D^{2}IM-Net: Learning Detail Disentangled Implicit Fields from Single Images Supplementary Material

Manyi Li Hao Zhang
Simon Fraser University

1. Implementation

In the implementation of D^{2}IM-Net, we take ResNet18 as our encoder to obtain the global feature and local feature map from the input image. The base decoder is an MLP with the architecture of IMNET [1]. The detail decoder follows the network in [4] to predict the two displacement maps from the local feature map. As for D^{2}IM-Net$_{GL}$, we take DISN [3] as the base decoder with both their global decoder and local decoder.

In the Laplacian computation, in order to balance the three loss terms, we scale the predicted and ground-truth derivitaves by the same factor with respect to $\frac{\partial u(p)}{\partial p'}$. Therefore, the Laplacian loss becomes (see Section 3.3 in the main paper for the denotations)

$$
\begin{aligned}
L_{lap} &= \frac{1}{|P_F|} \sum_{p_i \in P_F} \| \triangle f'_{DF}(u(p_i)) - l'(u(p_i)) \|_2^2 \\
l'(u(p)) &= \frac{N(u(p))}{\partial u_x} + \frac{N(u(p))}{\partial u_y} \\
\triangle f'_{DF}(u(p)) &= \frac{f_{DF}(u(p))}{\partial^2 u_x} \cdot \frac{\partial u_x}{\partial p'_x} + \frac{f_{DF}(u(p))}{\partial^2 u_y} \cdot \frac{\partial u_y}{\partial p'_y}.
\end{aligned} \tag{1}
$$

2. Detail transfer results

We present more results of detail transfer between two images. The details on the chairs' backs are transferred from the source images to the target images. Both the source images and target images are from the test set.

For the target images, we show the reconstructions and their part segmentation (axis-aligned bounding box per part) [2] in Figure 1, the detail transfer results in Figure 2. As described in Section 4.4 of our paper, the semantic segmentation of the reconstructed coarse shapes are used to provide the 3D correspondence for the transfer. Note that one can also interactively tune the bounding boxes to refine the transferred details.

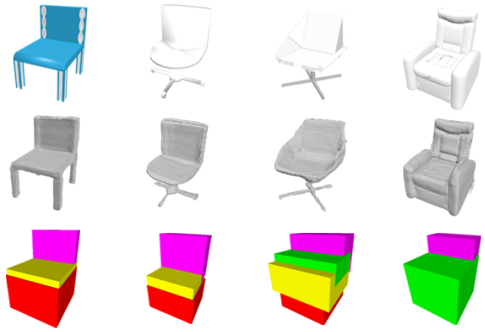


Figure 1. The reconstructions (middle row) and part segmentation (bottom row) of the input images (top row). The images are used as the target images in Figure 2.

3. More evaluations

Figure 3, 4, 5, 6, 7 show more qualitative results of single-view reconstruction. We mainly show the reconstructions of categories with clear details, such as chairs, sofas, cabinets, speakers. The results of D^{2}IM-Net recover the details while preserving the flatness of the other regions, which is preferred in the reconstruction scenarios.

As mentioned in the paper, the above comparisons focus on testing images whose views exhibit most geometric details to show the strength of our method. In addition, we present table 1 to show the comparison when the testing images are taken from arbitrary viewpoints. Our method still outperforms the others, while we can observe an across-board performance drop due to many more self-occlusions with objects captured from arbitrary views.

References

[1] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[2] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 2019. 1

[3] Qiangeng Xu, Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann. Disn: Deep implicit surface network

Figure 2. More results of detail transfer between semantic parts (chairs' backs). Top row: source image to provide details; Left column: target image to provide coarse shapes. Two views of each transferred reconstruction are shown.

