Supplementary Material for
Dense 3D Point Cloud Reconstruction Using a Deep Pyramid Network

Priyanka Mandikal R. Venkatesh Babu
Video Analytics Lab, CDS, Indian Institute of Science, Bangalore, India
priyanka.mandikal@gmail.com, venky@iisc.ac.in

1. Network Architectures

We provide network architecture details for the common image encoder (Table 1) and the point-cloud decoders of PSGN-FC (Table 2), PSGN-ConvFC (Table 3), and our approach (Table 4). Table 4 shows network details for various components of the dense reconstruction stage as well as the multi-stage network. It should be noted that the decoder in DensePCR has a total of 5.2M parameters (all three stages of the hierarchy), while both the variants of PSGN [2] - PSGN-FC and PSGN-ConvFC have nearly triple the number with 17.1M and 13M parameters respectively. It is worth noting that the majority of parameters in our setup are from the initial base point cloud prediction network. The dense reconstruction network has very few parameters (0.043M parameters per every stage of hierarchy). Hence, as we scale up, there will be negligible addition of parameters, making it highly efficient for dense prediction.

S.No.	Layer	Filter Size/ Stride	Output Size
E1	conv	3x3/1	128x128x32
E2	conv	3x3/1	128x128x32
E3	conv	3x3/2	64x64x64
E4	conv	3x3/1	64x64x64
E5	conv	3x3/1	64x64x64
E6	conv	3x3/2	32x32x128
E7	conv	3x3/1	32x32x128
E8	conv	3x3/1	32x32x128
E9	conv	3x3/2	16x16x256
E10	conv	3x3/1	16x16x256
E11	conv	3x3/1	16x16x256
E12	conv	3x3/2	8x8x512
E13	conv	3x3/2	8x8x512
E14	conv	3x3/2	8x8x512
E15	conv	3x3/2	8x8x512
E16	conv	3x3/2	4x4x512
E17	linear	-	512

Table 1: Image Encoder Architecture

S.No.	Layer	Output Size
D1	linear	256
D2	linear	256
D3	linear	16384x3

Table 2: PSGN-FC Decoder Architecture

S.No.	Layer	Output Size
D1	linear	256
D2	linear	256
D3	linear	4096x3

(a) FC branch

S.No.	Layer	Filter Size/ Stride	Output Size
D4	deconv(E16)	5x5/2	8x8x256
D5	conv(E15)	3x3/1	8x8x256
D6	conv(D4+D5)	3x3/1	8x8x256
D7	deconv(D6)	5x5/2	16x16x128
D8	conv(E11)	3x3/1	16x16x128
D9	conv(D7+D8)	3x3/1	16x16x128
D10	deconv(D9)	5x5/2	32x32x64
D11	conv(E8)	3x3/1	32x32x64
D12	conv(D10+D11)	3x3/1	32x32x64
D13	deconv(D12)	5x5/2	64x64x32
D14	conv(D8)	3x3/1	64x64x32
D15	conv(D13+D14)	3x3/1	64x64x32
D16	conv(D15)	3x3/1	64x64x32
D17	conv(D16)	3x3/1	64x64x9
D18	Reshape(D17)	-	12288x3
D19	Concat(D3,D18)	-	16384x3

(b) Conv branch

Table 3: PSGN-ConvFC Decoder Architecture

Sl. No.	Layer	Filter Size/ Stride	Output Size
D1	Input	-	nx3
Global Feature Learning			
D2	MLP(D1)	1x1/1	nx32
D3	MLP(D2)	1x1/1	nx64
D4	MLP(D3)	1x1/1	nx64
D5	MaxPool(D4)	-	1x64
D6	Tile(D5, 4096)	-	4nx64
Local Feature Learning			
D7	MLP(D1-Neighborhood)	1x1/1	nx32x8
D8	MLP(D7)	1x1/1	nx32x8
D9	MLP(D8)	1x1/1	nx64x8
D10	MaxPool(D10)	-	nx64
D11	Tile(D10, 4)	-	4nx64
Grid Conditioning			
D12	Grid(2x2)	-	4x1
D13	Tile(D12, 1024)	-	4nx1
Feature Aggregation			
D14	Concat(D1, D6, D11, D13)	-	4nx132
D15	MLP(D14)	1x1/1	4nx64
D16	MLP(D15)	1x1/1	4nx128
D17	MLP(D16)	1x1/1	4nx128
D18	MLP(D17)	1x1/1	4nx3

(a) Dense Reconstruction Stage

S.No.	Layer	Output Size
D1	linear	256
D2	linear	256
D3	linear	1024x3
D4	Dense(D3)	4096x3
D5	Dense(D4)	16384x3

(b) Multi-Stage Reconstruction

Table 4: DensePCR Decoder Architecture

2. Reconstructions on ShapeNet

Qualitative comparison with the two variants of PSGN [2] for single-view reconstruction on ShapeNet [1] are provided in Figs. 1, 2 and 3. Note that the samples are randomly selected.

