What Do Single-view 3D Reconstruction Networks Learn? Supplementary Material

1. Metrics and evaluation protocol

For completeness, we provide the definitions of the evaluation metrics used and additional details for converting different shape representations for evaluation.

1.1. Intersection over union (IoU)

In the context of 3D shape reconstruction, the IoU between two shapes $\mathcal{G}$ and $\mathcal{R}$, represented as binary occupancy maps, is commonly defined as

$$\text{IoU}(\mathcal{G}, \mathcal{R}) = \frac{|\mathcal{G} \cap \mathcal{R}|}{|\mathcal{G} \cup \mathcal{R}|}. \tag{3}$$

In our evaluation protocol, we compare shapes A, B at a resolution of 128^3 binary cells (voxels).

1.2. Chamfer Distance (CD)

The Chamfer Distance (CD) between the ground truth shape $\mathcal{G}$ and the reconstructed shape $\mathcal{R}$ (both represented as point clouds) is defined as

$$\begin{aligned}\text{CD}(\mathcal{G}, \mathcal{R}) = &\frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \min_{g \in \mathcal{G}} \|r - g\|_2 \\ &+ \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \min_{r \in \mathcal{R}} \|g - r\|_2 .\end{aligned} \tag{4}$$

1.3. F-score

Here we provide a full definition of the F-score measure. Consider a ground truth shape $\mathcal{G}$ and a reconstructed shape $\mathcal{R}$ both represented as point clouds. For every point $r \in \mathcal{R}$ its distance to $\mathcal{G}$ is calculated as

$$e_r = \min_{g \in \mathcal{G}} \|r - g\|_2 .$$

Subsequently, we calculate the percentage of points reconstructed better than a certain threshold d which results in the *precision* value

$$P(d) = \frac{100}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} [e_r < d].$$

The same procedure is repeated in the opposite direction to produce the *recall* value

$$e_g = \min_{r \in \mathcal{R}} \|g - r\|_2 , \qquad R(d) = \frac{100}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} [e_g < d].$$

The final F-score is given by the harmonic mean of the precision and recall values

$$F(d) = \frac{2P(d)R(d)}{P(d) + R(d)}. \tag{5}$$

In practice, we set d as a fraction of the side length of the reconstructed volume (*e.g.*, 1%).

To evaluate a method using the F-score, we convert each shape prediction to a mesh representation, from which we evenly sample 10K points from the surface. We show how predictions by different methods compare in terms of their visual quality, precision and recall for a qualitative example in Fig. 11. OGN [3], Matryoshka [2] and the clustering baseline completely miss parts of the plane, resulting in high precision but comparably low recall. AtlasNet [1] reconstructs a complete shape, but misplaces individual parts, resulting in both low precision and low recall. The retrieval baseline finds a reasonably similar model, leading to comparably high precision and recall values.

2. Qualitative examples

In addition to the qualitative examples for a selection of classes in the main paper, we show a randomly sampled qualitative example for each class of the ShapeNet dataset in Fig. 12. As in the main paper, we show, from left to right: input image, ground truth shape, and predictions from AtlasNet [1], OGN [3], Matryoshka [2], our clustering baseline, our retrieval baseline, and an Oracle Nearest Neighbor. Numbers in the bottom left of each prediction indicate the IoU (dark gray) and the F-score at a 1% threshold (bold), respectively.

3. Quantitative results

In Tab. 1 we provide the exact F-score values at 1% threshold in the viewer-centered mode.

4. Statistical evaluation

In the main paper we showed within-class IoU histograms for a selection of three classes. We visualize such histograms for all 55 classes in Fig. 13.

We performed the Kolmogorov-Smirnov test on the within-class distributions for each ShapeNet class and each pairing of methods. The null hypothesis assumes that two distributions exhibit no statistically significant difference. We plot the p-values for each test result in Fig. 14. The color of each cell indicates whether the null hypothesis can be rejected (orange) or not (green). Aggregated results can be found in Fig. 6 (right) in the main paper.

References

[1] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. AtlasNet: A papier-mâché approach to learning 3D surface generation. In *CVPR*, 2018. 1

[2] Stephan R. Richter and Stefan Roth. Matryoshka networks: Predicting 3D geometry via nested shape layers. In *CVPR*, 2018. 1

[3] Maxim Tatarchenko, Alexey Dosovitskiy, and Thomas Brox. Octree generating networks: Efficient convolutional architectures for high-resolution 3D outputs. In *ICCV*, 2017. 1