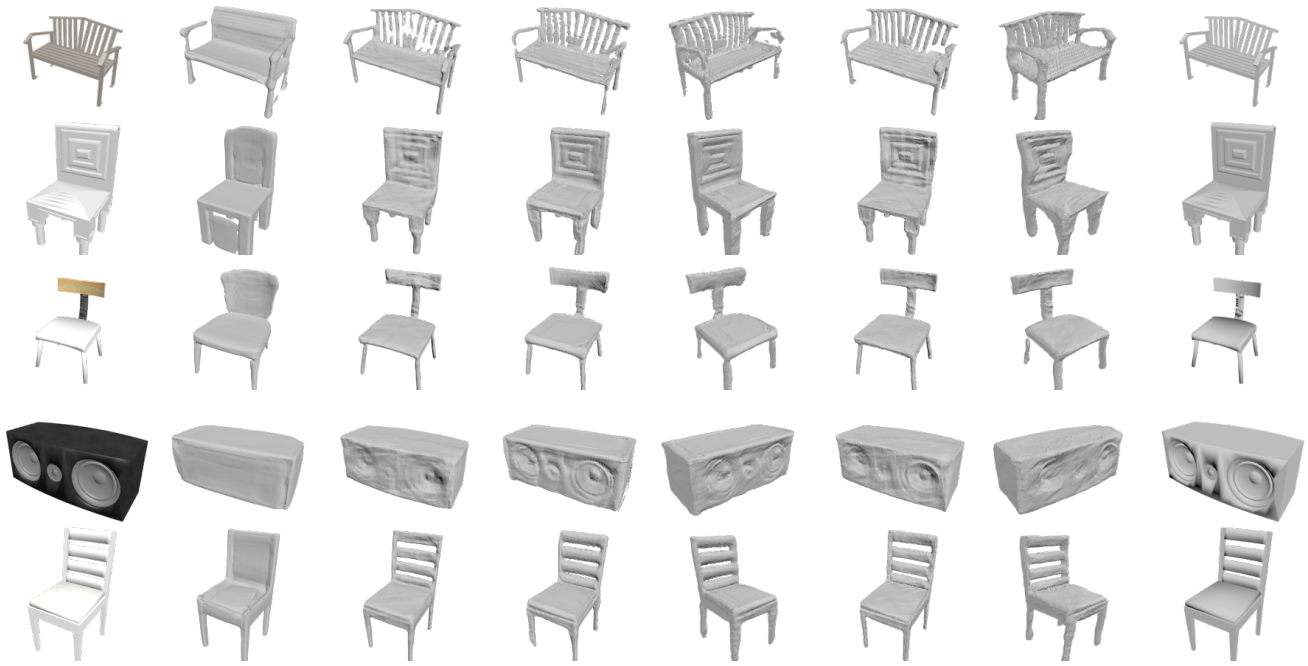