3. Reconstructions on Real-World Pix3D

Qualitative comparison with the two variants of PSGN [2] for single-view reconstruction on Pix3D [3] are provided in Fig. 4. Note that the samples are randomly selected.

References

[1] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, et al. Shapenet: An information-rich 3D model repository. *arXiv preprint arXiv:1512.03012*, 2015. 2

[2] H. Fan, H. Su, and L. Guibas. A point set generation network for 3D object reconstruction from a single image. In *CVPR*, volume 38, 2017. 1, 2

[3] X. Sun, J. Wu, X. Zhang, Z. Zhang, C. Zhang, T. Xue, J. B. Tenenbaum, and W. T. Freeman. Pix3D: Dataset and methods for single-image 3D shape modeling. In *CVPR*, 2018. 2

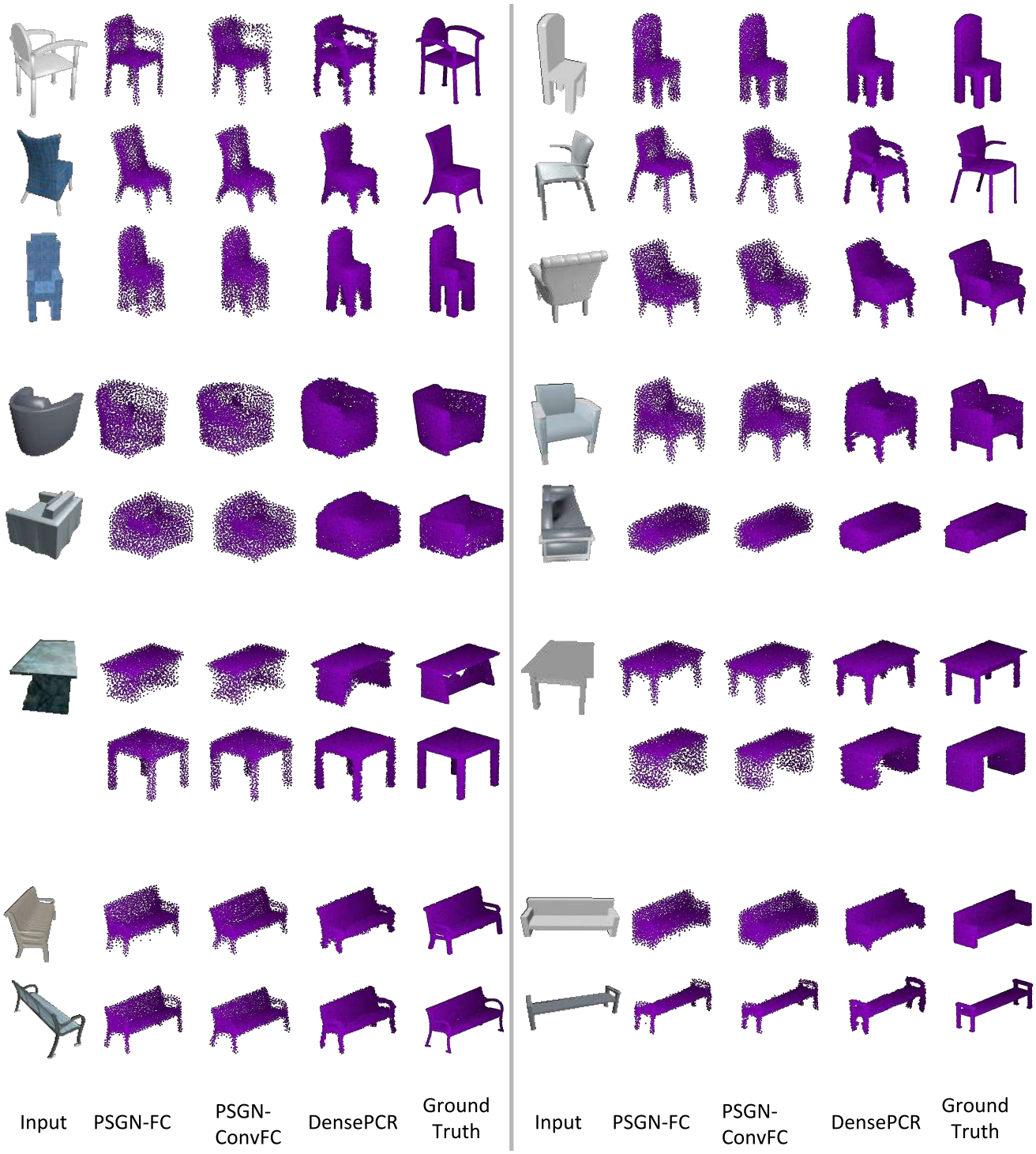

Figure 1: Reconstructions on ShapeNet (chair, sofa, table, bench). 3D reconstructions on randomly sampled input images from the validation set of ShapeNet. Note that both variants of PSGN have highly clustered regions resulting in high EMD scores (Table 1, main paper). On the other hand, DensePCR reconstructions are sharp and well distributed, and obtain lower EMD error metrics. Our reconstructions are also sharper and correspond better to the input image (handles and legs of chairs, benches and tables).

