SUPPLEMENTARY MATERIALS FOR MODEL2SCENE

Anonymous authors
Paper under double-blind review

We include the following information in the supplementary materials:

- **1**. Fine-tuning on the downstream tasks (semantic segmentation and object detection) when the voxel size is set to be 2 cm.
- **2**. Qualitatively evaluation of our method on the ScanNet and S3DIS datasets, we show more visualization results and the failure cases.
- **3**. The source code is included in the file of 'code.zip'. The backbone (MinkowskiNet14A) and other network details can be found in 'models.py'. We will release its clean version and all processed data in the near future.

1 FINE-TUNING ON THE DOWNSTREAM TASKS.

For efficient training, we set the voxel size to be 5 cm in the paper. To verify the effectiveness under different voxel resolutions, we also conduct the experiments with 2 cm voxel size. As shown in Table 1, 2 and 3, the performance are significantly improved in the semantic segmentation and object detection tasks. Note that only geometric features are used for training. Therefore, the baseline performance is slightly lower than that reported in pointContrast Xie et al. (2020). Besides, We find it can boost the performance of the unseen categories that are not available in ModelNet (Table 4). It is because the network learns common local structure features from seen classes that can be adapted to unseen classes.

Table 1: Fine-tuning on the Scannet. We omit the % to show the IoU performance. The number in () donates the improved accuracy compared with purely supervised training.

Model	MinkNet14	MinkNet34
Trained from scratch	67.60	70.13
PointContrast Xie et al. (2020)	69.40(1.80)	71.90(1.77)
Ours	**69.58(1.98)**	**71.94(1.81)**

Table 2: Fine-tuning on the S3DIS dataset for semantic segmentation task.

Model	MinkNet14	MinkNet34
Trained from scratch	64.15	65.79
PointContrast Xie et al. (2020)	66.18(2.03)	68.93(2.14)
Ours	**66.64(2.49)**	**68.99(2.20)**

Table 3: 3D object detection results on ScanNet dataset.

Model	mAP@0.5	mAP@0.25
Trained from scratch	34.57	56.27
PointContrast	36.52(1.95)	**58.45(2.18)**
Ours	**37.12(2.55)**	58.46(2.19)

2 QUALITATIVELY EVALUATION

More visualization results for all categories on two datasets are shown in this section, indicating the effectiveness of our method. We show each case's ground truth (the binary map) and the prediction result (the heat map). Besides, we also present the failure cases for each category, which may inspire

Table 4: Fine-tuning on the Scannet labelled data, seen classes are the types of synthetic models that existed in ModelNet.

Model	All classes	Seen classes	Unseen classes
Trained from scratch	67.60	71.22	63.17
Ours	**69.45**(**1.85**)	**73.23**(**2.01**)	**64.82**(**1.65**)

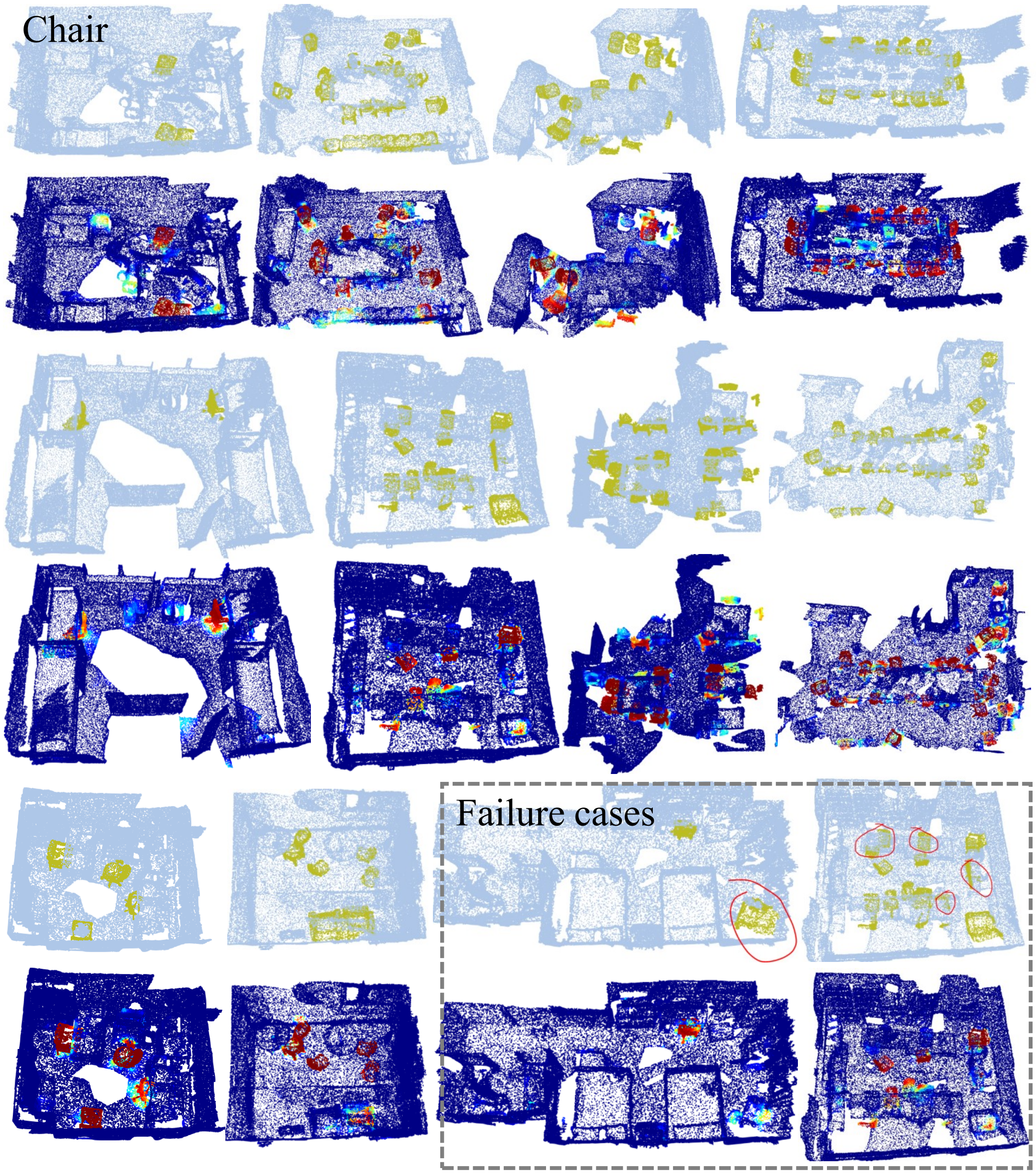


Figure 1: Visualization of inferring chair on ScanNet dataset by our method.

future works for improvement. As shown in Figure 1∼15, our method achieve promising results on both ScanNet and S3DIS dataset. However, our method sometimes misses and falsely detect the object in some hard cases.