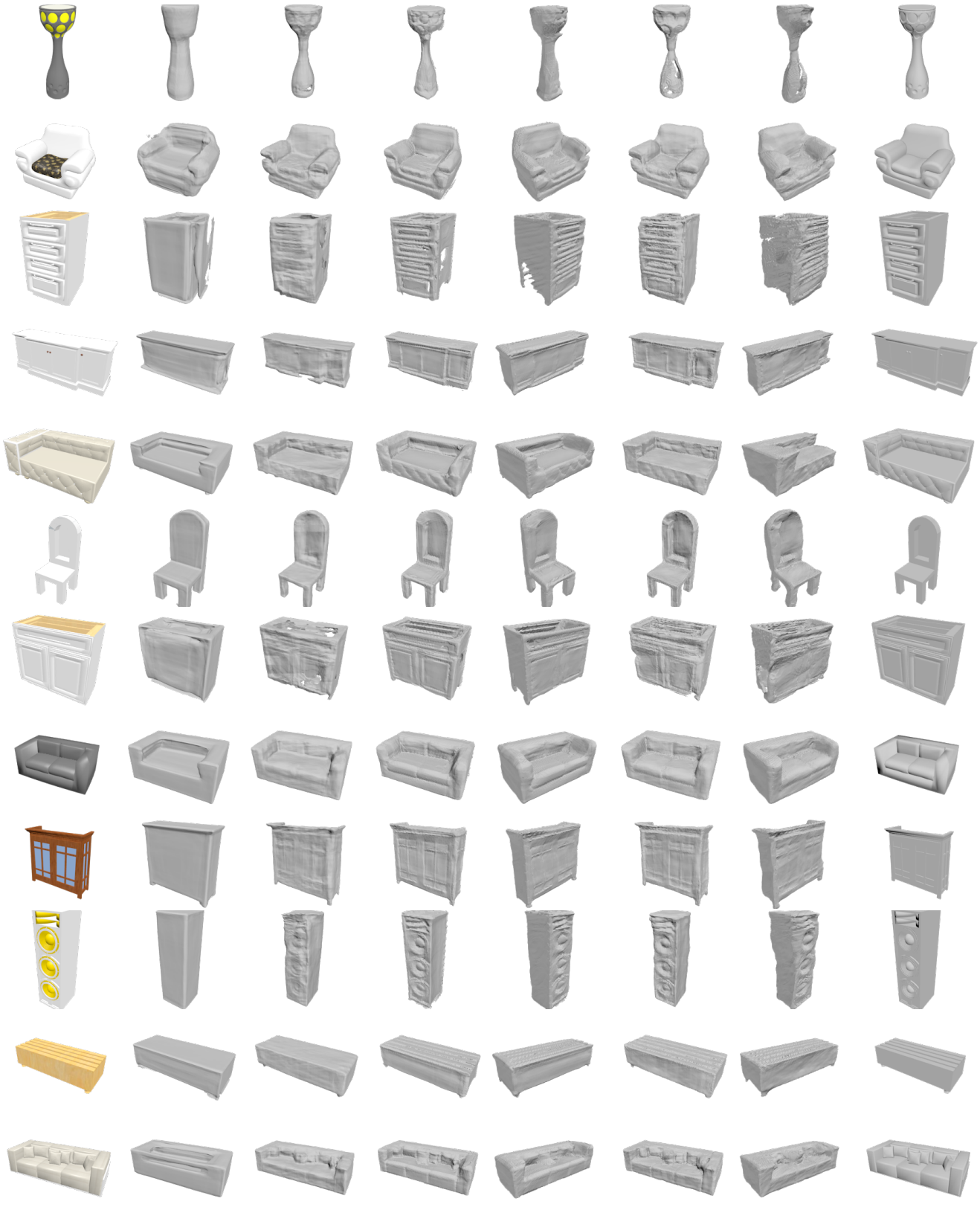


