Supplemental Materials of DPF-Net: Combining Explicit Shape Priors in Deformable Primitive Field for Unsupervised Structural Reconstruction of 3D Objects

Qingyao Shuai, Chi Zhang, Kaizhi Yang, Xuejin Chen
University of Science and Technology of China

1. Implementation Details

We now introduce the network details of each module in our DPF-Net.

Global feature projection MLP. After extracting the global feature from the input volume, we project the global feature to the part features. We first feed it into a weight-sharing one-layer MLP to obtain a $M \times n$ dimension embedding, then we divide this embedding into M features equally to obtain the part features $\{\mathbf{F}_i \in \mathbb{R}^n, i = 1, 2, \ldots, M\}$, in our experiments the dimension of global feature and part feature is 128.

Parameter Generation MLP f_{para}. To generate primitive parameters, we decode its part feature using a weight-sharing parameter generation MLP f_{para} which consists of five MLPs. For part i, the part feature is first fed into a three-layer MLP for high-level feature embedding extraction, then four one-layer MLPs are used to decode primitive parameters $\{\mathbf{R}_i, \mathbf{t}_i, \mathbf{s}_i, \rho_i\}$ respectively. In particular, the three-layer MLP is implemented with hidden layer size of 128 and no bias, output dimension 128. The bias of the four one-layer MLPs of $\mathbf{t}, \mathbf{s}, \rho$ is initialized as zeros.

HyperNet Φ and DeformNet D. To model the geometry detail of each part, we use a primitive deformation module (PDM) to deform the parameterized primitive fields. The PDM consists of a HyperNet Φ and a DeformNet D, both of which is adopted from [2], but we use different network settings. To decrease the parameters of our network, we apply a HyperNet Φ with one hidden layer and DeformNet D with two hidden layers of size 128. After each layer, we add the ReLU activation function except for the last layer.

2. Additional Experimental Results

Table 1. Quantitive results of using different maximum primitive number M.

	table		chair		airplane	
Method	CD	m-IoU	CD	m-IoU	CD	m-IoU
$M = 8$	0.4756	**91.3**	0.3918	83.6	0.1941	**66.0**
$M = 12$	0.4588	90.8	0.3702	83.8	0.1952	65.7
$M = 16$	0.4467	90.5	0.3633	**84.3**	0.1914	65.4
$M = 20$	**0.4451**	90.7	**0.3571**	83.9	**0.1903**	65.5

Options of the Maximum Primitive Number M. In our paper, we set the maximum primitive number M empirically for different categories. To evaluate the effect of M, we fix the other hyper-parameters and train several models with different M. The quantitative comparisons are shown in Table 1. For all categories, increasing M leads to better reconstruction results in terms of chamfer distance since more primitives can better fit the target shape and the primitive deformation module could focus more on the shape details. As expected, in the 'table' and 'airplane' categories, the m-IoU decreases with larger M

since the finer division of object parts undermines the structural consistency. In the 'chair' category, the m-IoU is the highest when $M = 16$ and then decreases as M increases, due to that chairs usually contain more parts. Based on the results in Table 1, we choose $M = 8$ for 'table' and 'airplane' and $M = 16$ for the 'chair' for better structure extraction.

Additional Structural Reconstruction Results We provide additional experimental results on the unsupervised structural reconstruction task on the 'table', 'chair', and 'airplane' categories in the ShapeNet dataset [1]. We set the maximum primitive number for table $M = 8$, chair $M = 16$, and airplane $M = 8$. The qualitative comparisons of the reconstruction results of these three categories are shown in Figure 1, 2 and 3 respectively. Compared to the primitive-based method CA [4] and implicit function-based method RIM-Net [3], we can observe that our DPF-Net shows consistent structural partition and more detailed reconstruction in the 'chair' and 'table' categories. For the 'airplane' category, our DPF-Net reconstructs more details and achieves comparable in structure partition performance with the other methods.