REFERENCES

Saining Xie, Jiatao Gu, Demi Guo, Charles R Qi, Leonidas Guibas, and Or Litany. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. In *European Conference on Com-*

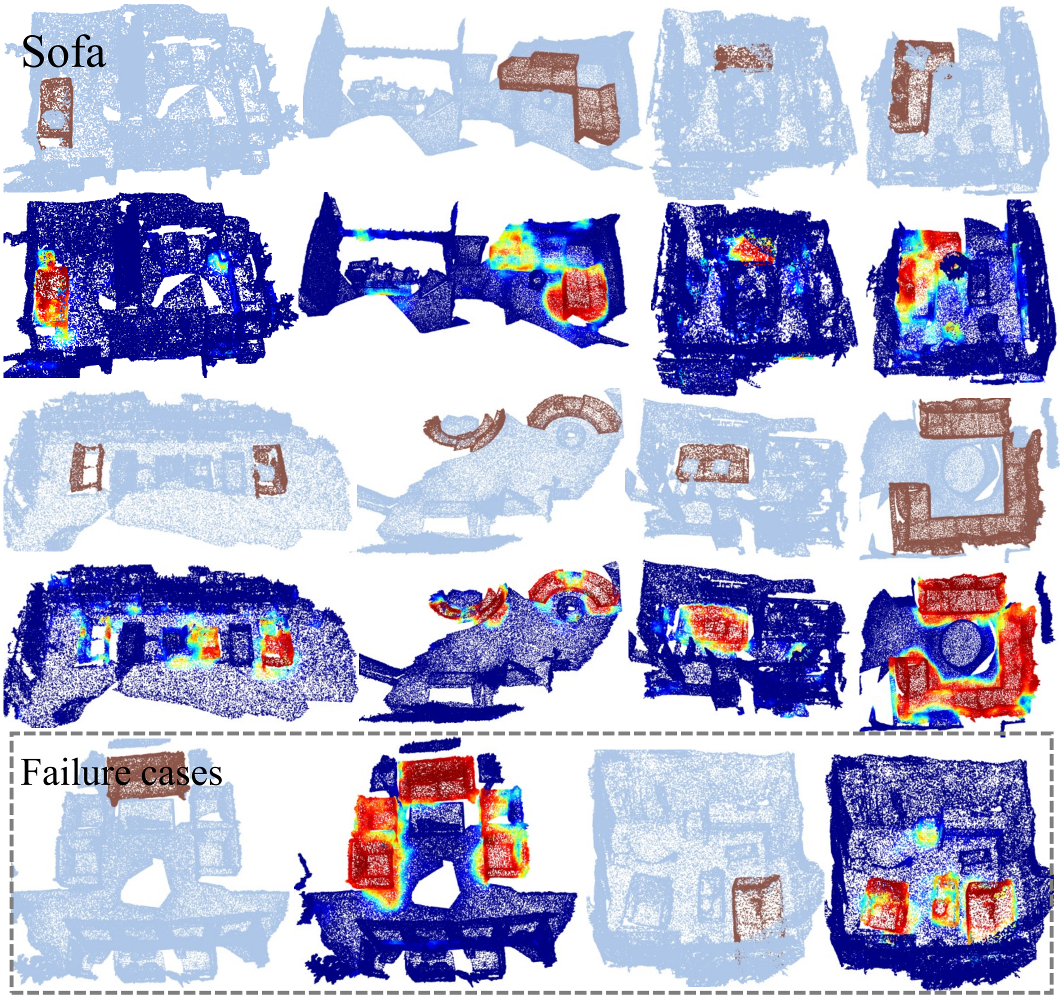


Figure 2: Visualization of inferring sofa on ScanNet dataset by our method.

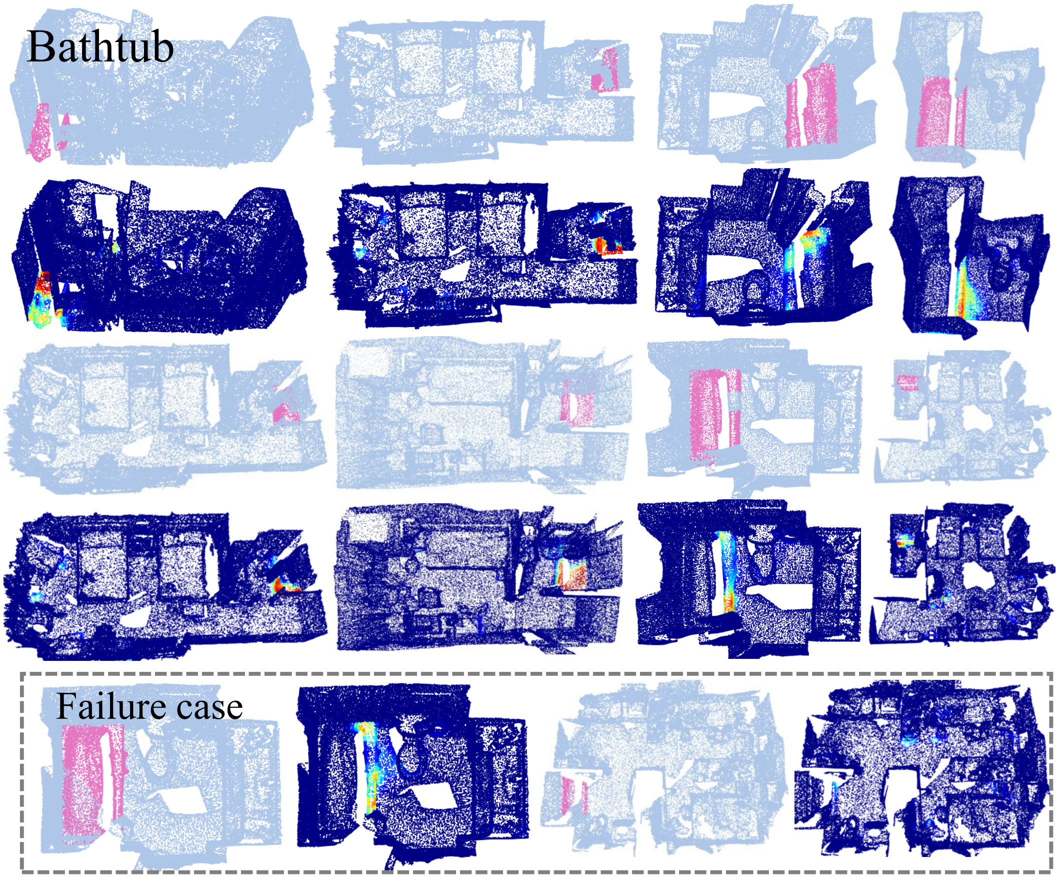


Figure 3: Visualization of inferring bathtub on ScanNet dataset by our method.