Figure 3. More qualitative results. Two views of D^{2}IM-Net and D^{2}IM-Net$_{GL}$ are presented to show the reconstruction and recovered details.

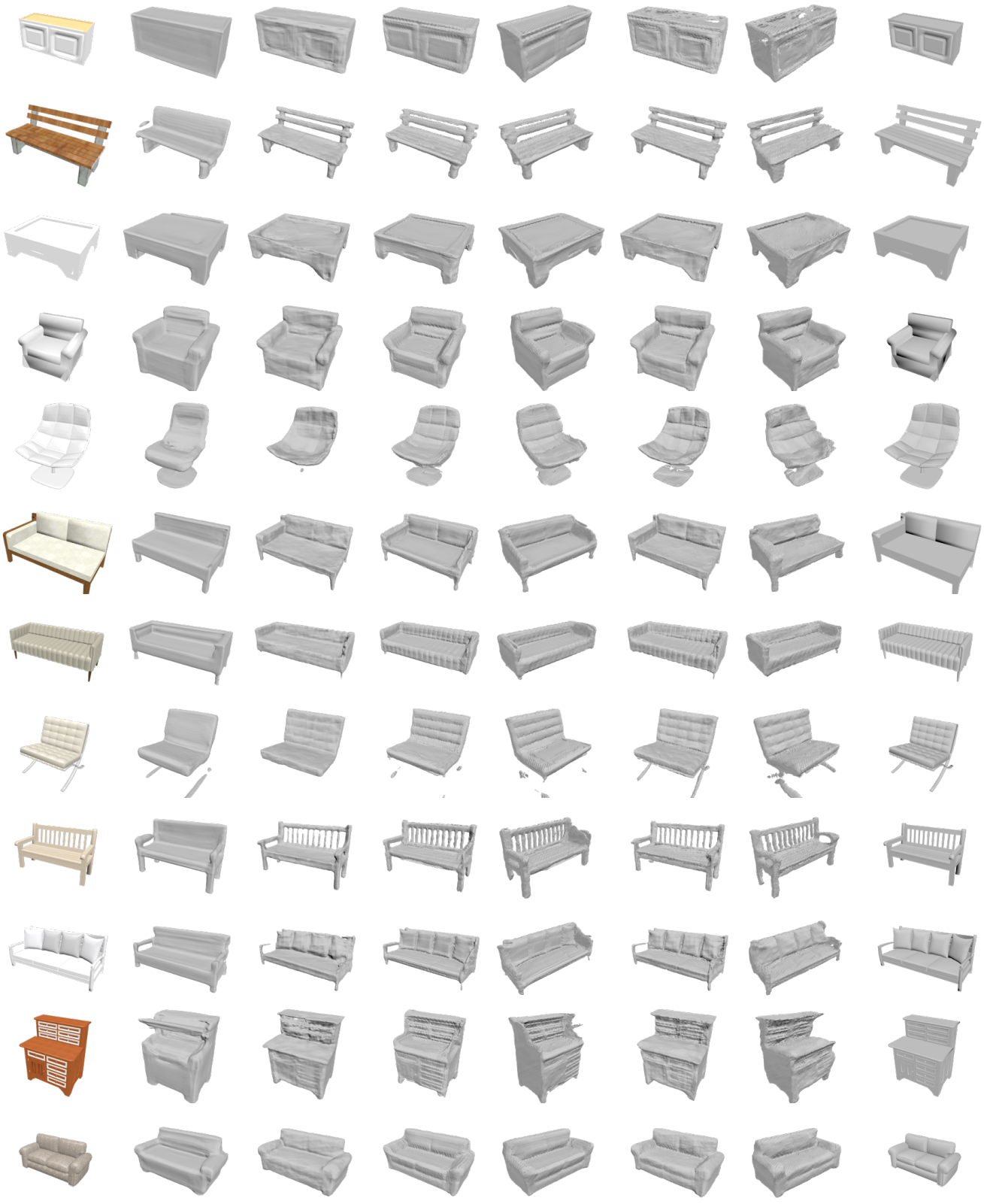
for high-quality single-view 3d reconstruction. In *NeurIPS*, pages 492–502, 2019. 1

[4] Zehao Yu, Jia Zheng, Dongze Lian, Zihan Zhou, and Shenghua Gao. Single-image piece-wise planar 3d reconstruction via associative embedding. In *CVPR*, pages 1029–1037, 2019. 1



(a) Input (b) IMNET (c) DISN (d) D²IM-Net (e) D²IM-Net$_{GL}$ (f) GT

Figure 4. More qualitative results. Two views of D^{2}IM-Net and D^{2}IM-Net$_{GL}$ are presented to show the reconstruction and recovered details.



(a) Input (b) IMNET (c) DISN (d) D²IM-Net (e) D²IM-Net$_{GL}$ (f) GT

Figure 5. More qualitative results. Two views of D^{2}IM-Net and D^{2}IM-Net$_{GL}$ are presented to show the reconstruction and recovered details.

