A. t-SNE Visualization

We visualize graph embeddings obtained by fixing the test masking ratio to 0 using t-SNE [55]. We could observe that plant and chair are two classes clustering close but easy to distinguish. The reason could be that both of them are tall cuboids. But chairs have a more regular appearance than plants. The piano and range hood are also the same cases. They have a similar outlook but are different when looking in detail. Usually, the mesh of the range hood has a hole inside its body. Another confusion to the model is the nightstand and dresser, two potentially similar objects. The t-SNE plot at ratios 0.3 and 0.6 are quite similar in clustering. While the plot at a ratio of 0.9 begins to confuse objects like desks and pianos (see the bed category move from the corner to the center).

Transitioning from a four-head to a three-head embedding suggests that the model is starting to merge certain aspects of objects from various categories into a single recognition region. This phenomenon elucidates why there is a decline in the model's performance when the training masking ratio is adjusted from 0.6 to 0.9.

B. Network Architecture

The overall architecture of our network is shown in Figure 7. It has a heavier encoder than the decoder. It follows the design logic in MaskMAE since, after pre-training, we no longer need the decoder. The reason we did not use batch normalization in the decoder is to follow [1]. And decoder is not the mean focus of our paper. For all the mesh graph attention blocks, we use 1-ring neighbors for neighboring lookups.

C. Masking Ratio Analysis

In Figure 8, we plot the accuracy curve under different training and test masking ratio. Three patterns of accuracy curve are found when the test masking ratios are fixed.

The first happens at test masking ratios of 0.0, 0.1, and 0.2. The accuracy goes up and down. The second one is at test masking ratios of 0.3, 0.4, 0.5, and 0.6. The accuracy goes up and done and up again. The last one happened at test masking ratios of 0.7, 0.8, and 0.9. The accuracy goes up. The reason is straightforward for the first and third patterns. For the first pattern, the models are trained with low masking ratios. When the training masking ratio increases, the models focus on extracting information other than just masking, which explains why there is an increasing curve at the beginning. And when the ratio is too high, there is not enough information. Thus, the curve begins to drop.

The third pattern is caused by test masking ratios being too high such that the models trained with low masking ratios could efficiently capture the information of testing meshes. And only models trained under high masking ratios could capture information from testing meshes.

The second pattern is generated when the first and third patterns merge.

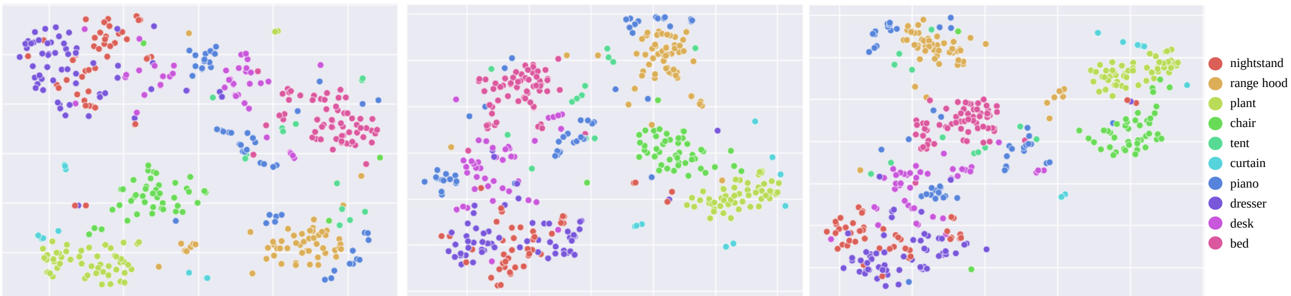
One pattern of accuracy curve is found when the training masking ratios are fixed.

D. Segmentation Visualization Results

More segmentation results are shown in Figure 9

E. Reconstruction Visualization Results

More reconstruction results are shown in Figure 10.



(a) Train mask ratio 0.3 (b) Train mask ratio 0.6 (c) Train mask ratio 0.9

Figure 6. **Visualization of t-SNE on ModelNet40.** We fix the test masking ratio to 0 and choose 10 random classes for rendering the figures.