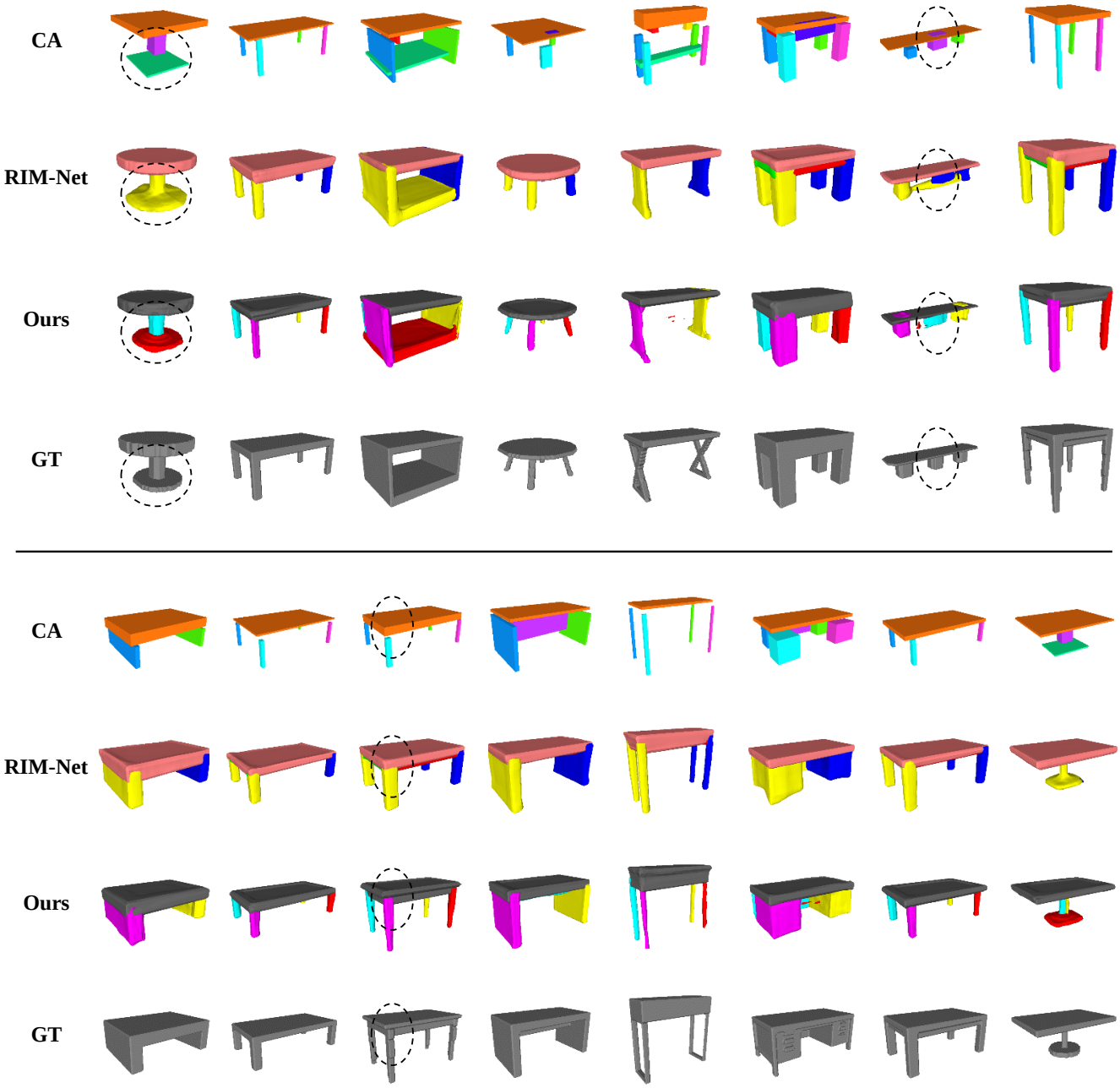

Figure 1. Qualitative comparison of unsupervised structural reconstruction methods on the 'table' category.