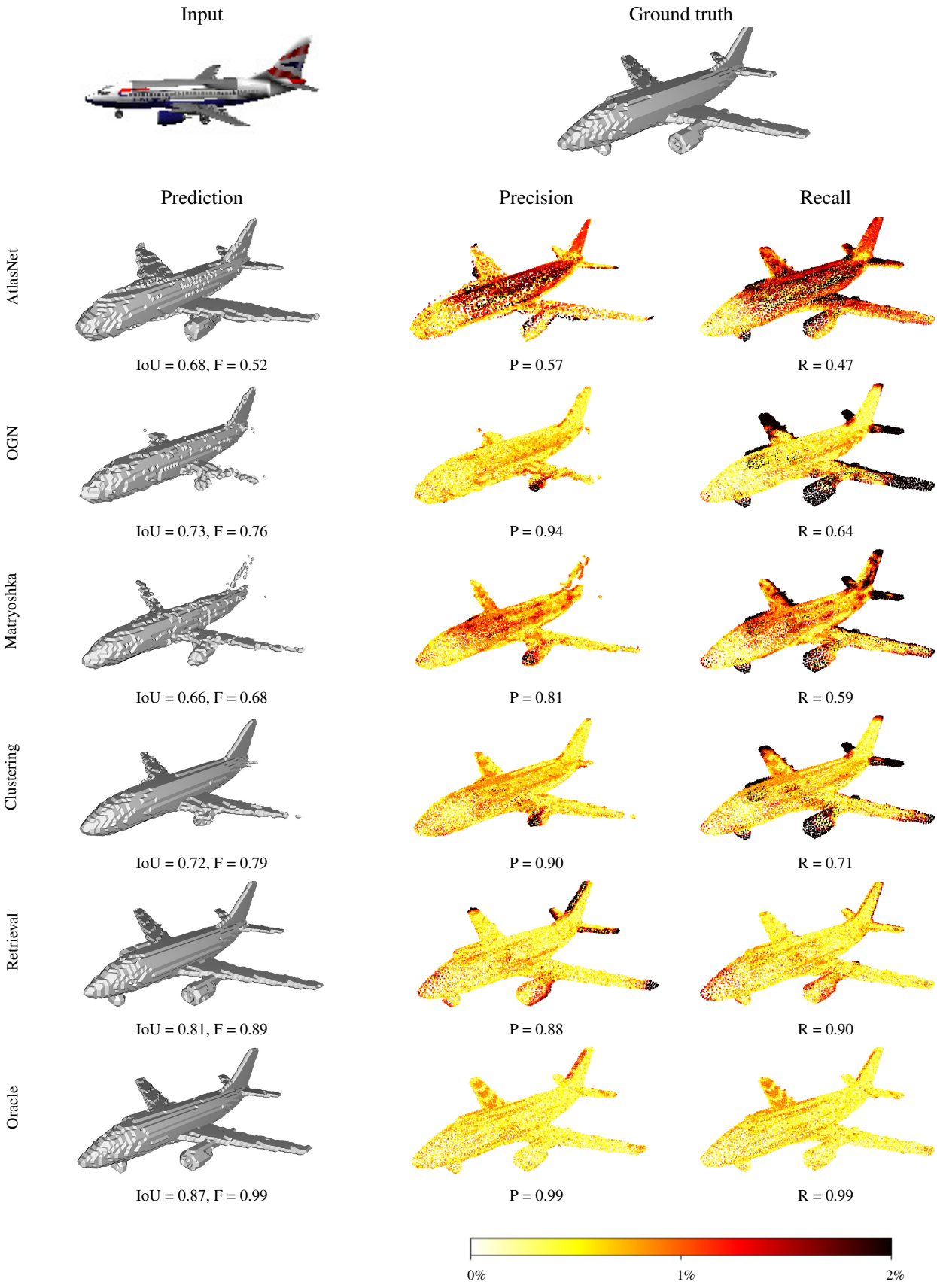

Figure 11: Exemplary predictions of different methods compared by visual quality, precision and recall. Colors encode the point-to-surface distance between shapes, normalized by the side length of the reconstructed volume.

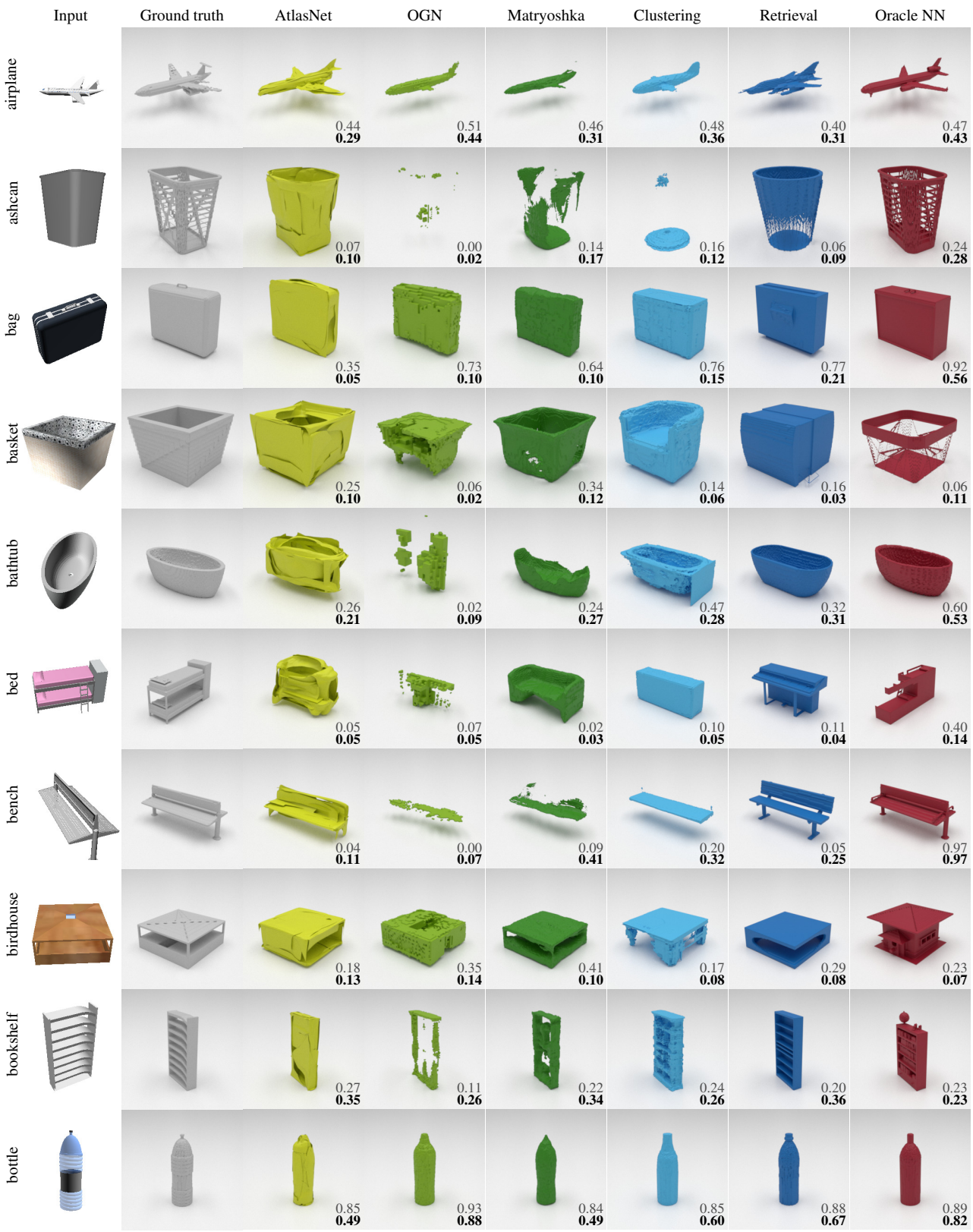

Figure 12: Qualitative results for all classes of ShapeNet. Numbers in the bottom right corner of each sample indicate IoU (dark gray) and F-score (bold), respectively.

Figure 12: Qualitative results for all classes of ShapeNet (continued).

