3D Point Capsule Networks Supplementary Material

Yongheng Zhao [†] [✣] [•] Tolga Birdal [†] [*] [•] Haowen Deng [†] [*] Federico Tombari [†]
[†] Technische Universität München, Germany [✣] University of Padova, Italy
[*] Siemens AG, München, Germany

This document supplements our paper entitled **3D Point-Capsule Networks** by providing further quantitative and qualitative insights into the results.

1. Semi-supervised Classification

We begin by showing semi-supervised classification results in Tab. 1. Note that our network can generate predictions that are on par with or better than FoldingNet [1].

Table 1. Part segmentation on ShapeNet-Part by learning on limited training data. The table shows the accuracies obtained by FoldingNet [1] and our approach for different amount of training data.

	1%	2%	5%	20%	100%
FoldingNet	56.15	67.05	75.97	84.06	88.41
Ours	59.24	67.67	76.49	84.48	88.91

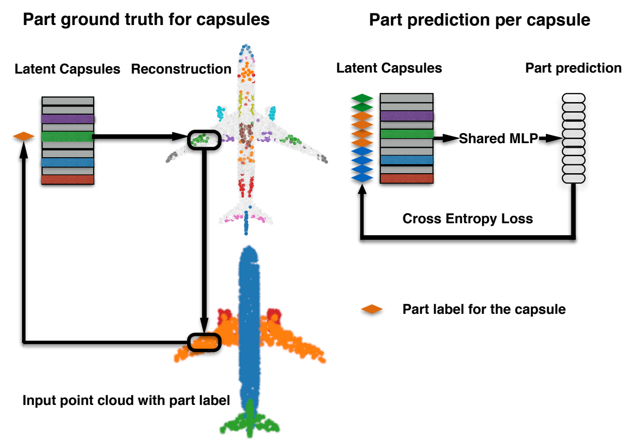

Figure 1. Supervising the 3d point capsule networks for part prediction. Instead of performing a point-wise part labeling, we use a capsule-wise association requiring less data annotation efforts.

2. Part Segmentation

We first give a small summary of the part association network for optional supervision. The input to this one-layer architecture is the latent capsules combined with one-hot vector of the object category. The output is the part prediction of each capsule. We use the cross entropy loss as our loss function and Adam as the optimizer with the learning rate of 0.01. The network structure is shown in Fig. 1.

Then we utilize the pre-trained decoder to reconstruct the object with the labeled capsules. Fig. 3 depicts further visualizations for different objects from the ShapeNet-Part dataset [2]. Our results are also qualitatively comparable to ground truth.

3. Part Interpolation

We first show an overview of how we perform part interpolation. While this part has been thoroughly explained in the paper, we have omitted this architecture illustration due to space considerations. We now provide this in Fig. 4.

Next we show, the part interpolation results on different objects. In this qualitative evaluation, we are given two shapes and the goal is to interpolate the source part towards the target. To do that we find the matching capsules that represent the part of interest in both shapes. We then linearly interpolate from the capsule(s) of the source to the one(s) of the target. This generates visually pleasing intermediate shapes, which our network has never seen before. Here we see that the learned embedding resemble a Euclidean space where linear latent space arithmetic is possible. It is also visible that such interpolation scheme can handle topological changes such as merging or branching legs. In the end of interpolation a new shape is generated in which the part is replaced completely with the target's. That brings us to our second and interesting application, part replacement.

[•] *First two authors contributed equally to this work.*

4. Part Replacement

We now supplement our paper by presenting additional qualitative results on the task of part replacement. Fig. 6 shows numerous object pairs where a part-of-interest is selected in both and exchanged by the help of latent space capsule arithmetic. Analogous to the ones in the paper we also show a cut-and-paste operation that is a mere exchange of the parts in 3D space, obviously resulting in undesired disconnected shapes. Thanks to our decoder's capability in generating high fidelity shapes, our capsule-replacement respects the overall coherence of the resulting point cloud.