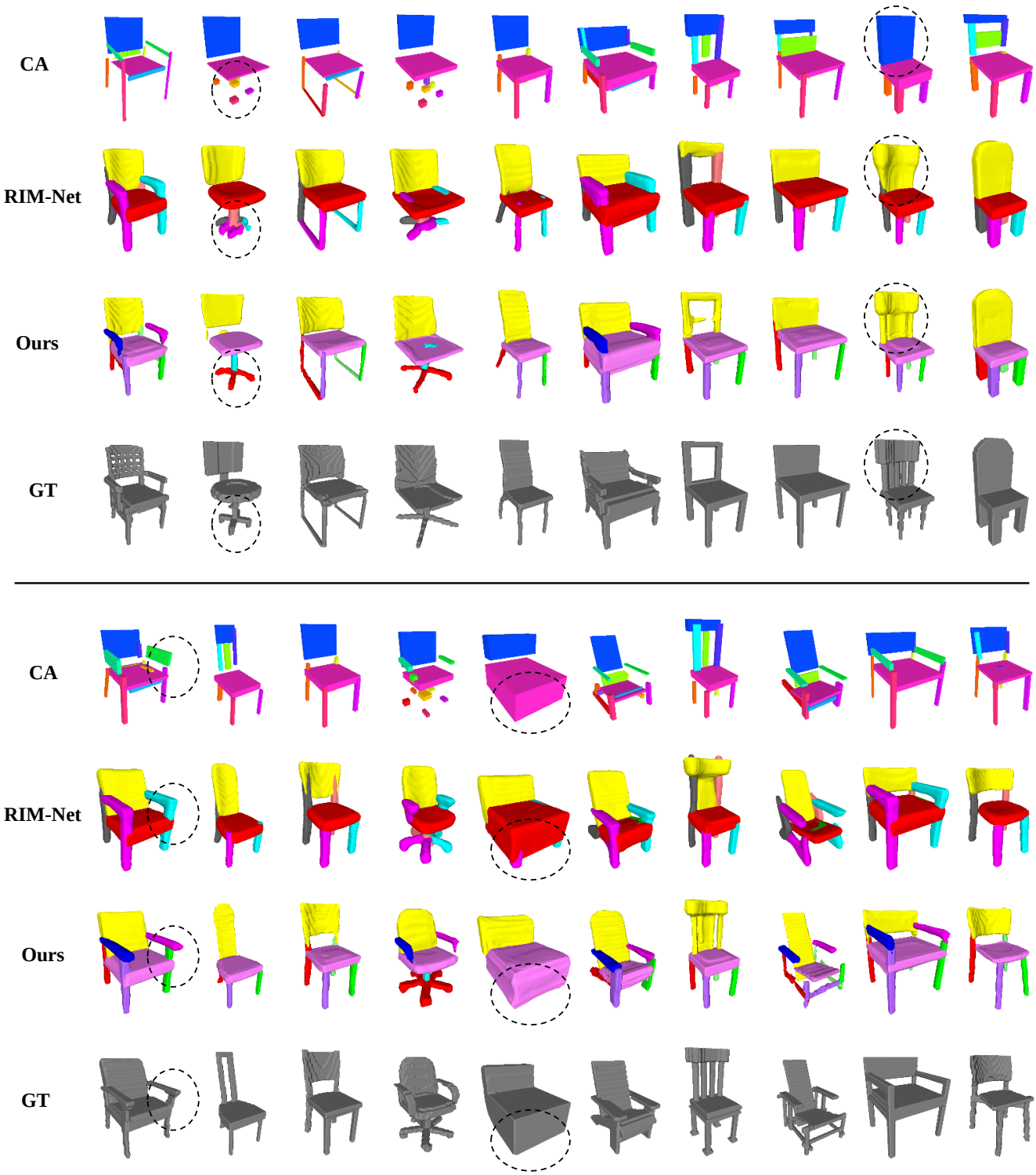

Figure 2. Qualitative comparison of unsupervised structural reconstruction methods on the 'chair' category.