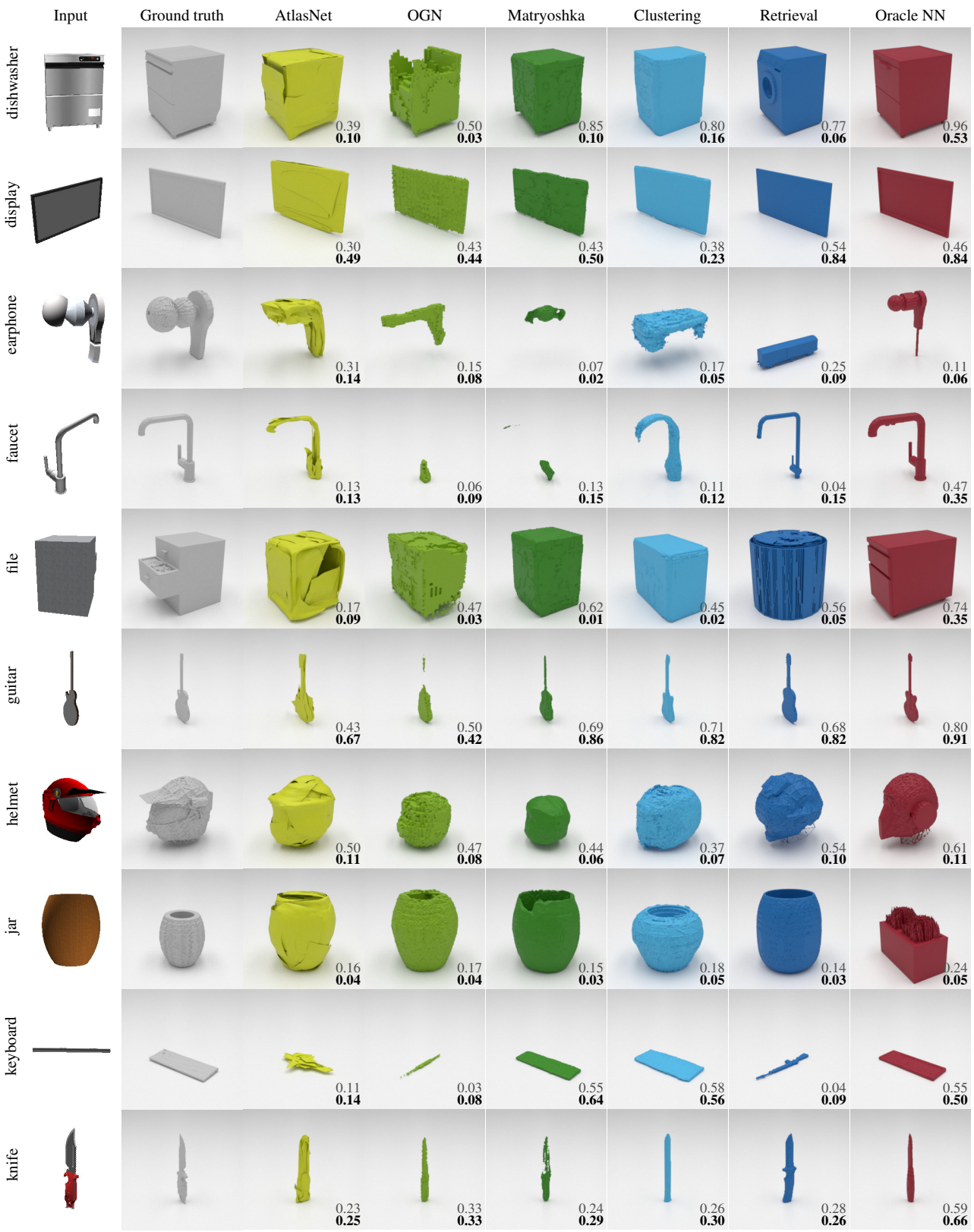

Figure 12: Qualitative results for all classes of ShapeNet (continued).