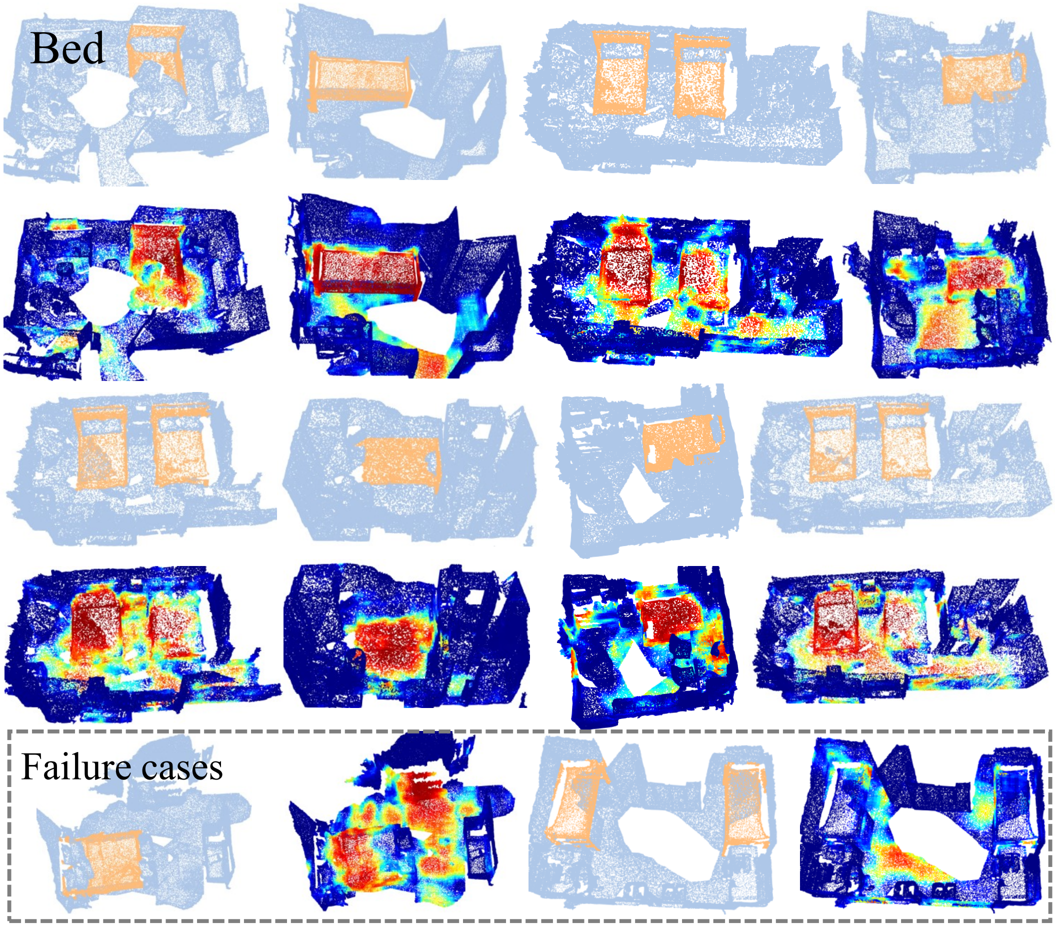


Figure 4: Visualization of inferring bed on ScanNet dataset by our method.

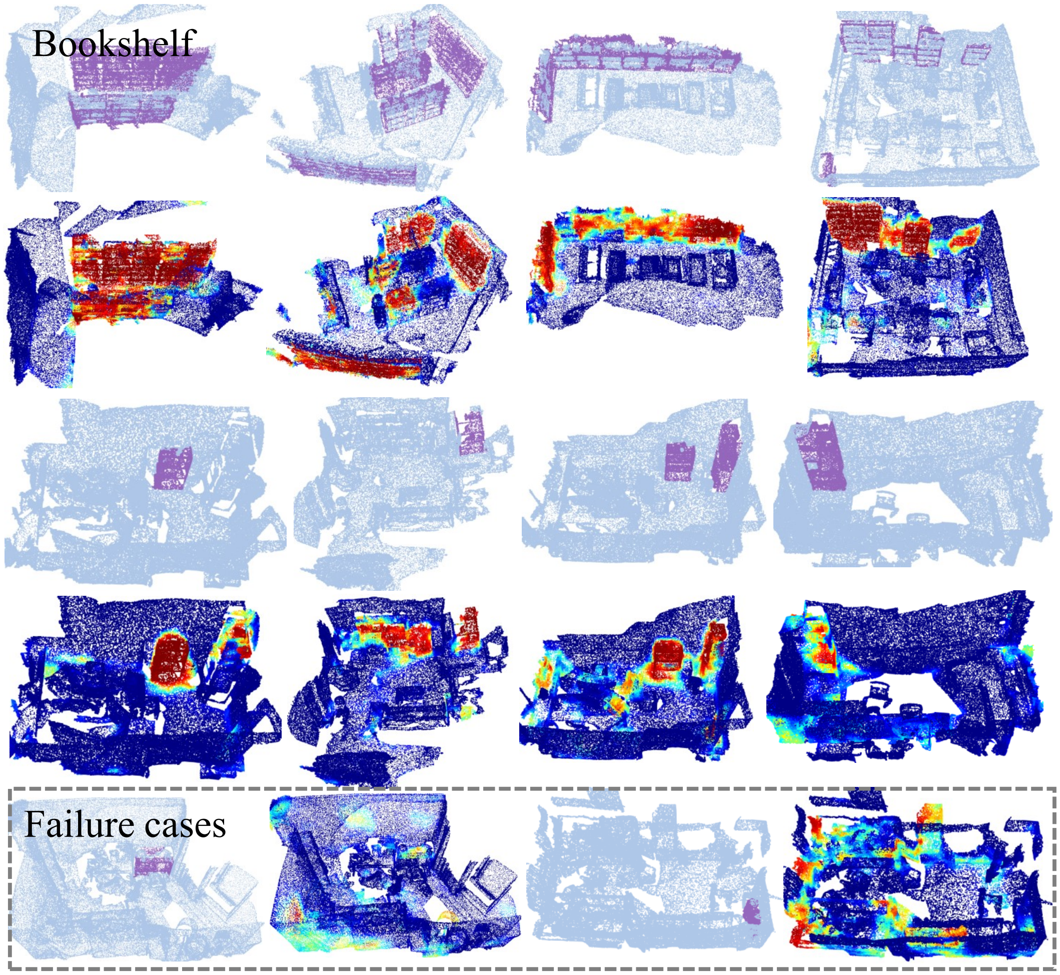
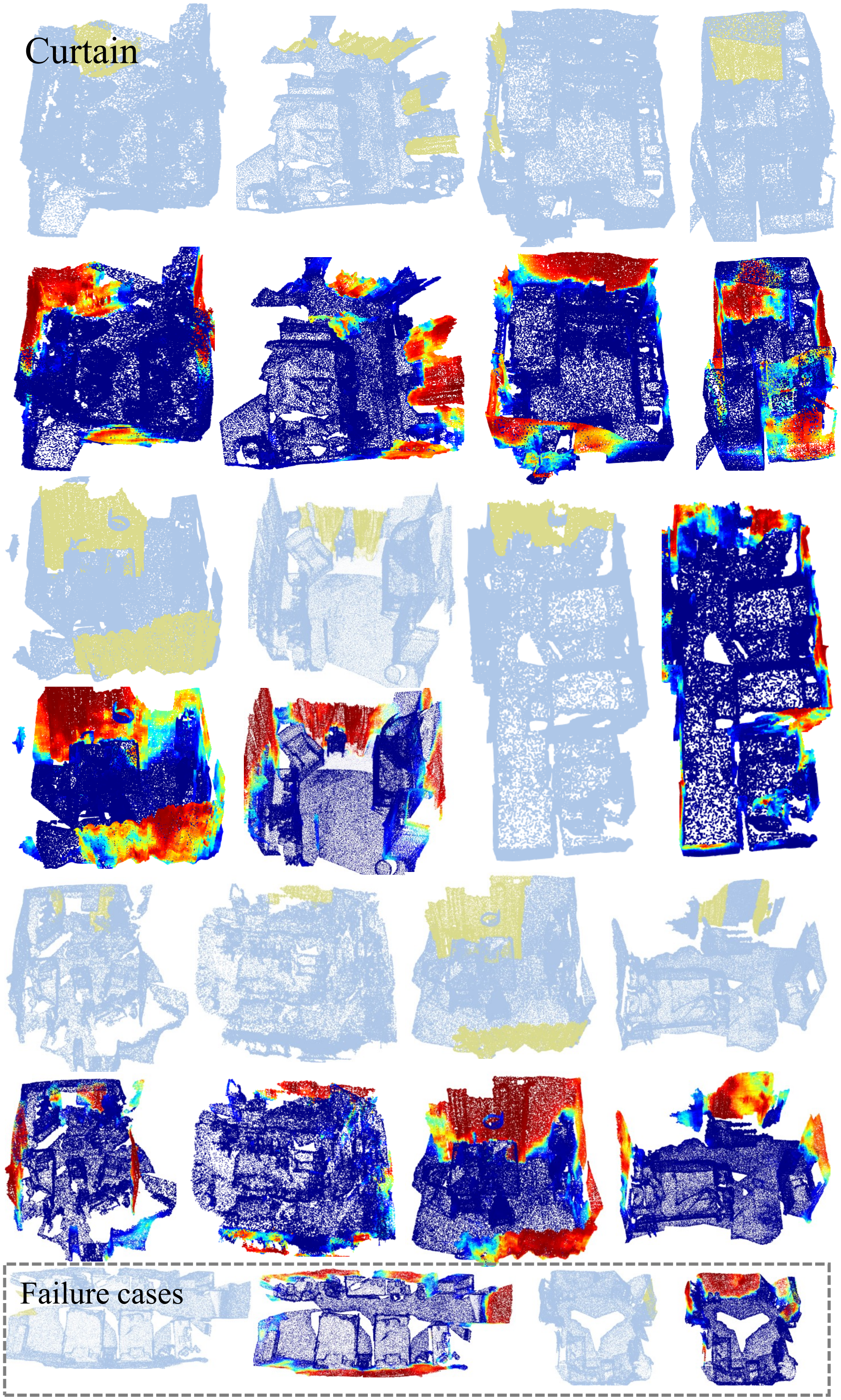


