

A Implementation Details

Experiments on PartNet-E follow the rendering protocol and training-testing splits used in PartSLIP, where we render point clouds into 800×800 resolution images from 10 selected views. Then we utilize a 64×64 grid of point prompts in SAM to generate segment masks. For ShapeNetPart, we follow the approach of PointCLIPv2, projecting point clouds into 224×224 resolution images and using a 16×16 grid of points as prompts for SAM to produce masks. Since there is no official few-shot split provided for ShapeNetPart, we randomly sample 8 instances per category from the training set as the few-shot training data. We adopt the method from 3-by-2 to further subdivide the overlapping regions among SAM-generated masks, ensuring that all masks are mutually non-overlapping.

For the training on PartNet-E, we used the Adam optimizer with a learning rate of 0.001 and a weight decay of 0.01. A StepLR scheduler was applied with a step size of 5 and a gamma of 0.9. Each category in PartNet-E was trained for 20 epochs. For ShapeNetPart, we also used the Adam optimizer, with a learning rate of 0.01 and no weight decay. A StepLR scheduler with a step size of 5 and gamma of 0.9 was used as well. Each category in ShapeNetPart was trained for 30 epochs. All experiments are conducted on a single NVIDIA V100 GPU. We will release the code for a reference of implementation details.

B Per-category Results for Comparisons and Ablations

The complete comparison results of the proposed method SegGraph and the baselines across all 45 categories on the PartNet-E dataset are presented in Tab. A1.

Table A1: Few-shot part segmentation results on PartNet-E

Categories	few-shot w/ additional data (45x8+28k)			few-shot (45x8)						
	PointNet++	SoftGroup	PointNext	PartSLIP	PartSLIP++	PartSTAD	PartDistill	3-By-2	Ours	(std)
Bottle	48.8	41.4	68.4	83.4	85.5	83.6	84.6	80.9	90.0	0.7
Box	18.6	8.8	84.2	84.5	85.5	81.1	87.9	76.1	87.9	0.6
Bucket	0.0	25.0	4.1	36.5	85.5	83.6	50.7	78.4	85.6	2.2
Camera	6.5	23.6	33.2	58.3	63.2	64.4	60.1	62.6	70.8	1.0
Cart	6.4	23.9	36.3	88.1	84.9	85.0	90.1	81.2	87.1	0.3
Chair	84.7	88.3	91.8	85.3	85.3	85.3	88.4	84.4	89.5	0.4
Clock	19.2	2.5	28.4	37.6	54.1	47.4	37.2	45.8	49.8	3.1
Coffee machine	34.6	8.3	17.9	37.8	37.6	35.8	40.2	34.2	41.7	1.2
Dishwasher	49.5	53.0	69.2	62.5	60.9	60.6	60.2	53.6	63.1	0.9
Dispenser	12.1	18.9	26.0	73.8	72.0	73.7	74.7	78.2	77.6	0.4
Display	78.3	62.1	89.4	84.8	85.1	82.3	87.4	72.6	81.9	0.1
Door	45.7	53.1	43.8	40.8	45.1	61.4	55.5	54.4	73.6	0.2
Eyeglasses	76.2	72.4	88.1	88.3	88.3	92.5	91.1	92.8	97.0	0.4
Faucet	67.2	68.4	85.0	71.4	65.9	65.3	73.5	66.9	68.4	0.6
Folding chair	10.9	14.7	96.4	86.3	89.9	91.6	90.7	93.6	94.7	0.1
Globe	46.5	59.0	92.3	95.7	92.8	93.5	97.4	95.2	98.4	0.3
Kettle	20.9	57.4	45.1	77.0	85.6	84.2	78.6	81.5	87.8	0.5
Keyboard	74.5	58.9	45.0	53.6	72.4	82.4	70.8	89.6	88.7	0.0
Kitchenpot	15.8	45.5	57.0	69.6	72.9	73.5	69.7	65.0	85.2	0.7
Knife	35.4	31.3	58.7	65.2	64.3	63.8	71.4	75.1	77.3	0.1
Lamp	68.0	82.2	64.9	66.1	68.0	68.4	69.2	59.5	78.5	0.7
Laptop	55.4	18.4	32.5	29.7	29.7	34.6	40.0	45.3	58.8	0.5
Lighter	35.0	30.2	16.3	64.7	66.2	65.9	64.9	65.0	64.5	0.4
Microwave	43.6	38.3	40.5	42.7	49.5	59.1	43.8	40.2	59.4	2.4
Mouse	16.7	55.9	32.5	44.0	58.7	48.9	46.9	68.4	46.9	1.3
Oven	34.4	13.7	37.8	73.5	70.3	71.9	72.8	60.0	75.7	1.3
Pen	46.5	28.9	39.5	71.5	66.4	60.6	74.4	62.8	70.4	1.0
Phone	20.0	2.4	39.5	48.4	59.0	63.2	50.8	41.0	59.2	2.7
Pliers	57.7	74.2	99.6	33.2	29.7	99.3	90.4	99.4	99.6	0.0
Printer	0.0	1.2	0.0	4.3	6.2	7.9	6.3	8.5	27.1	2.1
Refrigerator	43.4	46.9	76.2	55.8	55.7	53.6	58.1	51.7	59.1	2.3
Remote	3.6	37.1	57.8	38.3	36.4	53.4	40.7	54.1	83.5	0.8
Safe	31.9	5.3	25.6	32.2	37.2	36.8	58.6	34.7	66.1	1.4
Scissors	50.0	76.0	57.3	60.3	60.5	68.5	68.8	65.7	75.2	0.6
Stapler	51.6	80.1	88.6	84.8	63.0	85.8	86.3	90.1	95.1	0.3
Storage furniture	46.9	60.2	68.5	53.6	57.3	59.5	56.5	51.7	60.1	0.4
Suitcase	40.7	18.3	13.6	70.4	70.0	68.3	73.4	65.2	77.3	1.3
Switch	1.8	21.0	39.7	59.4	56.1	57.9	60.7	54.6	76.1	1.2
Table	57.7	59.1	61.6	42.5	52.1	47.8	63.3	56.6	63.9	0.3
Toaster	14.7	26.4	8.4	60.0	50.8	58.6	58.7	56.5	66.1	6.2
Toilet	18.0	18.0	16.5	53.8	70.5	57.5	55.0	56.9	59.4	3.7