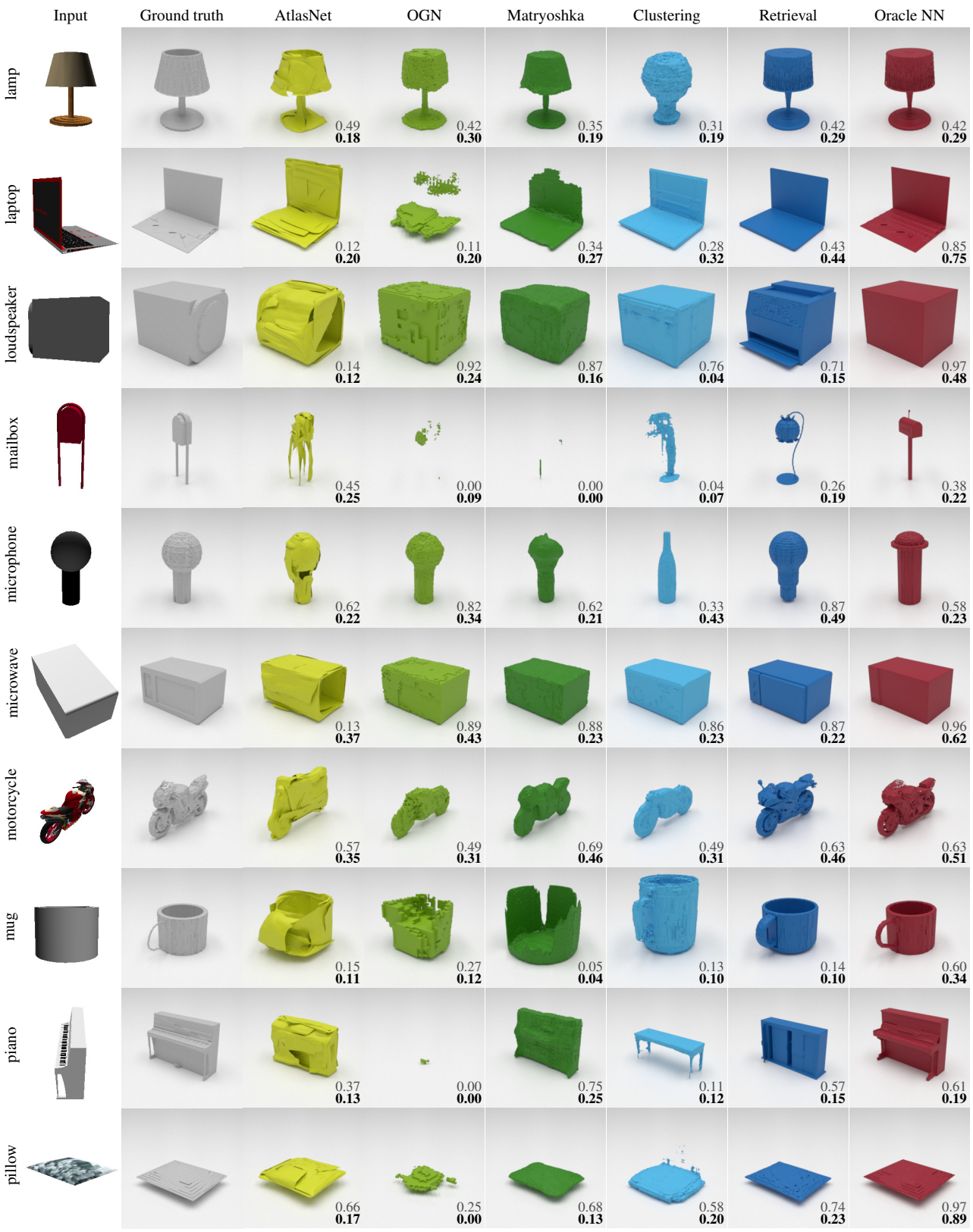

Figure 12: Qualitative results for all classes of ShapeNet (continued).

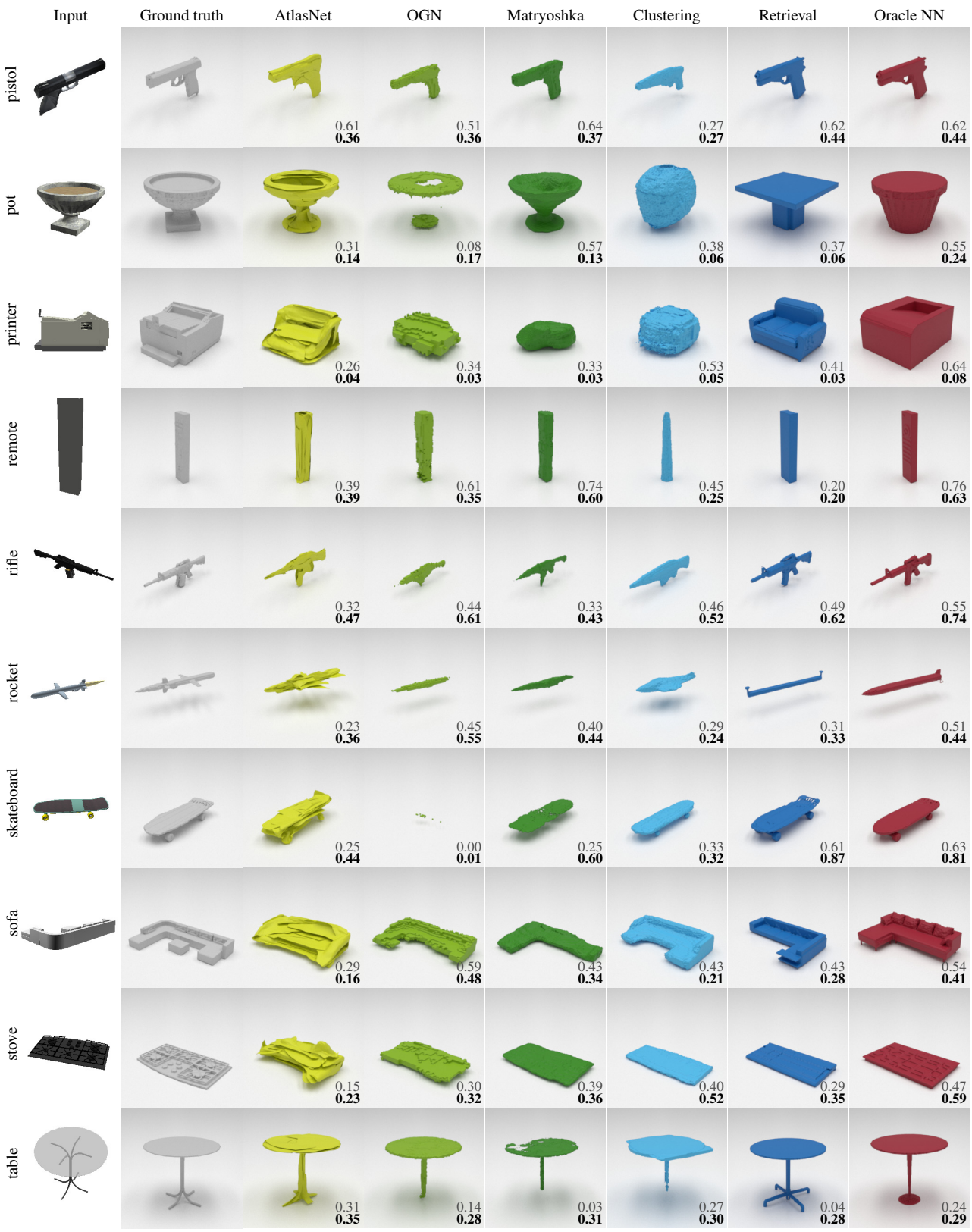

Figure 12: Qualitative results for all classes of ShapeNet (continued).

