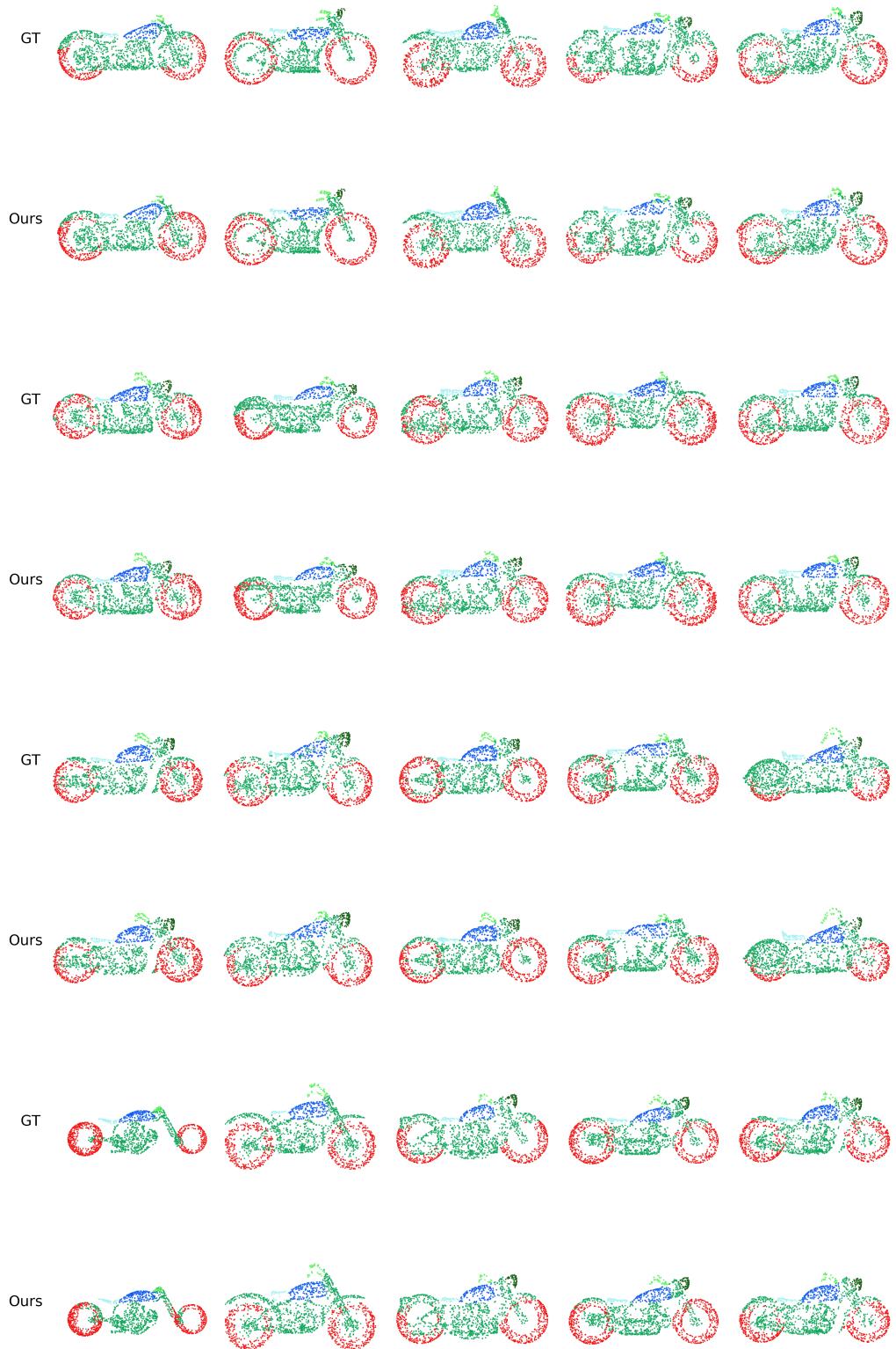


Figure 25. The visual result of semantic segmentation on ShapeNet part dataset (Motorbike). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