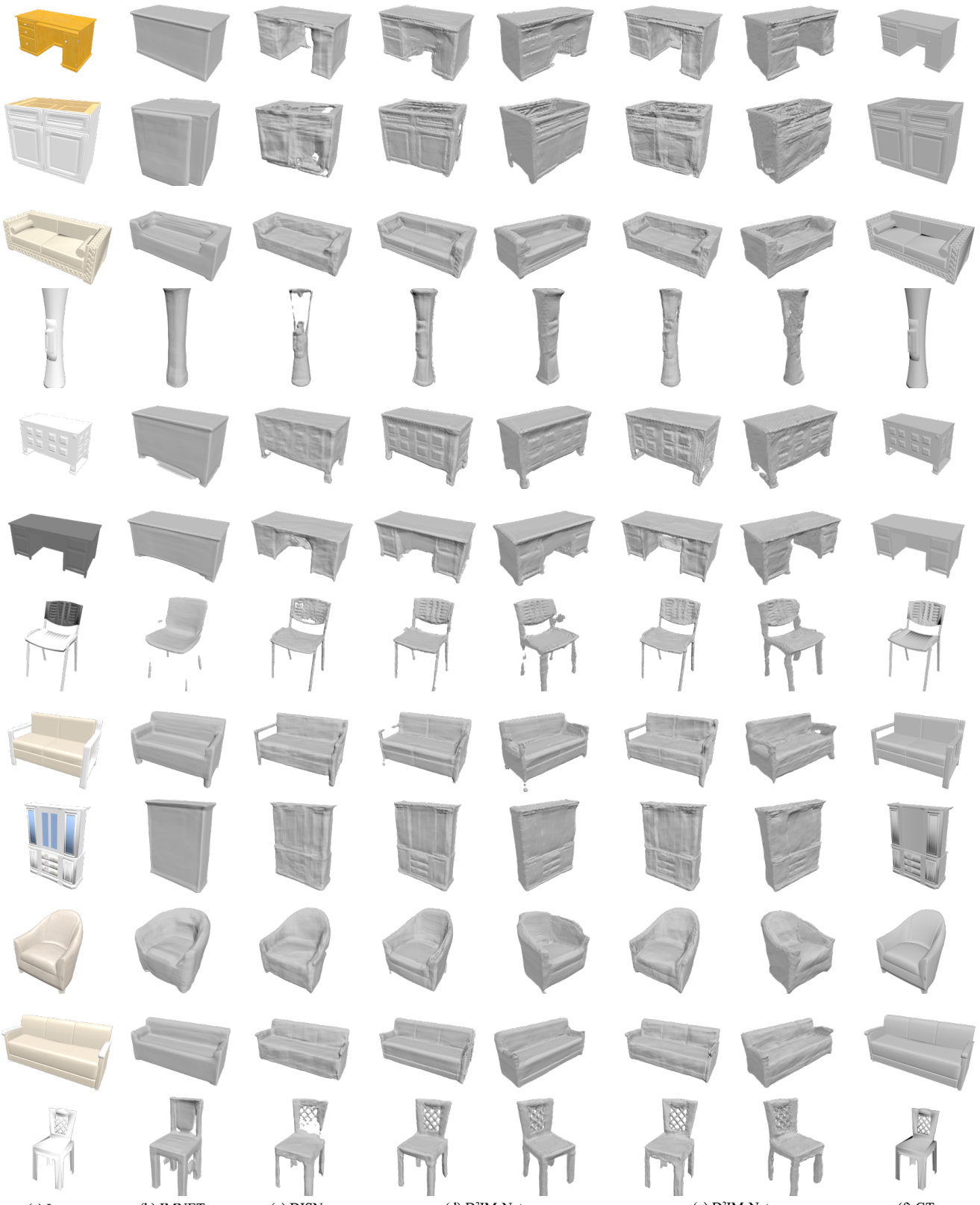


Figure 6. More qualitative results. Two views of D^{2}IM-Net and D^{2}IM-Net$_{GL}$ are presented to show the reconstruction and recovered details.