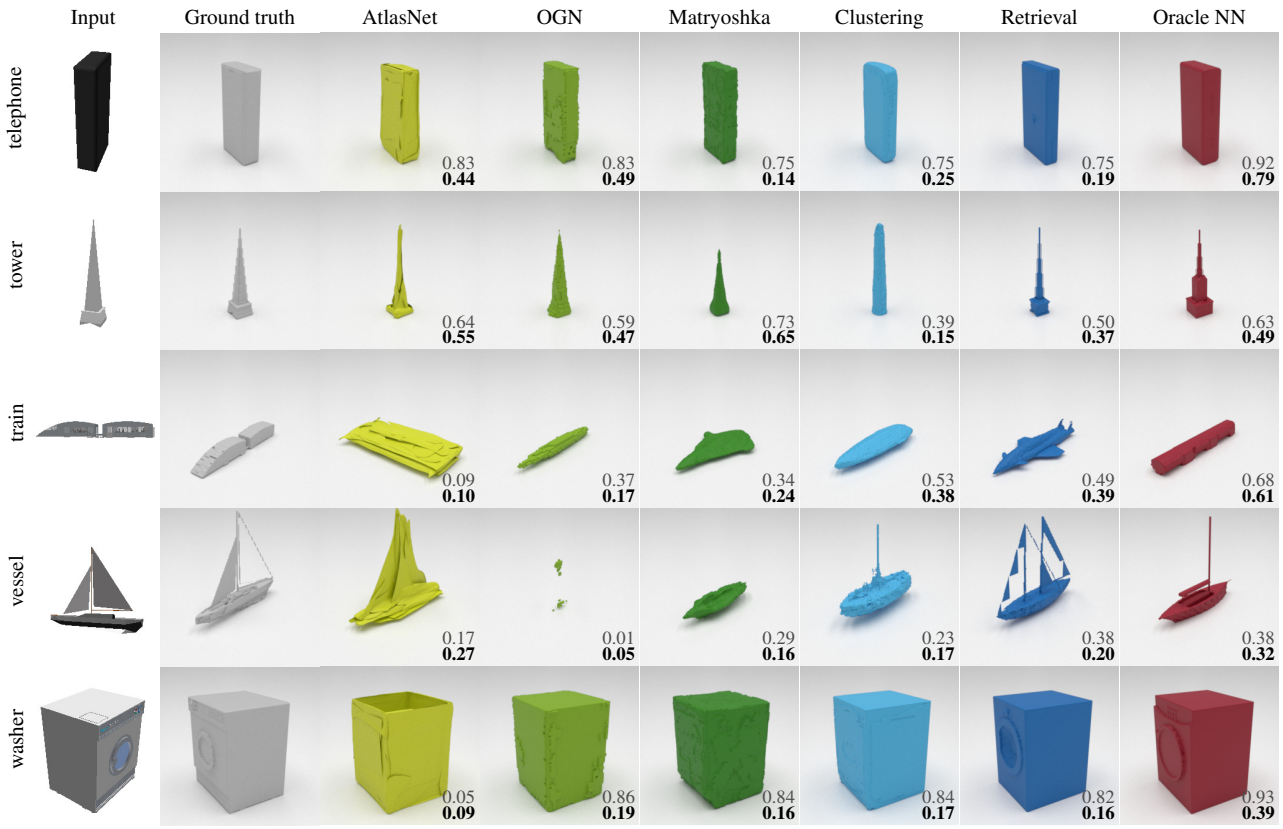

Figure 12: Qualitative results for all classes of ShapeNet (continued).

	AtlasNet	OGN	Matryoshka	Retrieval	Oracle NN
airplane	0.39	0.26	0.33	0.37	0.45
ashcan	0.18	0.23	0.26	0.21	0.24
bag	0.16	0.14	0.18	0.13	0.15
basket	0.19	0.16	0.21	0.15	0.15
bathtub	0.25	0.13	0.26	0.22	0.26
bed	0.19	0.12	0.18	0.15	0.17
bench	0.34	0.09	0.32	0.3	0.34
birdhouse	0.17	0.13	0.18	0.15	0.15
bookshelf	0.24	0.18	0.25	0.2	0.2
bottle	0.34	0.54	0.45	0.46	0.55
bowl	0.22	0.18	0.24	0.2	0.25
bus	0.35	0.38	0.41	0.36	0.44
cabinet	0.25	0.29	0.33	0.23	0.27
camera	0.13	0.08	0.12	0.11	0.12
can	0.23	0.46	0.44	0.36	0.44
cap	0.18	0.02	0.15	0.19	0.25
car	0.3	0.37	0.38	0.33	0.39
cellular	0.34	0.45	0.47	0.41	0.5
chair	0.25	0.15	0.27	0.2	0.23
clock	0.24	0.21	0.25	0.22	0.27
dishwasher	0.2	0.29	0.31	0.22	0.26
display	0.22	0.15	0.23	0.19	0.24
earphone	0.14	0.07	0.11	0.11	0.13
faucet	0.19	0.06	0.13	0.14	0.2
file	0.22	0.33	0.36	0.24	0.25
guitar	0.45	0.35	0.36	0.41	0.58
helmet	0.1	0.06	0.09	0.08	0.12
jar	0.21	0.22	0.25	0.19	0.22
keyboard	0.36	0.25	0.37	0.35	0.49
knife	0.46	0.26	0.21	0.37	0.54
lamp	0.26	0.13	0.2	0.21	0.27
laptop	0.29	0.21	0.33	0.26	0.33
loudspeaker	0.2	0.26	0.27	0.19	0.23
mailbox	0.21	0.2	0.23	0.2	0.19
microphone	0.23	0.22	0.19	0.18	0.21
microwave	0.23	0.36	0.35	0.22	0.25
motorcycle	0.27	0.12	0.22	0.24	0.28
mug	0.13	0.11	0.15	0.11	0.17
piano	0.17	0.11	0.16	0.14	0.17
pillow	0.19	0.14	0.17	0.18	0.3
pistol	0.29	0.22	0.23	0.25	0.3
pot	0.19	0.15	0.19	0.14	0.16
printer	0.13	0.11	0.13	0.11	0.14
remote	0.3	0.33	0.31	0.31	0.37
rifle	0.43	0.28	0.3	0.36	0.48
rocket	0.34	0.2	0.23	0.26	0.32
skateboard	0.39	0.11	0.39	0.35	0.47
sofa	0.24	0.23	0.27	0.21	0.27
stove	0.2	0.19	0.24	0.18	0.19
table	0.31	0.24	0.34	0.26	0.34
telephone	0.33	0.42	0.45	0.4	0.5
tower	0.24	0.2	0.25	0.25	0.25
train	0.34	0.29	0.3	0.32	0.38
vessel	0.28	0.19	0.22	0.23	0.29
washer	0.2	0.31	0.31	0.21	0.25

Table 1: F-score evaluation (@1%) in the viewer-centered mode.