5. Ablation Study

In order to show the prosperity of the dynamic routing, we compare the reconstruction result by replacing the DR with PointNet-like set of convolutional layers. In this ablation study, the primary point capsules (1024×16) are considered as 1024 point-features and each point has the feature dimension of 16. We utilize a shared MLP to increase the feature dimension from 16 to 64. After conducting max pooling, we can obtain a vector of length 64. With multiple MLPs and max-pooling, we are able to generate 64 vectors which have the same dimensions as the latent capsules produced by dynamic routing. The structure of this comparison module is shown in Fig. 2. To carry out our fair evaluation, we re-train the whole AE with this module. The result of the reconstruction is shown in Fig. 5 of the main paper.

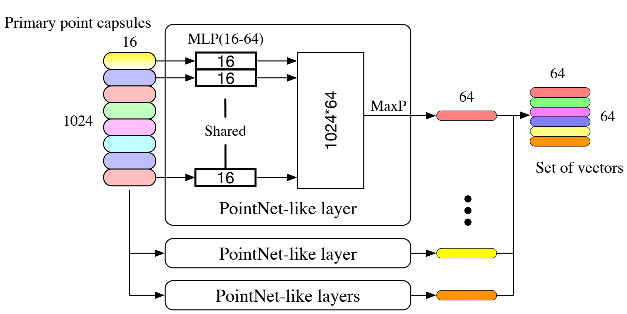

Figure 2. The structure of the comparison module that operates on the primary point capsules and generates a set of vectors having the same dimensionality as the latent capsule output of DR.

6. A Discussion on the Local Spatial Attention

Our network consists of multiple MLPs acting on a single capsule. It encodes the part information inside that capsule rather than the MLPs themselves. For that reason, the local attention stems from both the organization of primary point capsules (in our case obtained by dynamic routing) and potentially the decoder (see Fig. 5 of the main paper). Thus, we are able to control and represent the shape instantiation in the latent space as shown in *part interpolation/replacement* evaluations. Contrarily, AtlasNet reconstructs different local patches with different MLPs from the same latent vector. This embeds the part knowledge into the MLPs, making it challenging to control the shape properties.

References

[1] Y. Yang, C. Feng, Y. Shen, and D. Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[2] L. Yi, V. G. Kim, D. Ceylan, I. Shen, M. Yan, H. Su, C. Lu, Q. Huang, A. Sheffer, L. Guibas, et al. A scalable active framework for region annotation in 3d shape collections. *ACM Transactions on Graphics (TOG)*, 2016.