Figure 5: Visualization of inferring bookshelf on ScanNet dataset by our method.


Curtain
Failure cases

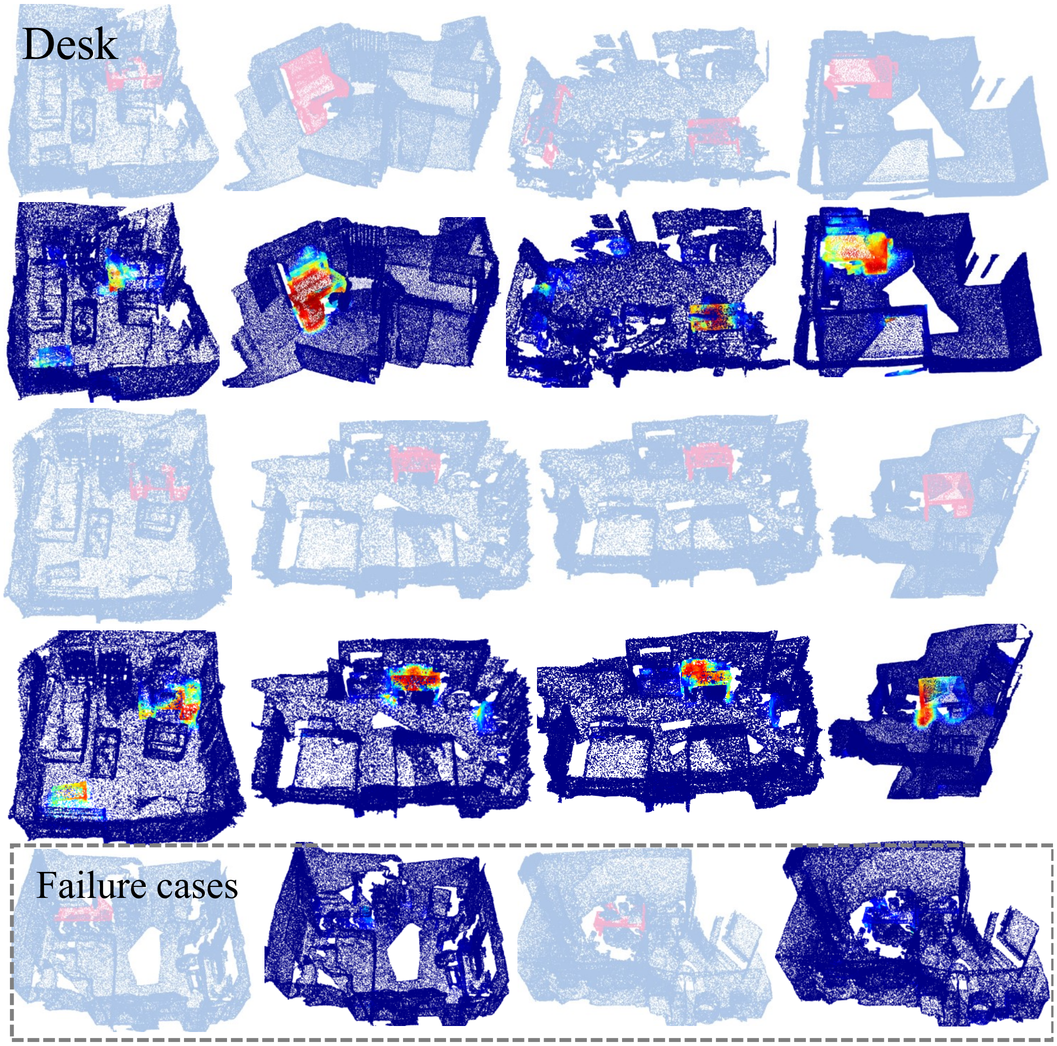


Figure 7: Visualization of inferring desk on ScanNet dataset by our method.