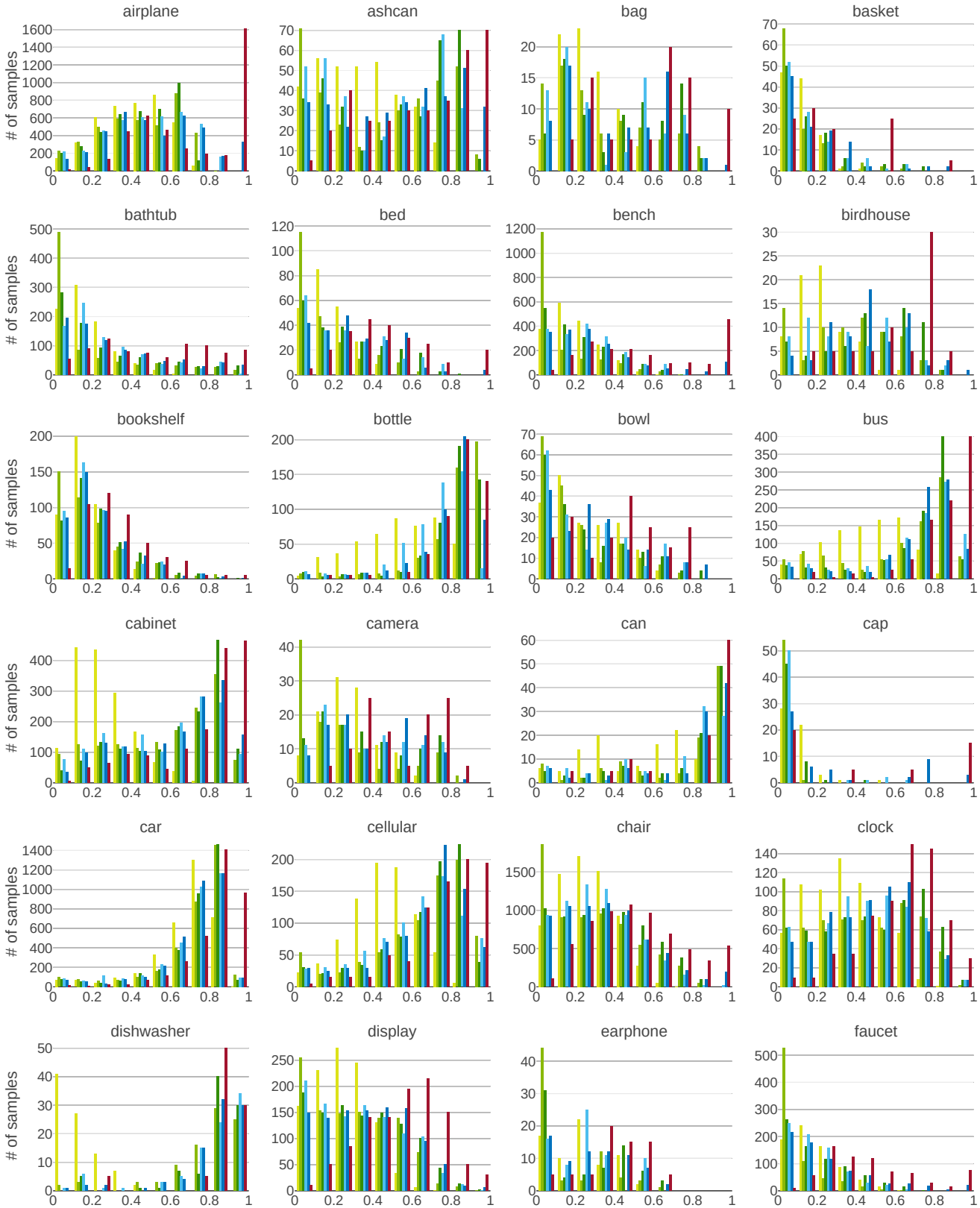

Figure 13: Distribution of within-class reconstruction performance for all ShapeNet classes, measured by IoU.