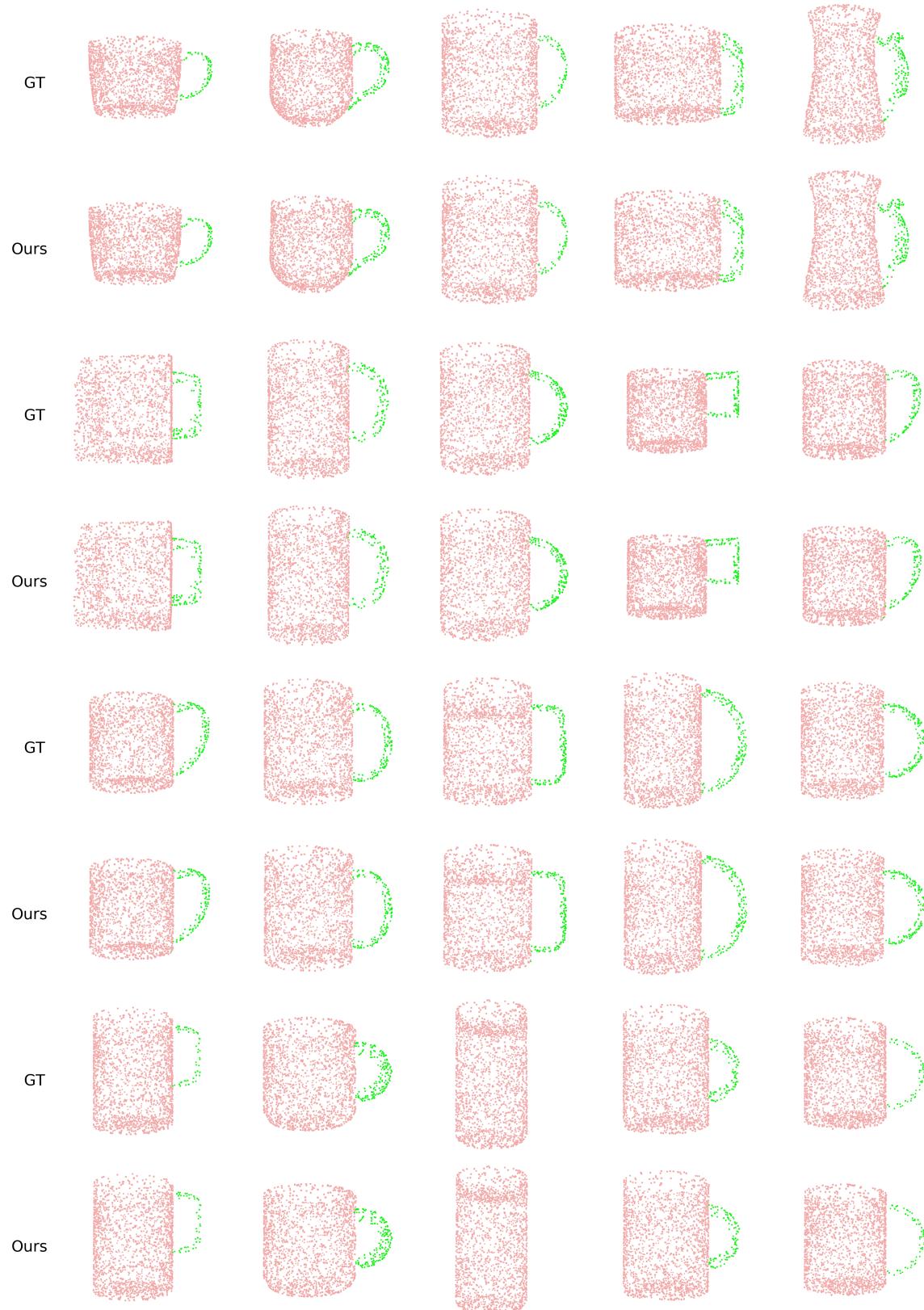


Figure 26. The visual result of semantic segmentation on ShapeNet part dataset (Mug). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).