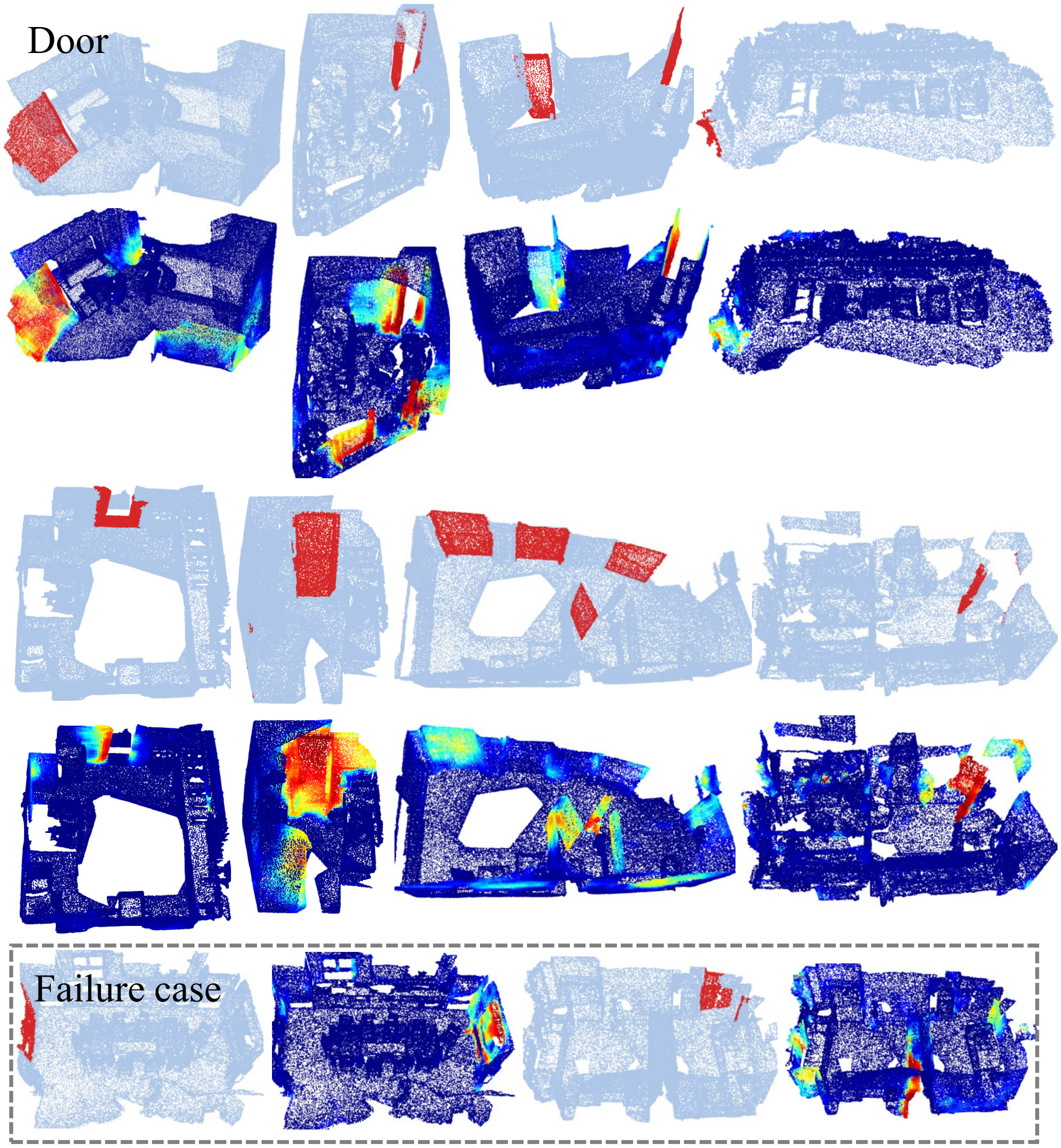


Figure 8: Visualization of inferring door on ScanNet dataset by our method.

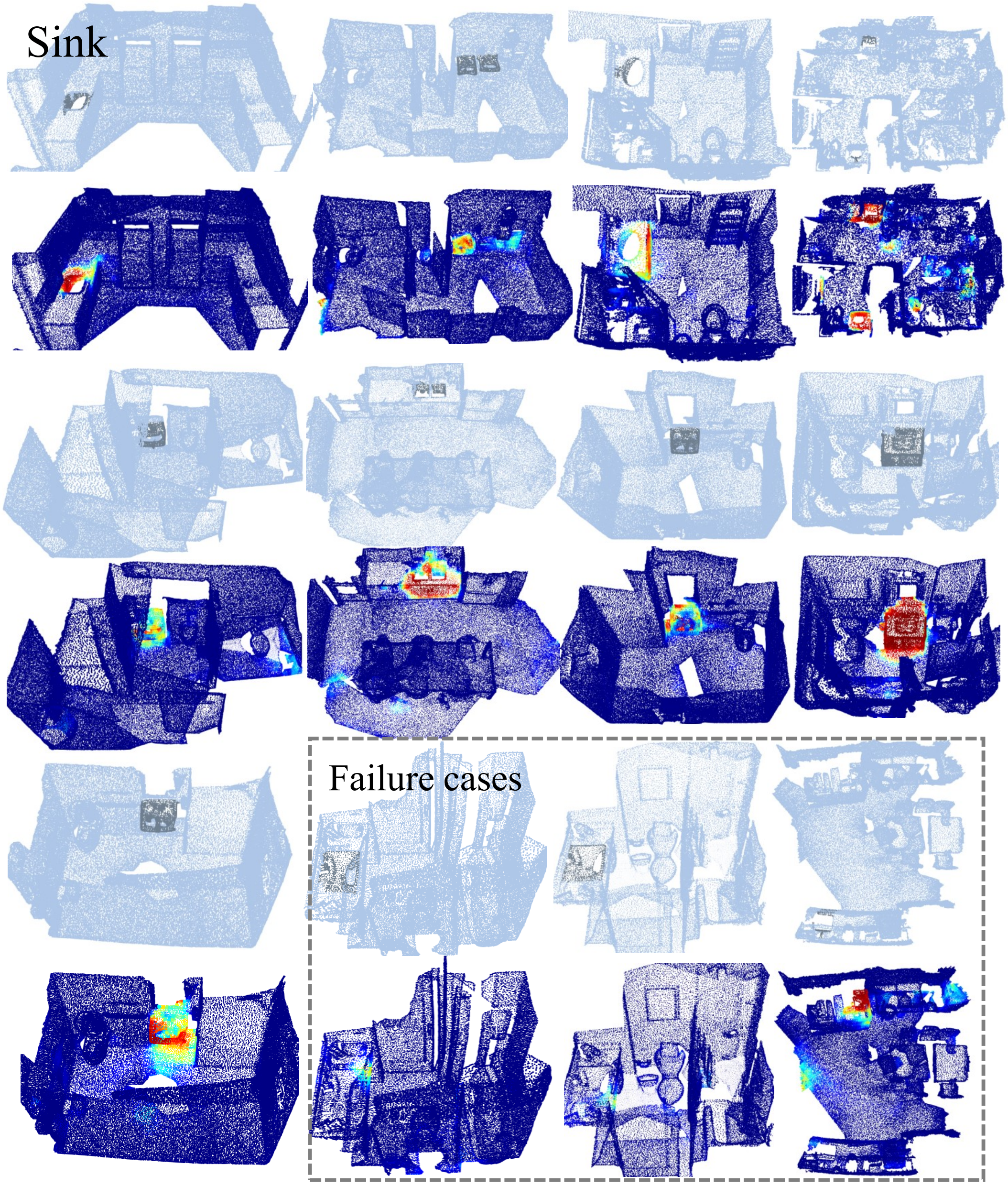


Figure 9: Visualization of inferring sink on ScanNet dataset by our method.