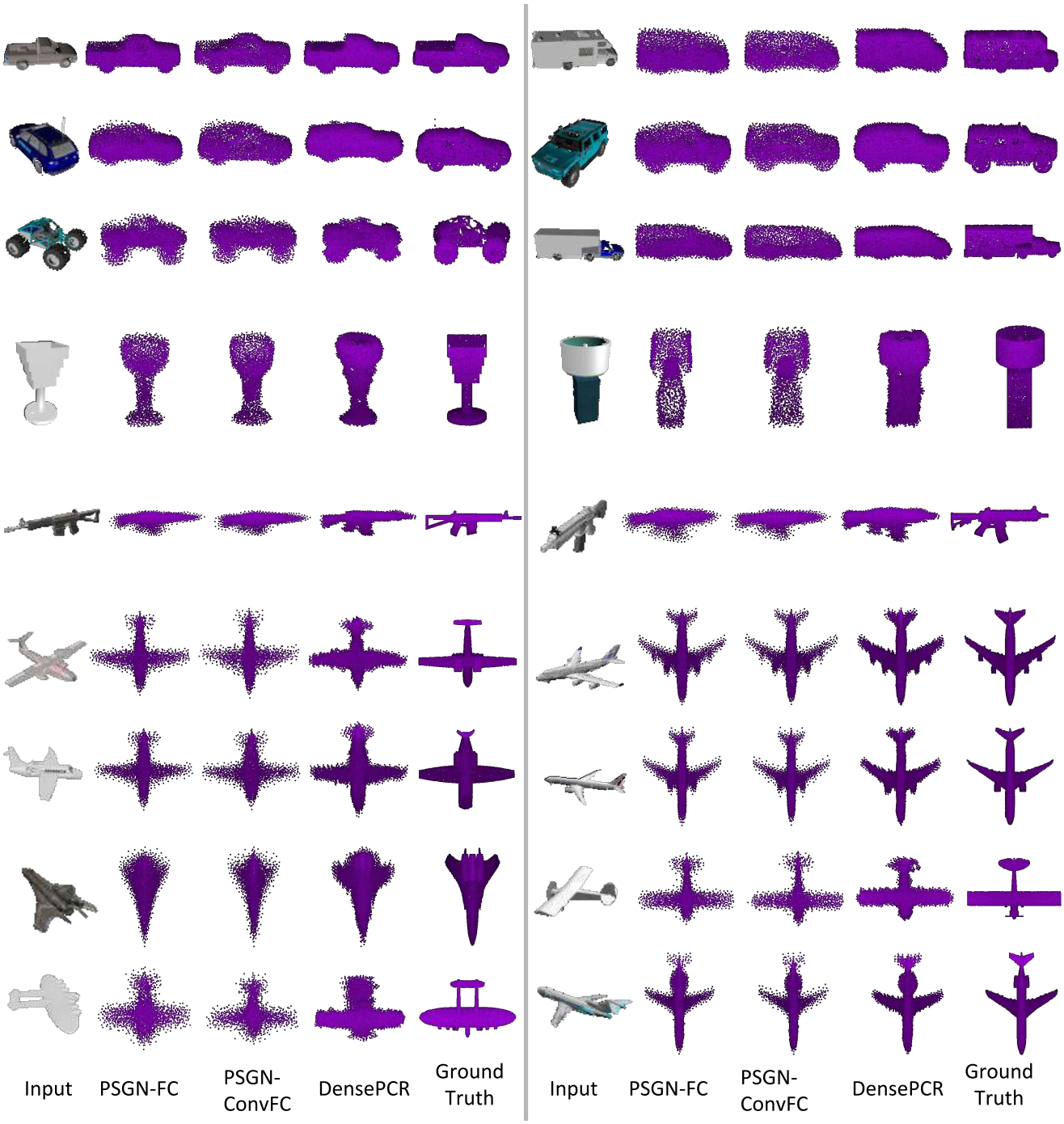

Figure 2: Reconstructions on ShapeNet (car, lamp, rifle, airplane). 3D reconstructions on randomly sampled input images from the validation set of ShapeNet. Note that both variants of PSGN have highly clustered regions resulting in high EMD scores (Table 1, main paper). On the other hand, DensePCR reconstructions are sharp and well distributed, and obtain lower EMD error metrics. Our reconstructions are also sharper and correspond better to the input image (wing and tail of airplanes, trigger of rifles).