Figure 27. The visual result of semantic segmentation on ShapeNet part dataset (Pistol). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

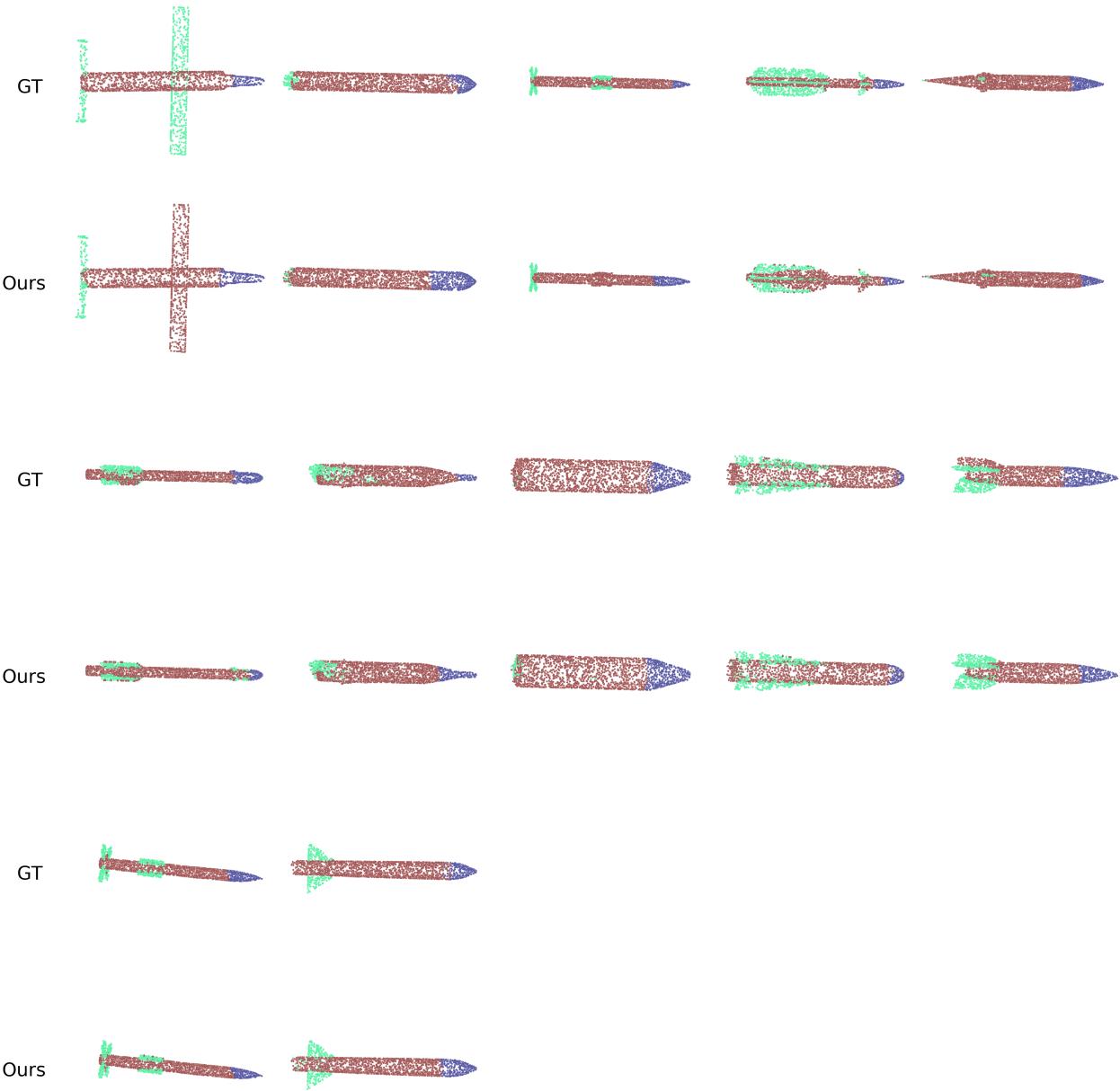


Figure 28. The visual result of semantic segmentation on ShapeNet part dataset (Rocket). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