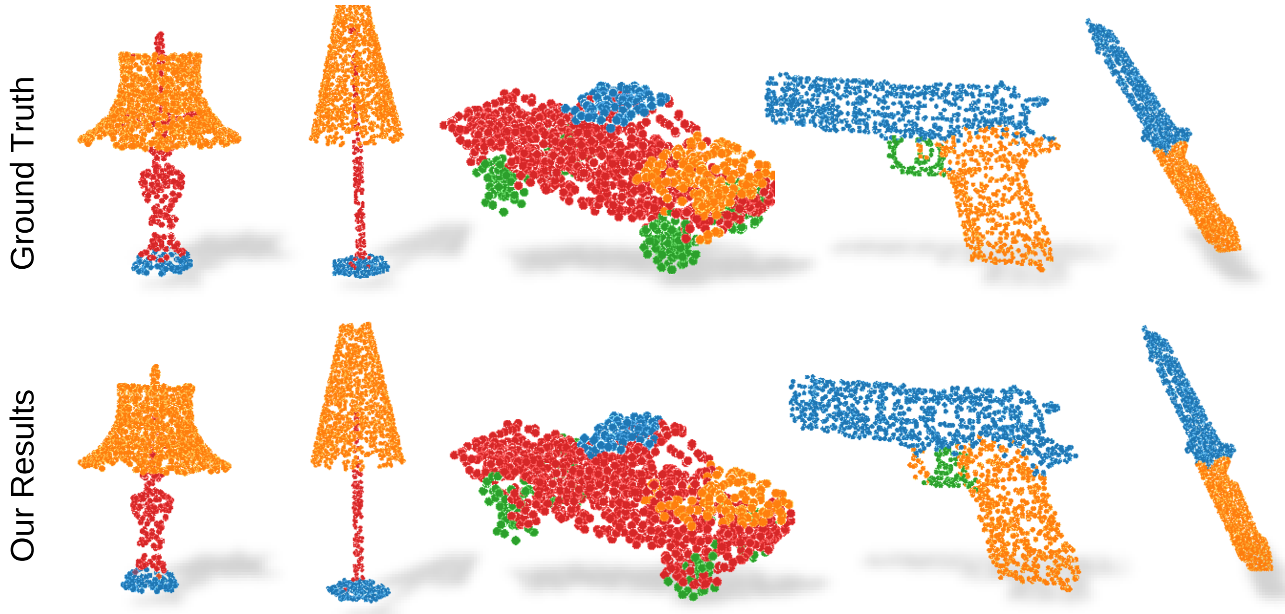

Figure 3. Part segmentation on limited amount of training data.

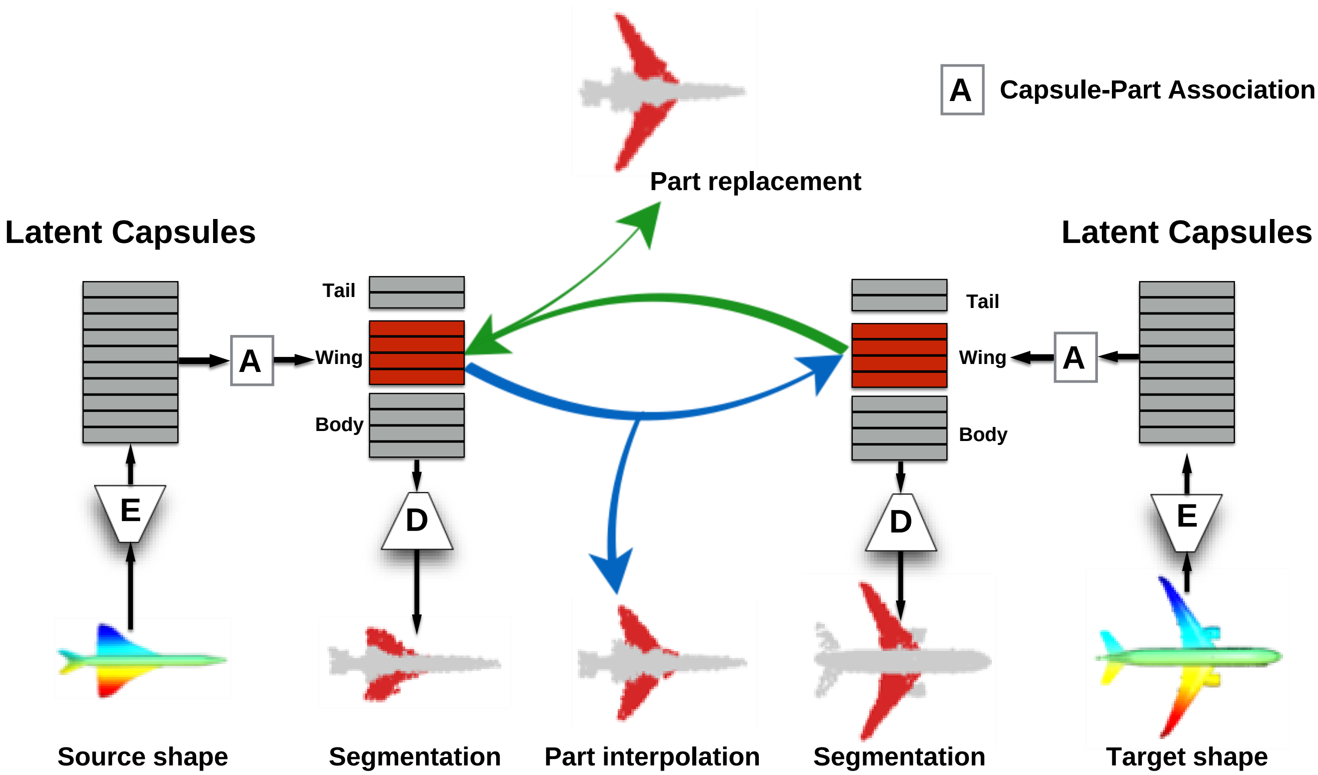

Figure 4. Our interpolation / replacement pipeline.