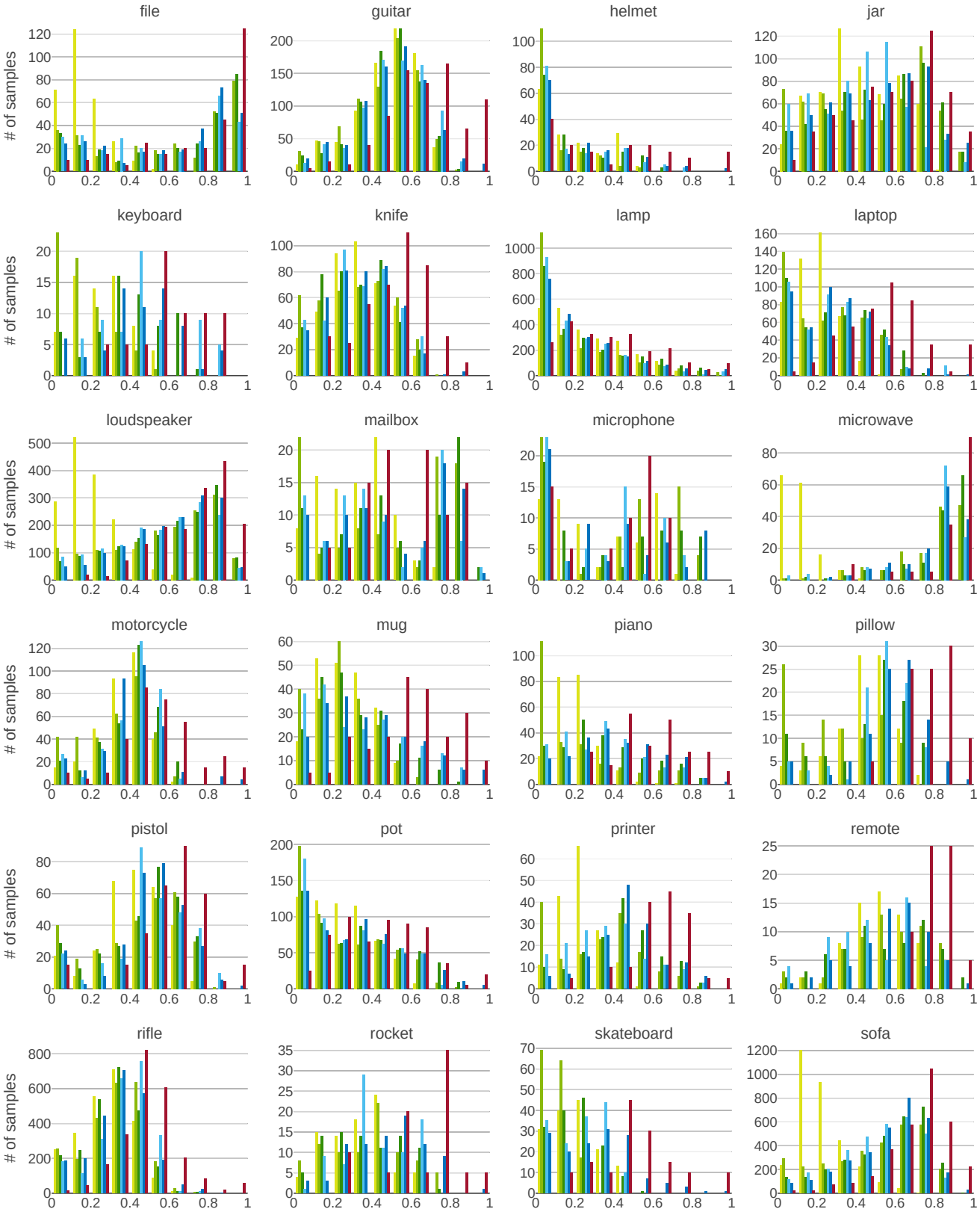

Figure 13: Distribution of within-class reconstruction performance for all ShapeNet classes, measured by IoU (continued).

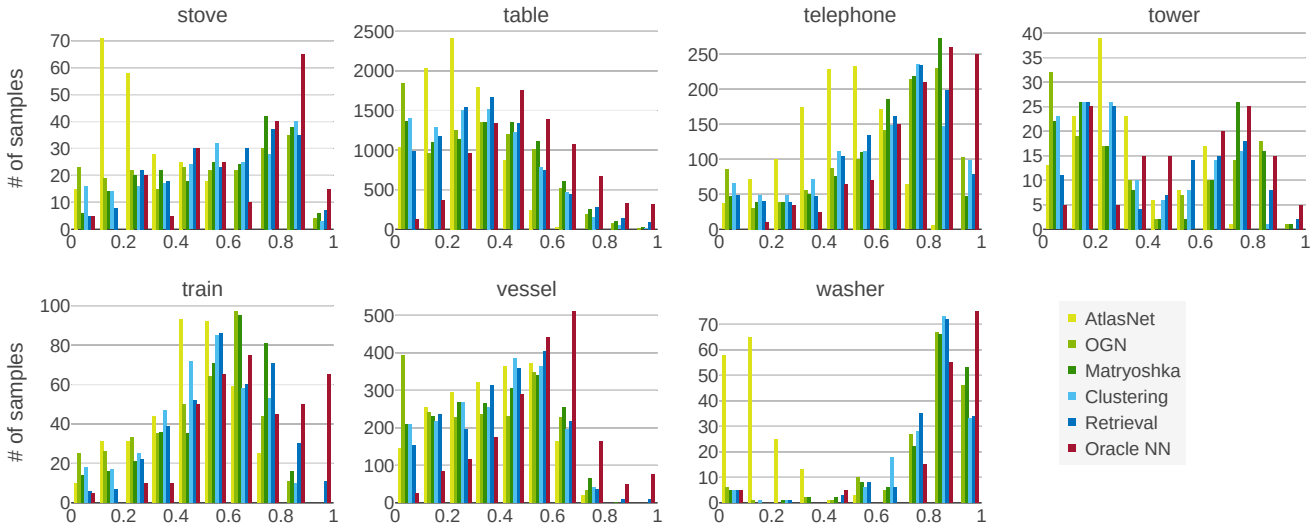

Figure 13: Distribution of within-class reconstruction performance for all ShapeNet classes, measured by IoU (continued).