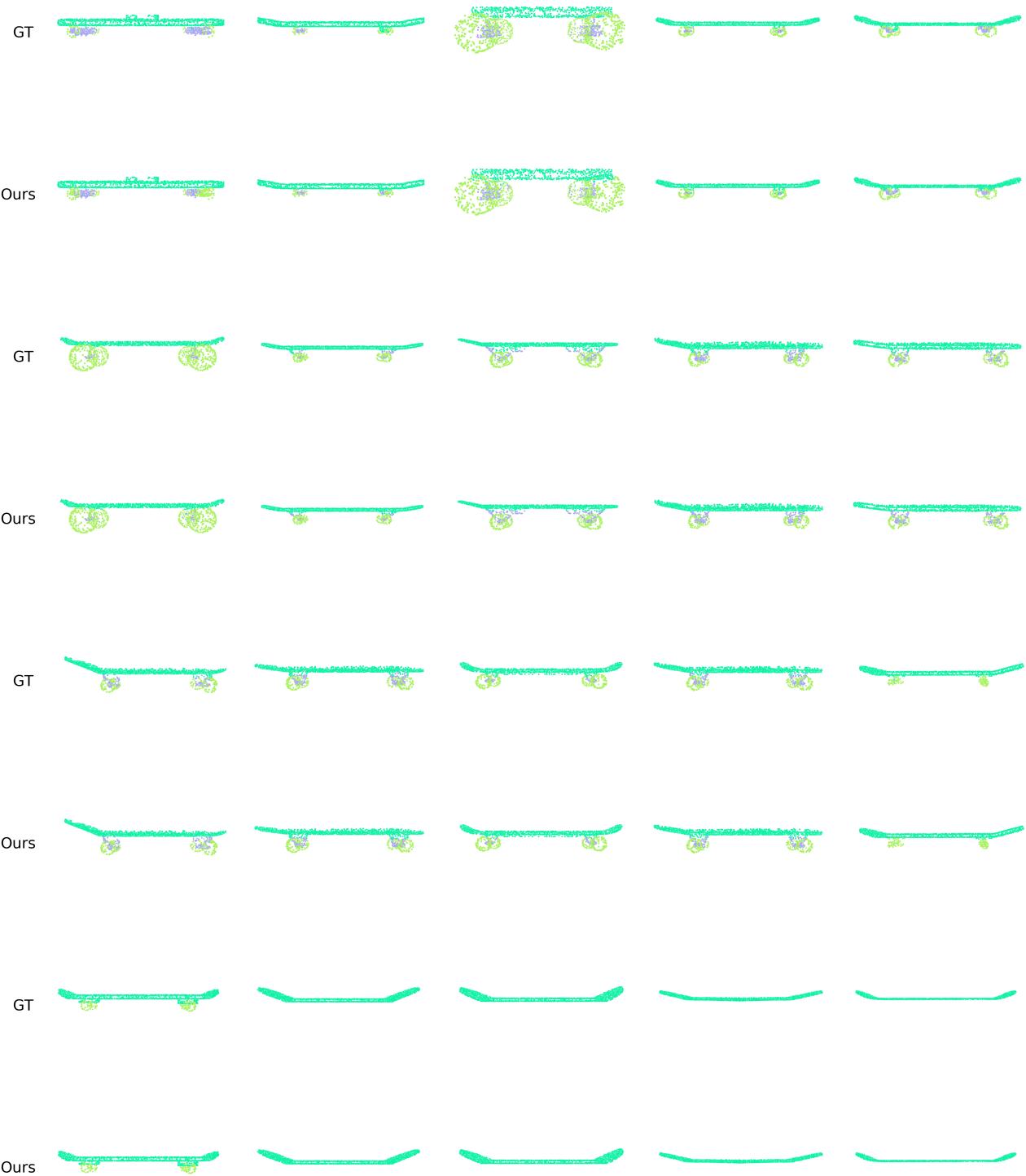


Figure 29. The visual result of semantic segmentation on ShapeNet part dataset (Skateboard). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).