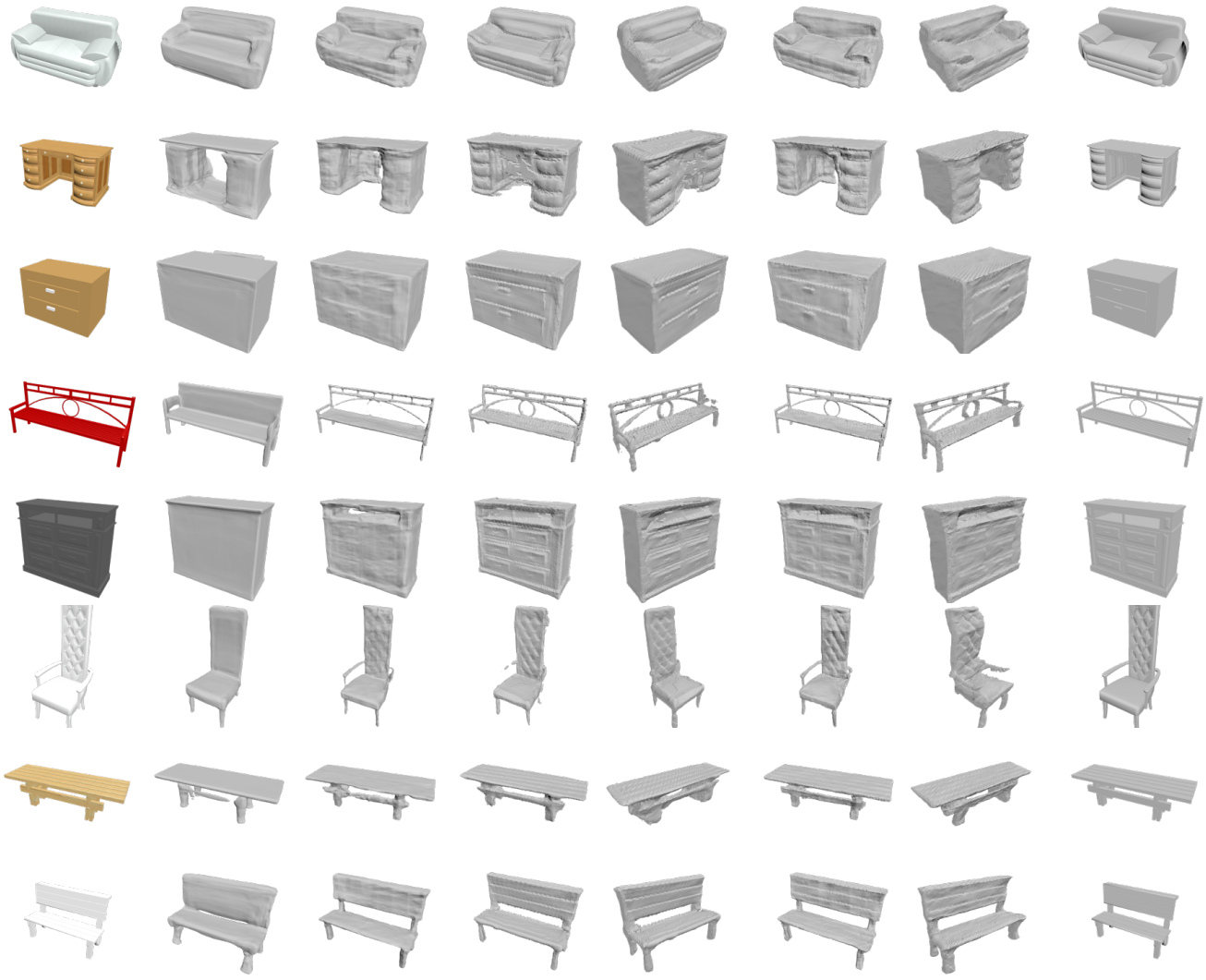


Figure 7. More qualitative results. Two views of D^{2}IM-Net and D^{2}IM-Net$_{GL}$ are presented to show the reconstruction and recovered details.