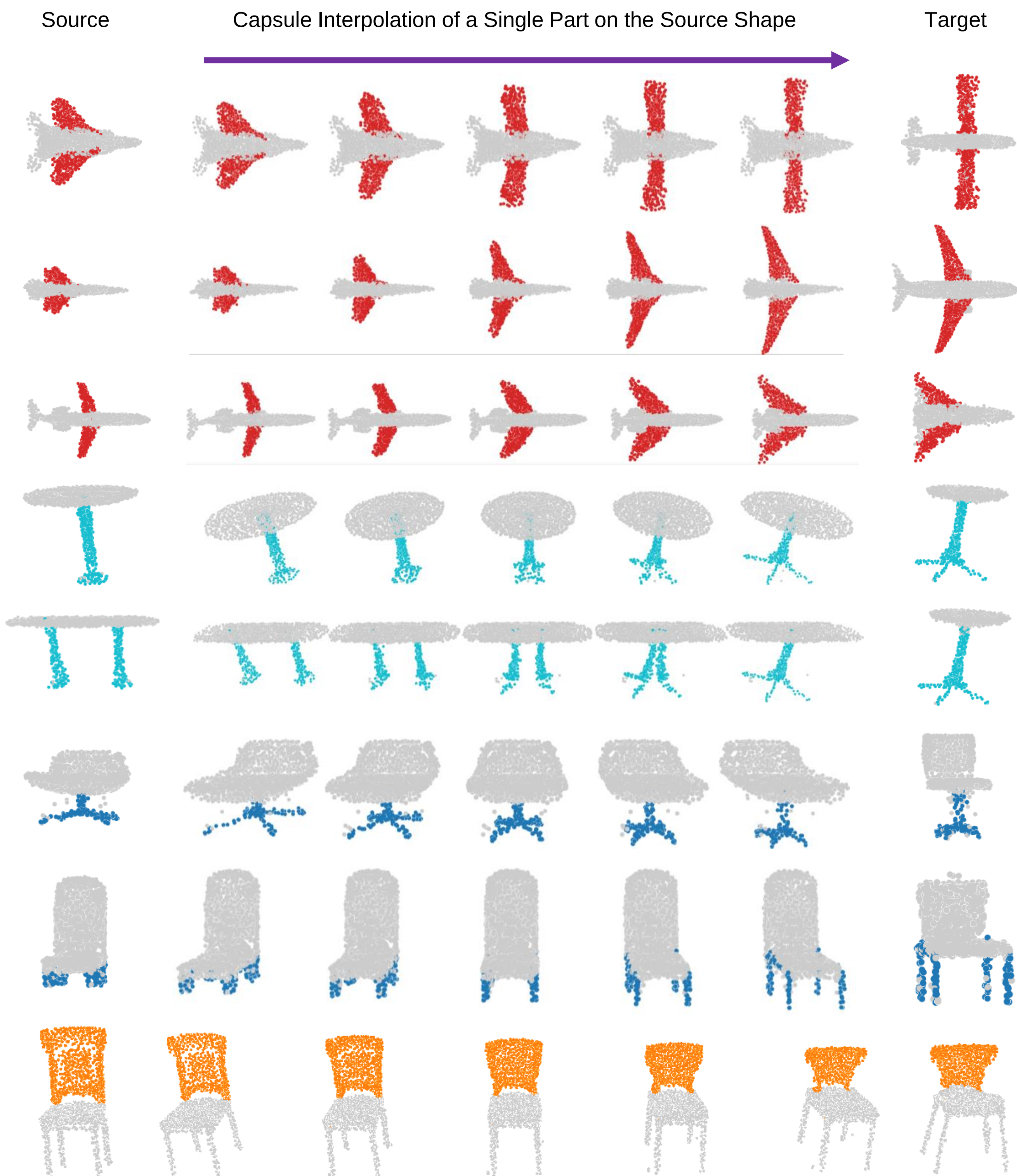

Figure 5. Visualization of part interpolation from source shape part to target. By simple linear interpolation on the correspondent capsule(s), smooth intermediate topologies could be generated.