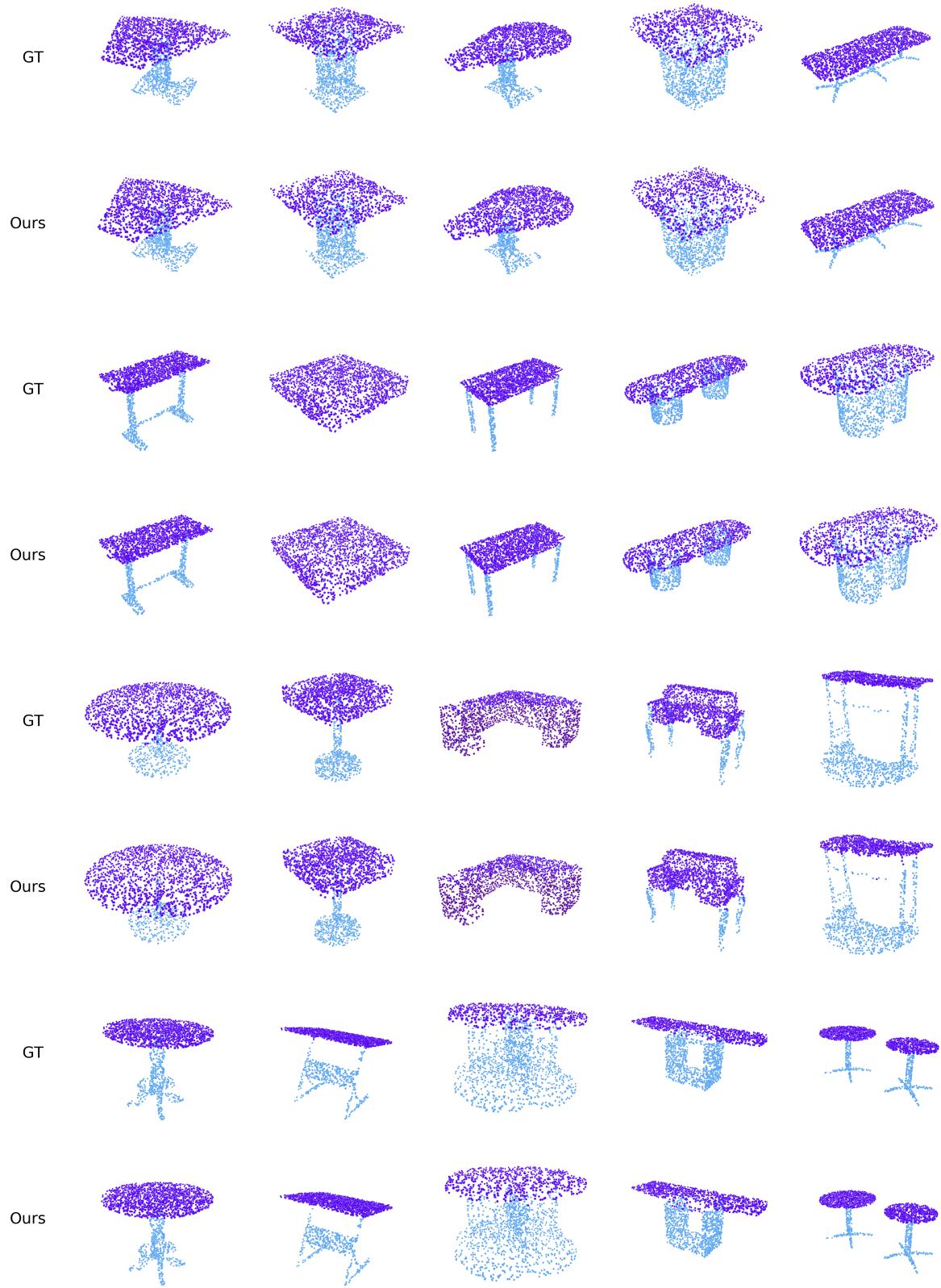


Figure 30. The visual result of semantic segmentation on ShapeNet part dataset (Table). For comparison, we show for each shape the semantic segmentation result (bottom) and the corresponding ground-truth (top).