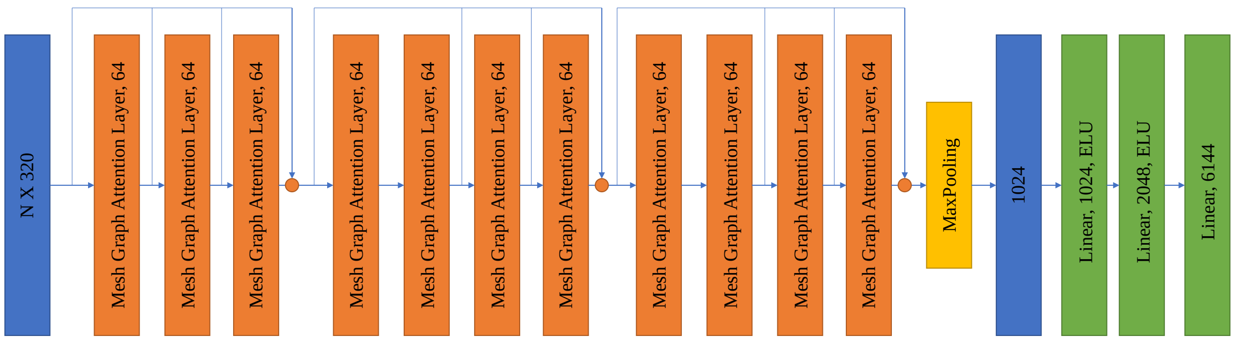


Figure 7. **Neural Network Architecture.** N stands for the number of face nodes. The number after each layer stands for the dimension of the output embedding. The orange dot stands for concatenating the forward embeddings from previous layers. Each mesh graph attention layer has batch normalization and ReLU layer following.

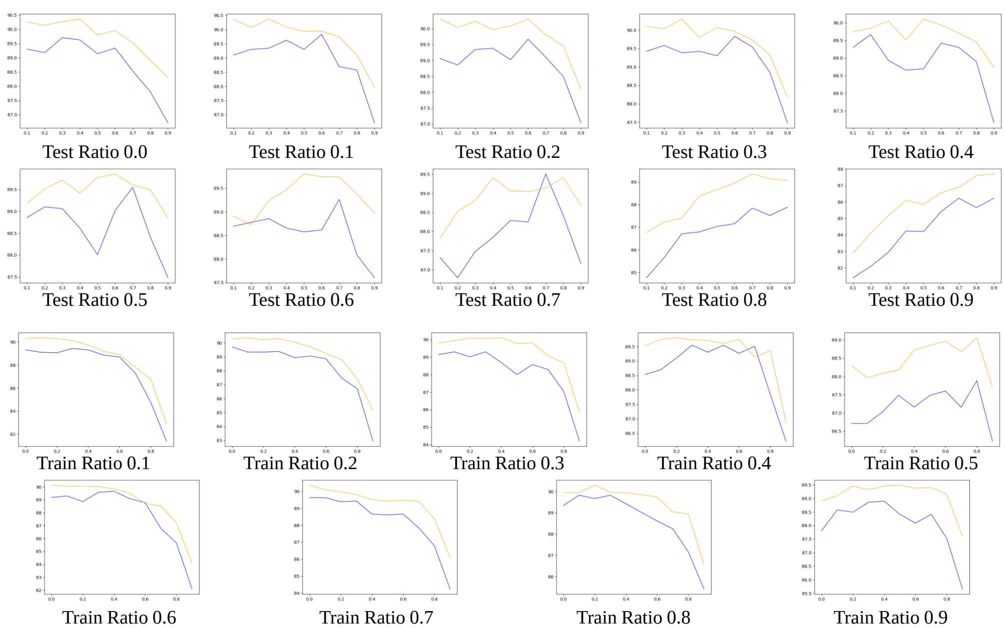


Figure 8. **Masking ratio's effect on accuracy curves.** The validation accuracy curve (in yellow) and test accuracy curve (in blue) are plotted by fixing different test and training masking ratios.