	airplane	ashcan	bag	basket	bathtub	bed	bench	birdhouse	bookshelf	bottle	bowl	bus	cabinet	camera	can	cap	car	cellular	chair	clock	dishwasher	display	earphone	faucet	file	guitar	helmet	jar	keyboard	knife	lamp	laptop	loudspeaker	mailbox	microphone	microwave	motorcycle	mug	piano	pillow	pistol	pot	printer	remote	rifle	rocket	skateboard	sofa	stove	table	telephone	tower	train	vessel	washer
ATL-OGN	0.1	0.15	0.15	0.68	0.0	0.68	0.15	0.1	0.02	0.0	0.36	0.03	0.0	0.95	0.0	0.51	0.95	0.51	0.01	0.1	0.51	0.06	0.68	0.24	0.0	0.36	0.51	0.36	1.0	0.84	0.03	0.24	0.0	1.0	1.0	0.1	1.0	0.84	0.24	0.84	0.15	0.36	0.02	1.0	1.0	0.68	0.1	0.0	0.02	0.02	0.36	0.51	0.68	0.15	0.84
ATL-MN	0.68	0.15	0.36	0.15	0.0	0.02	0.51	0.1	0.02	0.0	0.68	0.03	0.03	0.51	0.03	1.0	1.0	0.68	0.01	0.1	1.0	0.06	1.0	0.68	0.0	0.95	1.0	0.36	0.84	0.95	0.36	0.06	0.0	0.84	1.0	0.1	0.68	0.15	0.01	0.84	0.51	0.1	0.01	0.51	1.0	1.0	0.95	0.0	0.06	0.01	0.68	1.0	0.84	0.84	1.0
ATL-CL	0.06	0.68	1.0	1.0	0.0	0.24	0.51	0.95	0.51	0.01	1.0	0.15	0.0	1.0	0.0	0.95	0.68	0.36	0.1	0.36	1.0	0.1	0.51	1.0	0.0	0.68	0.95	0.68	0.95	1.0	0.68	0.1	0.0	1.0	0.03	0.02	1.0	0.51	0.06	0.51	1.0	0.68	0.03	0.68	0.95	0.36	0.95	0.01	0.01	0.02	0.36	0.95	1.0	0.68	1.0
ATL-RET	0.02	0.15	0.1	0.51	0.0	0.06	0.0	0.51	0.03	0.03	0.95	0.06	0.0	0.84	0.03	1.0	0.68	0.36	0.0	0.24	0.95	0.03	0.95	0.06	0.0	0.24	0.95	0.15	0.1	0.84	0.36	0.02	0.0	0.51	1.0	0.1	0.84	0.06	0.0	1.0	0.68	0.1	0.0	1.0	0.68	0.68	0.15	0.0	0.03	0.0	0.36	0.68	0.51	0.95	1.0
ATL-ORA	0.02	0.06	0.15	0.95	0.0	0.36	0.0	0.15	0.15	0.0	0.51	0.0	0.15	0.03	0.0	0.84	0.68	0.51	0.0	0.36	0.24	0.03	0.36	0.03	0.68	0.95	0.36	0.24	0.24	0.84	0.1	0.24	0.15	0.0	0.01	0.84	0.24	0.84	0.03	0.0	0.51	0.15	0.84	0.0	0.03	0.0	0.68	0.06	0.84	0.0	0.36	0.03	0.51	0.36	0.24
OGN-MN	0.51	0.95	1.0	0.95	0.06	0.36	0.03	1.0	1.0	1.0	0.68	0.03	0.36	0.24	1.0	0.95	0.95	0.84	0.36	0.84	0.36	0.95	0.51	0.1	0.95	0.84	0.36	0.51	0.84	0.51	0.36	0.95	0.84	1.0	1.0	0.51	0.95	0.51	0.1	1.0	1.0	0.84	1.0	0.95	1.0	1.0	0.15	0.15	0.84	0.84	0.84	1.0	0.24	0.24	1.0
OGN-CL	0.84	0.68	0.03	1.0	0.06	0.68	0.02	0.68	0.68	0.36	0.51	0.84	0.95	0.95	0.1	1.0	0.95	0.84	0.03	0.68	0.84	0.95	0.95	0.36	0.95	1.0	0.68	0.68	0.84	0.51	0.51	0.84	0.84	0.95	0.24	0.84	0.95	0.95	0.36	0.24	0.1	0.36	1.0	0.24	0.84	0.01	0.68	0.1	1.0	0.24	0.84	0.51	0.36	0.68	0.51
OGN-RET	0.24	0.51	1.0	0.84	0.0	0.24	0.0	1.0	0.95	0.84	0.95	0.15	1.0	0.51	1.0	0.1	0.95	1.0	0.02	1.0	1.0	0.68	0.1	0.02	0.84	0.68	0.36	1.0	0.15	0.84	0.36	0.84	0.36	1.0	1.0	1.0	0.84	0.24	0.36	0.68	0.15	0.84	1.0	0.84	0.95	1.0	0.0	0.03	0.95	0.15	1.0	0.68	0.68	0.51	1.0
OGN-ORA	0.02	0.1	0.0	0.36	0.0	0.06	0.0	0.0	0.36	0.01	0.06	0.0	0.0	0.0	0.0	0.68	0.1	0.1	0.51	0.02	0.0	0.68	0.24	0.0	0.0	0.36	0.0	0.24	0.15	0.68	0.02	0.84	0.01	0.0	0.1	0.0	0.68	0.68	0.06	0.0	0.1	0.51	0.0	0.0	0.15	0.0	0.36	0.0	0.0	0.24	0.03	0.0	0.1	0.03	0.01
MN-CL	0.68	0.36	0.1	0.68	1.0	0.95	0.95	0.68	0.68	0.51	0.68	0.24	0.51	0.84	0.24	1.0	0.84	0.36	0.03	0.36	1.0	1.0	0.15	0.68	1.0	0.95	1.0	0.51	0.15	0.68	0.36	0.95	0.84	0.68	0.06	0.51	0.36	1.0	0.95	0.24	0.15	0.36	1.0	0.02	0.95	0.1	0.95	0.84	1.0	0.36	0.15	0.51	1.0	0.84	0.95
MN-RET	0.06	0.68	0.68	1.0	0.68	1.0	0.03	1.0	1.0	0.68	1.0	0.95	0.36	1.0	1.0	0.68	0.95	1.0	0.03	0.84	0.84	0.84	0.68	0.36	0.95	0.84	1.0	0.84	0.84	0.95	0.51	0.95	0.51	1.0	1.0	0.51	1.0	0.84	1.0	0.68	0.51	1.0	1.0	0.68	0.84	0.95	0.24	0.24	0.68	0.24	0.95	0.95	0.15	0.84	1.0
MN-ORA	0.03	0.24	0.0	0.06	0.03	0.15	0.03	0.0	0.36	0.03	0.03	0.01	0.15	0.0	0.0	1.0	0.24	0.15	0.15	0.0	0.36	0.68	0.1	0.15	0.0	0.84	0.06	0.06	0.01	0.51	0.03	0.95	0.03	0.0	0.02	0.0	0.1	0.02	0.06	0.0	0.15	0.68	0.0	0.0	0.06	0.0	0.68	0.03	0.0	0.68	0.01	0.0	0.03	0.06	0.06
CL-RET	0.68	0.15	0.02	1.0	0.68	0.84	0.02	0.95	0.84	0.95	0.84	0.68	0.95	0.84	0.24	0.36	0.95	0.68	0.15	0.84	1.0	0.95	0.06	0.03	0.95	0.68	0.68	0.68	0.0	0.95	0.36	0.95	0.68	0.36	0.1	1.0	0.68	0.68	0.95	0.51	0.24	0.36	1.0	0.51	0.51	0.01	0.02	0.68	0.84	0.51	0.36	0.51	0.68	0.84	0.84
CL-ORA	0.01	0.68	0.51	0.84	0.02	0.51	0.03	0.01	0.84	0.06	0.68	0.0	0.0	0.01	0.15	1.0	0.1	0.1	0.24	0.1	0.1	0.51	0.68	0.02	0.0	0.84	0.06	0.68	0.95	0.68	0.06	0.68	0.02	0.01	0.36	0.0	0.1	0.1	0.36	0.24	0.84	0.51	0.0	0.15	0.24	0.51	0.95	0.1	0.0	0.1	0.0	0.06	0.15	0.1	0.51
RET-ORA	0.02	0.02	0.0	0.24	0.36	0.24	0.24	0.0	0.51	0.02	0.06	0.0	0.0	0.0	0.0	0.24	0.1	0.06	0.36	0.06	0.02	0.51	0.03	0.68	0.0	0.36	0.1	0.15	0.0	0.84	0.03	0.51	0.24	0.0	0.03	0.0	0.06	0.0	0.03	0.0	0.15	0.68	0.0	0.0	0.68	0.0	0.03	0.15	0.0	0.95	0.0	0.0	0.01	0.36	0.03

Figure 14: P-values of the pairwise Kolmogorov-Smirnov test on per-class IoU performance histograms. The null-hypotheses of two distributions being the same can be rejected for $p < 0.05$ (orange) and cannot be rejected for $p > 0.05$ (green).