		plane	bench	box	car	chair	display	lamp	speaker	rifle	sofa	table	phone	boat	Mean
IoU ↑	IMNET	0.5387	0.5044	0.4375	0.7707	0.5369	*0.4898*	0.4275	0.4833	0.5780	0.5303	0.4942	*0.7214*	0.5734	0.5451
	DISN	0.5527	0.5131	0.4509	*0.7942*	0.5518	0.4848	0.3903	0.4723	**0.6277**	0.6031	0.5269	0.6793	**0.6292**	0.5597
	D^{2}IM-Net	*0.5645*	**0.5389**	**0.5012**	0.7919	*0.5523*	**0.5386**	*0.4331*	*0.5248*	0.6016	**0.6384**	*0.5292*	**0.7535**	0.6092	**0.5829**
	D^{2}IM-Net$_{GL}$	**0.5687**	*0.5241*	*0.4942*	**0.7961**	**0.5648**	0.4876	**0.4626**	**0.5392**	*0.6028*	*0.6211*	**0.5342**	0.6818	*0.6181*	*0.5766*
CD ↓	IMNET	0.0406	0.0388	0.0489	0.0415	0.0374	*0.0428*	0.0611	0.0623	0.0329	0.0491	0.0428	0.0329	0.0436	0.0442
	DISN	0.0371	0.0396	0.0411	**0.0329**	0.0341	0.0515	0.0786	0.0604	0.0315	**0.0386**	*0.0358*	0.0325	*0.0361*	0.0423
	D^{2}IM-Net	*0.0349*	**0.0329**	*0.0394*	0.0361	*0.0338*	**0.0402**	*0.0522*	**0.0561**	**0.0282**	0.0410	0.0378	**0.0260**	0.0371	**0.0381**
	D^{2}IM-Net$_{GL}$	**0.0347**	*0.0357*	**0.0390**	*0.0334*	**0.0316**	0.0460	**0.0493**	**0.0561**	*0.0305*	*0.0391*	**0.0342**	*0.0308*	**0.0347**	**0.0381**
ECD-3D ↓	IMNET	0.0746	0.0693	0.0865	0.0874	0.0654	0.0795	0.1034	0.1066	0.0719	0.0804	0.0712	0.0771	0.0795	0.0810
	DISN	0.0650	0.0627	0.0723	0.0693	0.0604	0.0864	0.1131	0.1080	0.0537	*0.0649*	0.0650	0.0779	0.0640	0.0741
	D^{2}IM-Net	**0.0521**	**0.0472**	**0.0677**	**0.0684**	*0.0543*	**0.0644**	*0.0864*	**0.0917**	**0.0395**	0.0678	*0.0640*	**0.0627**	*0.0553*	**0.0632**
	D^{2}IM-Net$_{GL}$	*0.0575*	*0.0536*	*0.0716*	*0.0688*	**0.0535**	*0.0732*	**0.0845**	*0.1004*	*0.0461*	**0.0640**	**0.0614**	*0.0721*	**0.0538**	*0.0662*
ECD-2D ↓	IMNET	2.704	2.996	3.102	3.054	2.465	*3.198*	3.854	3.984	2.660	3.093	**2.603**	*2.458*	2.734	2.993
	DISN	2.564	2.444	2.805	2.653	2.015	3.583	4.918	3.514	2.466	2.910	2.949	2.768	2.605	2.938
	D^{2}IM-Net	**2.165**	**2.014**	**2.491**	**2.522**	*1.683*	**2.683**	*3.441*	**3.256**	**2.073**	*2.890*	*2.618*	**2.047**	*2.422*	**2.485**
	D^{2}IM-Net$_{GL}$	*2.193*	*2.235*	*2.558*	*2.602*	**1.621**	3.269	**3.339**	*3.392*	*2.174*	**2.881**	2.795	2.770	**2.282**	*2.624*

Table 1. Quantitative comparison on input images with arbitrary viewpoints. Top numbers are in bold and second place is indicated in italic.