Figure 9. **Segmentation results.**

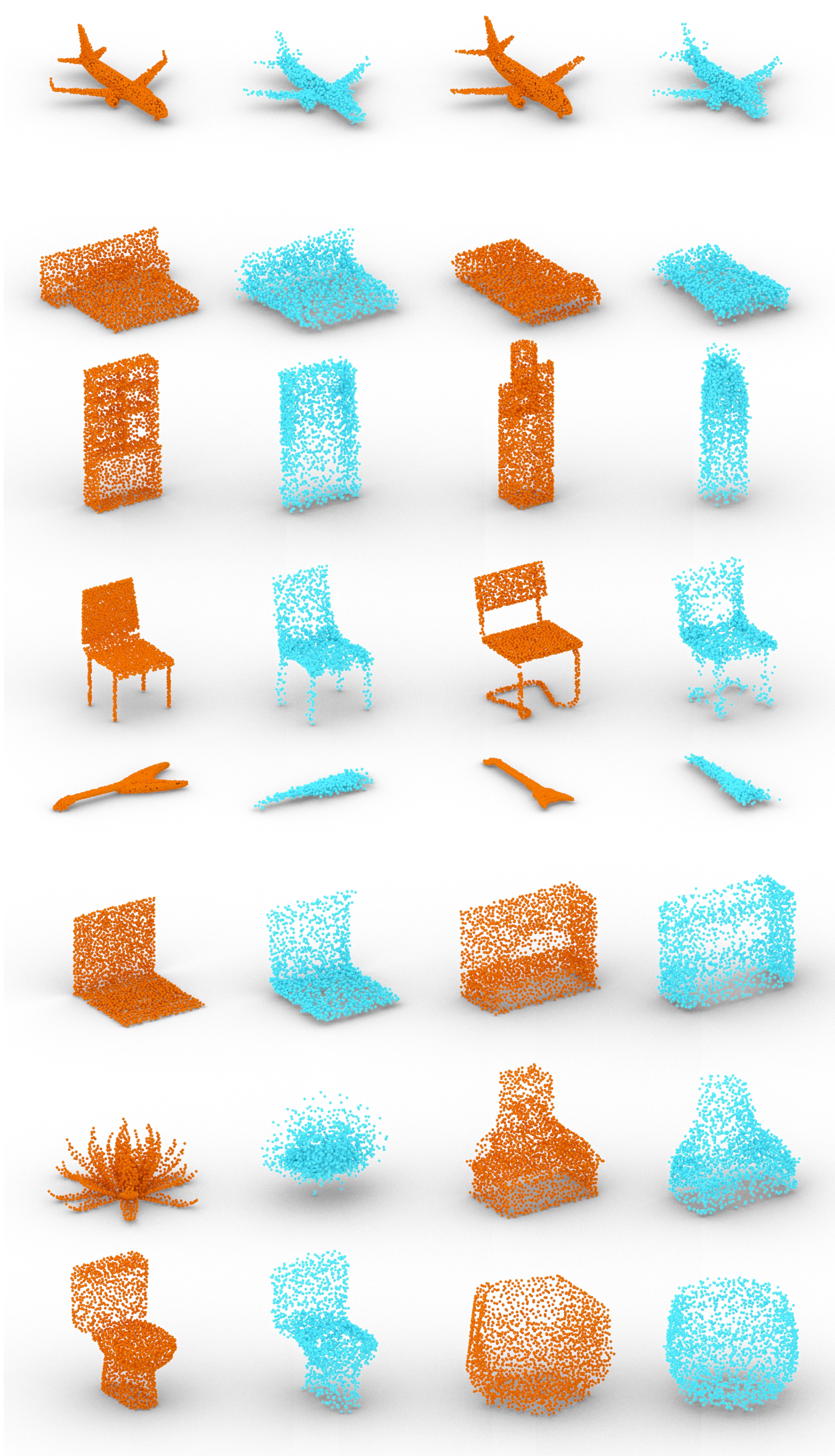


Figure 10. **Reconstruction results.** Orange represents the ground truth point cloud, and blue depicts the reconstructed point cloud after masking 40% of the input mesh faces.