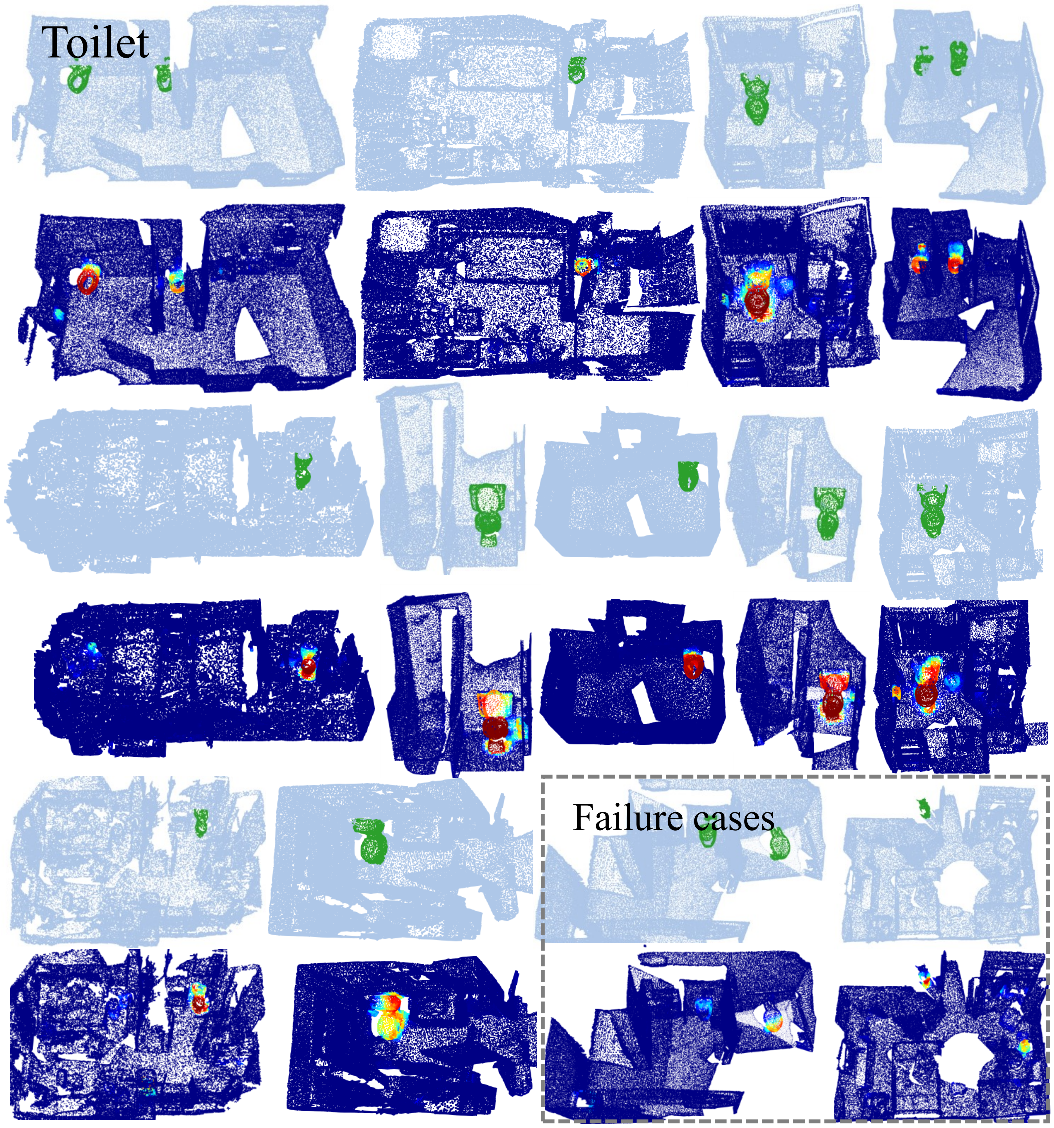


Figure 10: Visualization of inferring toilet on ScanNet dataset by our method.

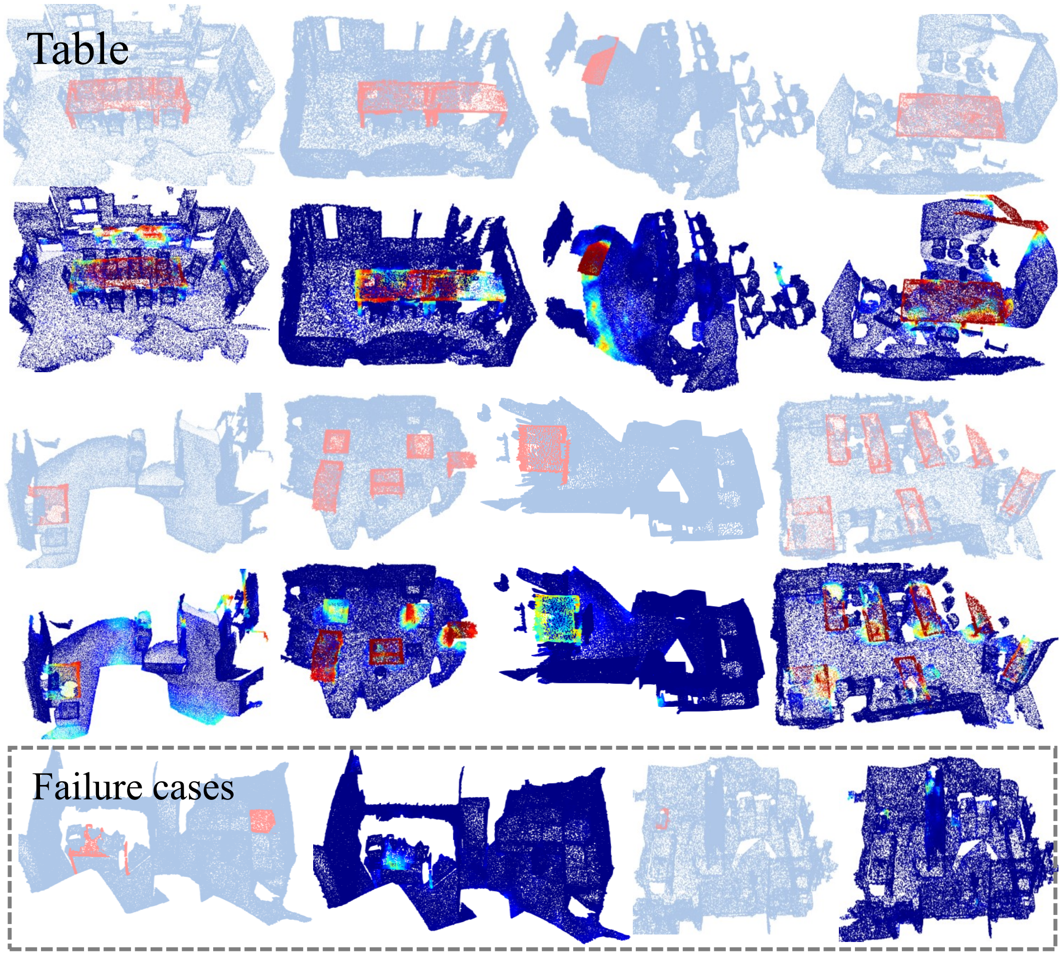


Figure 11: Visualization of inferring table on ScanNet dataset by our method.