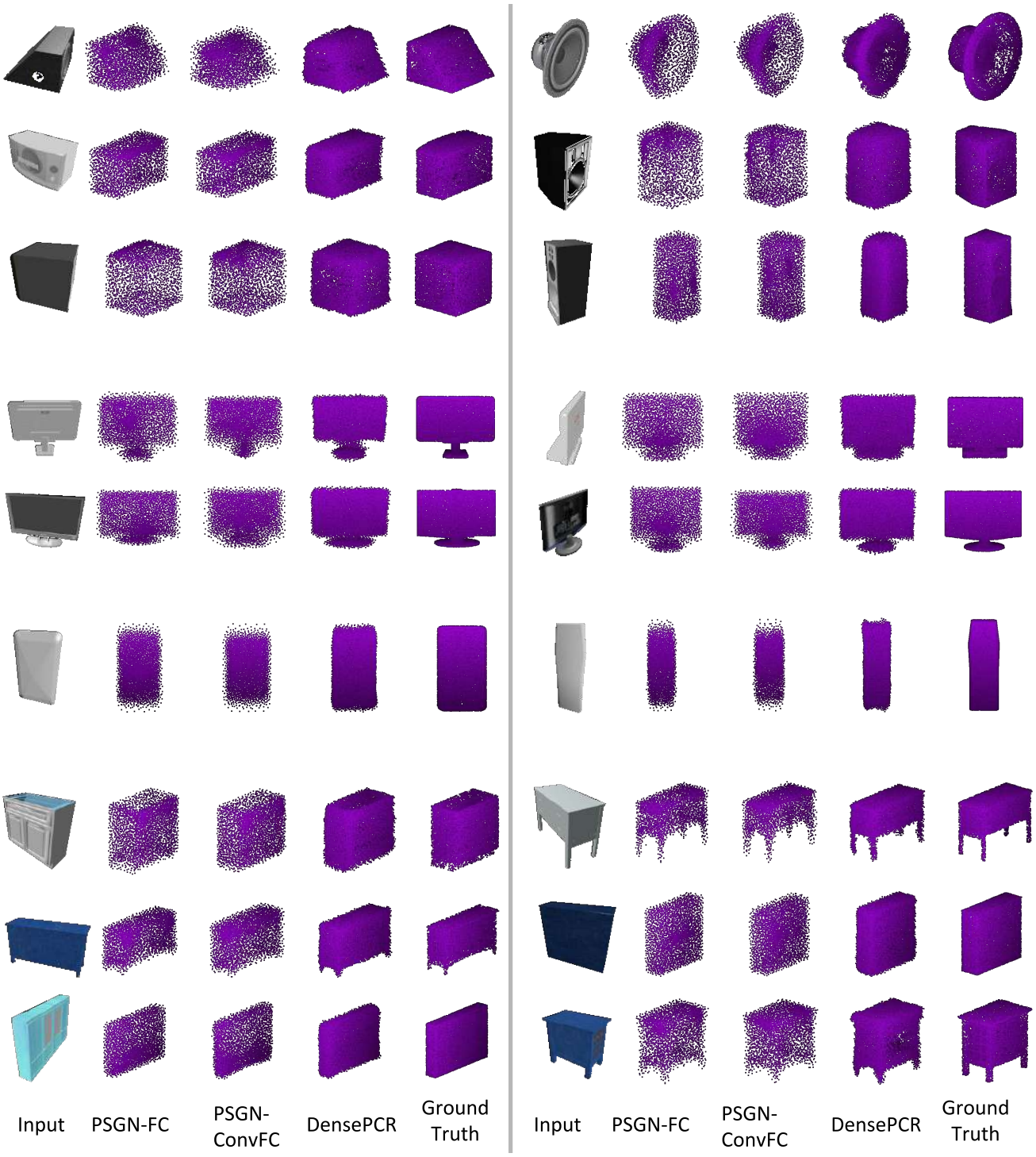

Figure 3: Reconstructions on ShapeNet (speaker, monitor, telephone, cabinet). 3D reconstructions on randomly sampled input images from the validation set of ShapeNet. Note that both variants of PSGN have highly clustered regions resulting in high EMD scores (Table 1, main paper). On the other hand, DensePCR reconstructions are sharp and well distributed, and obtain lower EMD error metrics.

Figure 4: Reconstructions on real-world Pix3D (chair, sofa, table). Surprisingly, both variants of PSGN have very poor generalizability, predicting highly incoherent shapes that often do not correspond to the input image (especially in chairs and sofas). On the other hand, DensePCR has very good generalization capability and predicts shapes that display high correspondence with the input image, despite the input space coming from a different distribution. Note that all three networks are trained on the same ShapeNet training set and tested on Pix3D.