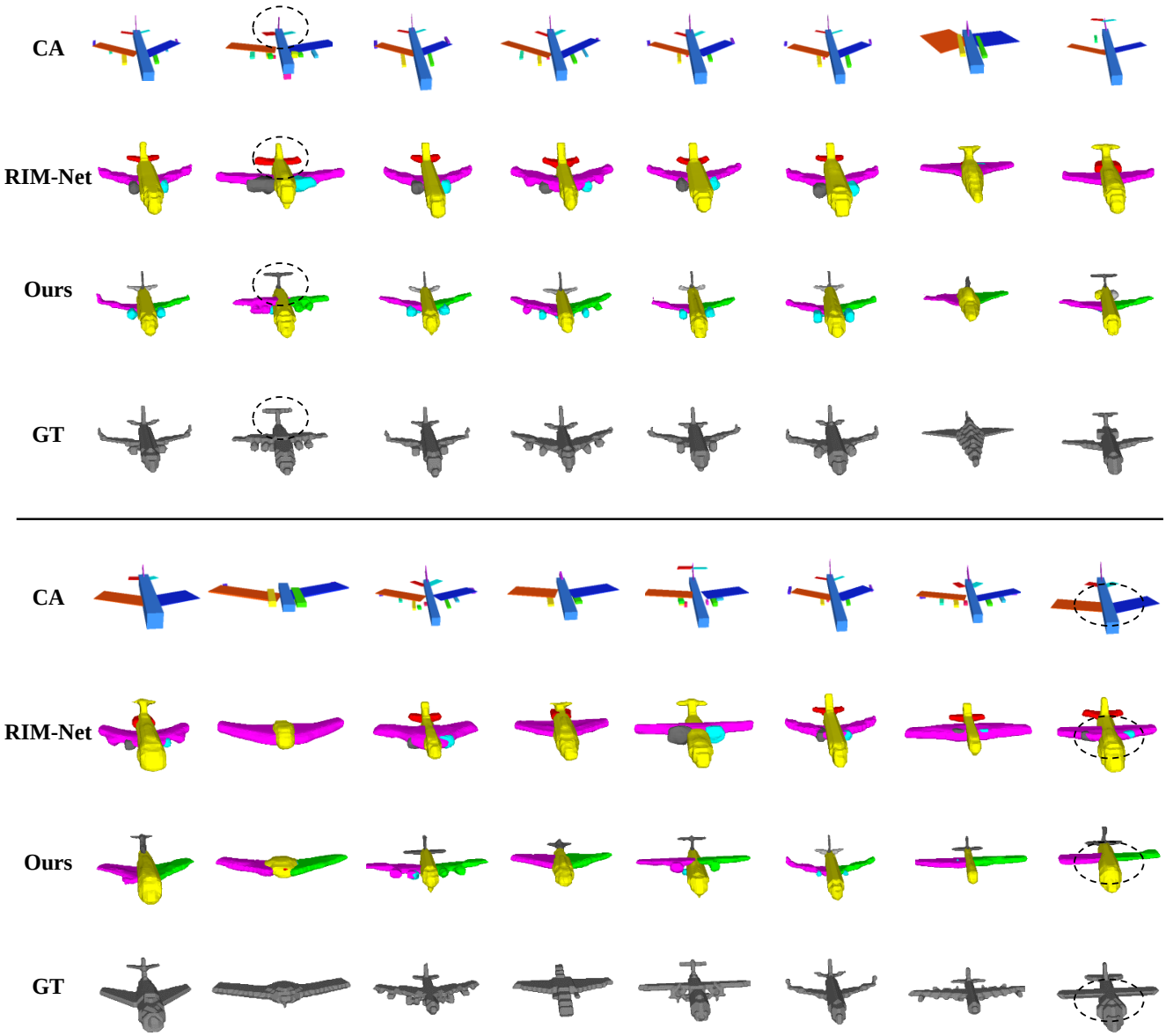

Figure 3. Qualitative comparison of unsupervised structural reconstruction methods on the ‘airplane’ category.

References

[1] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[2] Yu Deng, Jiaolong Yang, and Xin Tong. Deformed Implicit Field: Modeling 3D shapes with learned dense correspondence. In *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2021.

[3] Chengjie Niu, Manyi Li, Kai Xu, and Hao Zhang. RIM-Net: recursive implicit fields for unsupervised learning of hierarchical shape structures. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022.

[4] Kaizhi Yang and Xuejin Chen. Unsupervised learning for cuboid shape abstraction via joint segmentation from point clouds. *ACM Trans. Graph.*, 40(4), 2021.