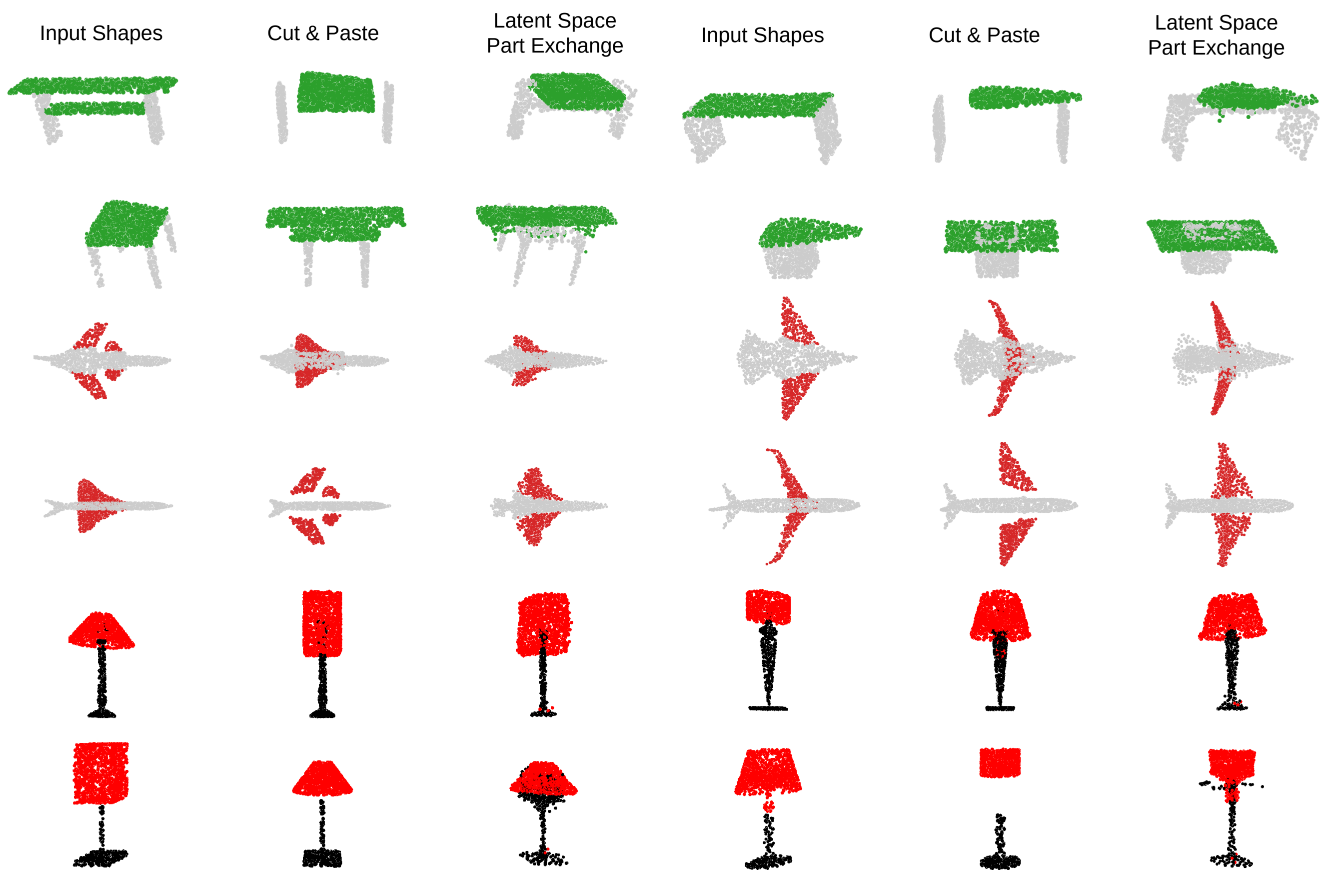

Figure 6. Part replacement visualization and comparison. By operating in the latent space, more natural replacement results could be obtained, without suffering from the detachment problems as with simple Cut & Paste method.