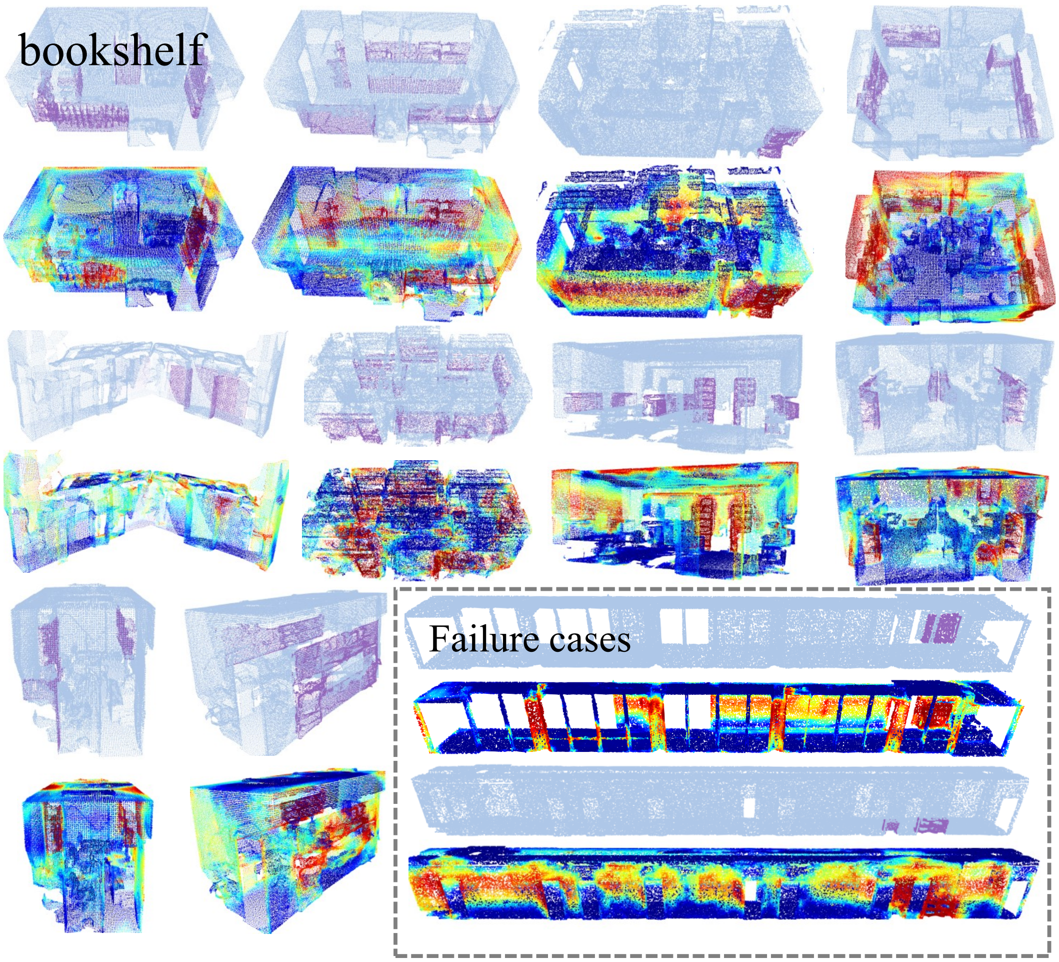


Figure 12: Visualization of inferring bookshelf on S3DIS dataset by our method.

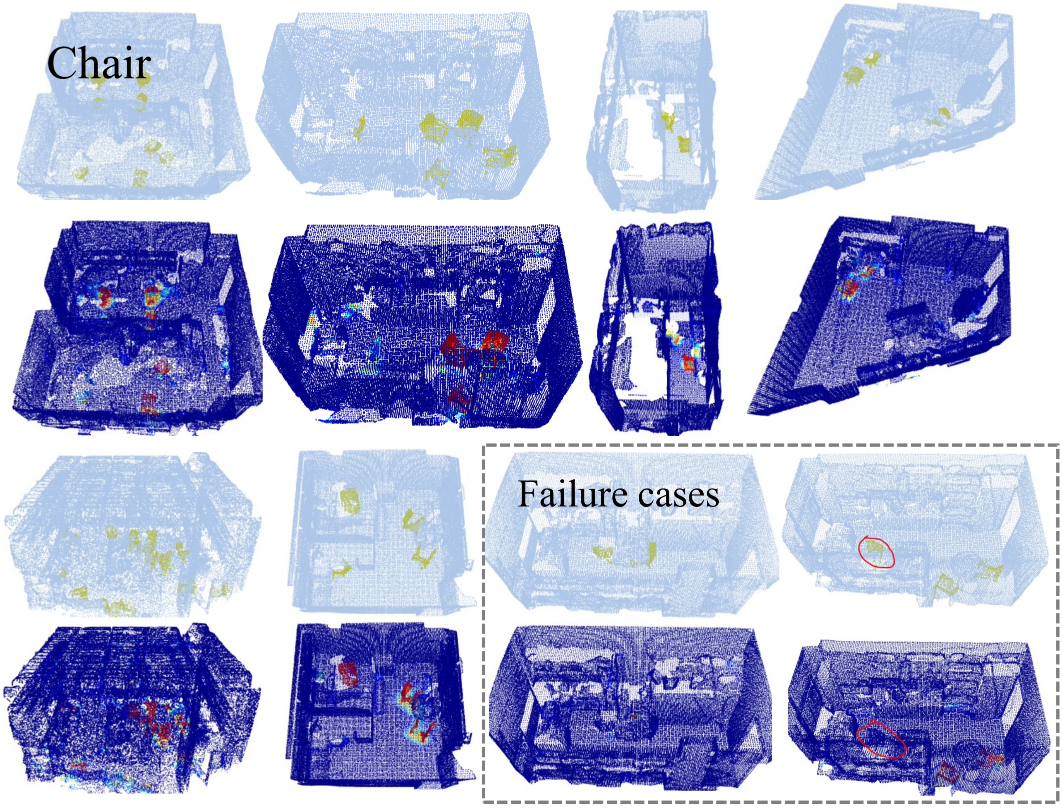


Figure 13: Visualization of inferring chair on S3DIS dataset by our method.

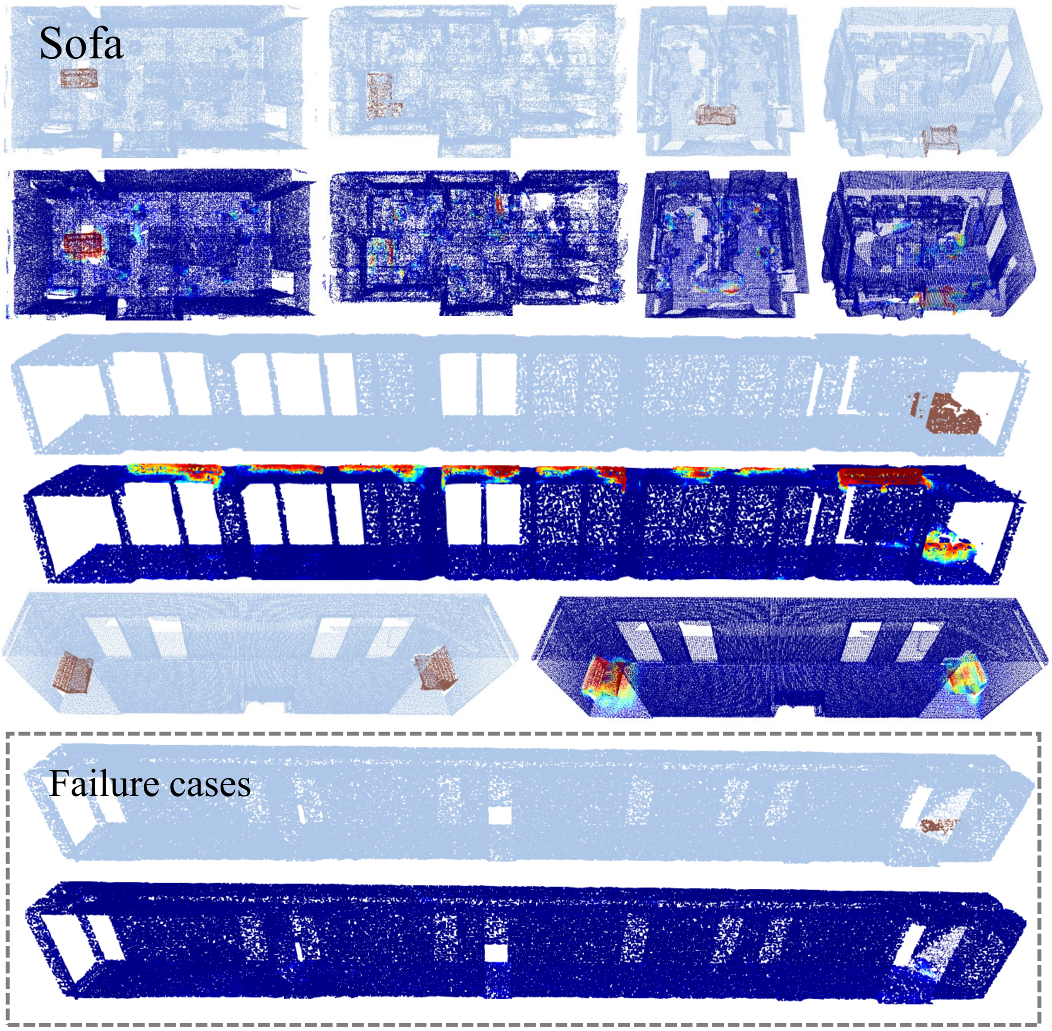


Figure 14: Visualization of inferring sofa on S3DIS dataset by our method.

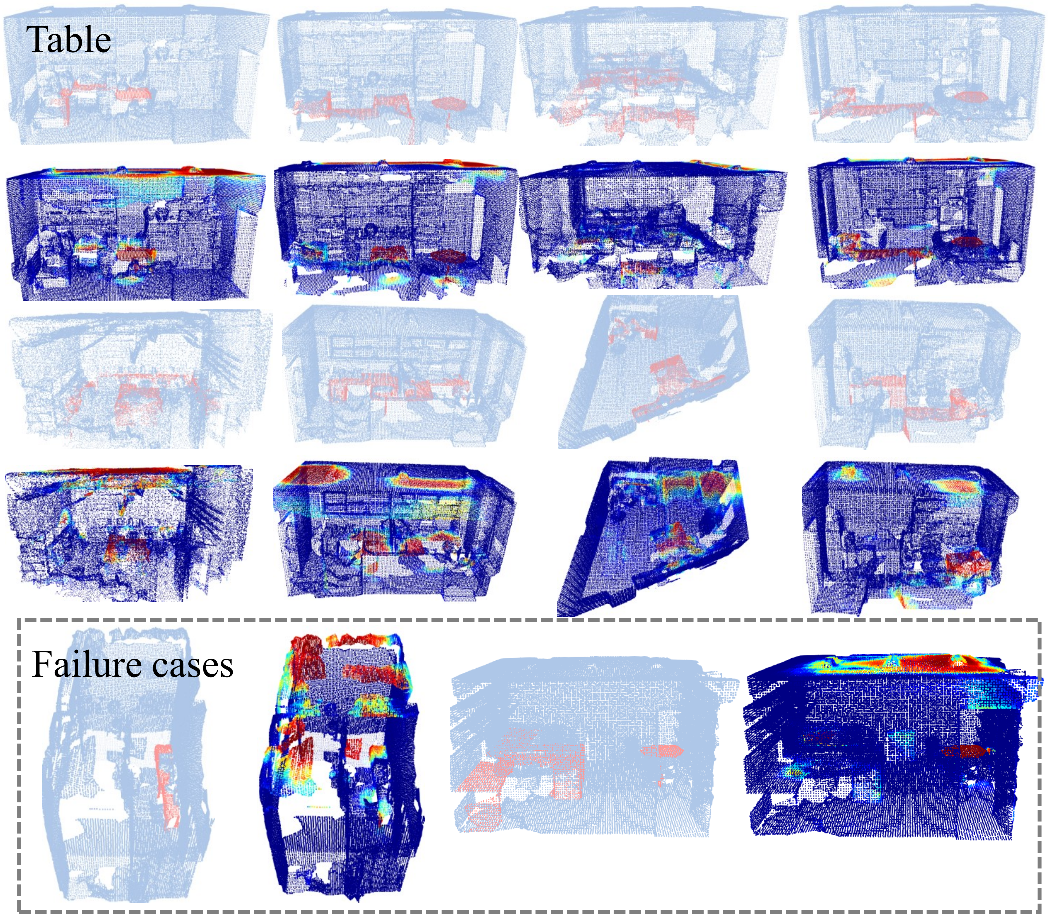


Figure 15: Visualization of inferring table on S3DIS dataset by our method.

puter Vision, pp. 574–591. Springer, 2020.