Table A1 – continued from previous page

Categories	few-shot w/ additional data (45x8+28k)			few-shot (45x8)						
	PointNet++	SoftGroup	PointNext	PartSLIP	PartSLIP++	PartSTAD	PartDistill	3-By-2	Ours	(std)
Trash can	71.7	17.0	22.8	22.3	24.2	21.2	70.0	33.1	55.3	1.1
USB	52.4	44.1	67.9	54.4	57.5	59.9	64.3	79.0	63.8	1.8
Washing machine	0.6	24.1	27.3	53.5	48.7	48.2	55.1	52.6	55.6	0.9
Window	26.3	39.2	83.3	75.4	72.8	76.1	78.1	73.9	84.3	0.8
Overall	36.5	38.5	50.6	59.4	62.1	65.0	65.9	64.2	72.8	1.1

Table A2: Ablation study of each module on PartNet-E

	Base		AVE		SE & eq(2)		graph		SE & eq(2) & graph	
	mIoU	std	mIoU	std	mIoU	std	mIoU	std	mIoU	std
Bottle	83.4	0.3	89.0	0.8	87.9	1.9	87.8	1.2	90.0	2.7
Box	84.2	0.3	84.9	3.8	90.4	3.9	87.7	1.5	87.9	1.2
Bucket	75.3	0.5	77.0	5.4	79.1	5.5	81.7	1.0	85.6	0.5
Camera	58.8	0.1	70.1	0.5	69.4	1.5	68.1	0.6	70.8	0.7
Cart	78.1	0.2	86.7	2.9	88.7	0.4	87.1	0.0	87.1	0.3
Chair	86.1	0.1	88.1	1.3	89.0	0.3	90.1	0.1	89.5	2.3
Clock	22.7	1.0	46.6	5.0	32.8	5.1	43.9	1.9	49.8	0.6
CoffeeMachine	36.1	0.2	32.2	1.0	35.0	3.8	41.8	1.0	41.7	2.2
Dishwasher	58.8	0.2	61.5	2.1	61.5	1.1	62.8	0.5	63.1	1.0
Dispenser	69.8	0.5	77.2	0.4	79.6	1.5	78.2	2.0	77.6	0.4
Display	75.9	0.3	81.4	1.8	82.9	0.3	81.9	0.5	81.9	3.1
Door	62.3	0.2	72.7	0.2	72.3	1.5	70.5	1.0	73.6	0.9
Eyeglasses	90.8	0.4	94.8	0.5	94.2	0.4	95.9	0.7	97.0	0.4
Faucet	64.6	0.7	69.5	1.4	70.2	1.0	68.9	0.8	68.4	0.1
FoldingChair	91.6	0.3	94.6	0.5	94.9	0.2	94.6	0.2	94.7	0.2
Globe	96.5	0.4	98.4	0.5	98.2	0.2	98.4	0.0	98.4	0.4
Kettle	80.2	0.2	87.2	0.3	87.3	0.6	85.9	0.2	87.8	0.6
Keyboard	84.0	0.1	85.8	0.2	85.9	0.4	87.9	0.1	88.7	0.1
KitchenPot	80.5	0.2	82.0	0.5	82.9	2.1	84.4	0.3	85.2	0.3
Knife	71.5	0.6	78.9	1.1	77.0	2.3	76.6	1.0	77.3	0.5
Lamp	71.9	0.1	79.9	2.0	79.9	0.3	77.6	0.6	78.5	0.0
Laptop	54.8	0.5	54.3	0.7	55.7	0.8	57.6	1.3	58.8	0.7
Lighter	57.9	0.5	64.7	0.5	63.8	2.2	65.7	1.8	64.5	0.1
Microwave	50.0	0.6	58.6	2.3	60.3	3.4	56.1	2.7	59.4	0.7
Mouse	31.4	0.2	42.9	0.4	42.9	0.7	43.0	0.9	46.9	0.4
Oven	67.3	0.2	74.8	0.7	76.1	0.5	74.9	0.7	75.7	2.4
Pen	64.1	0.3	71.9	1.9	74.6	1.9	68.3	1.0	70.4	1.3
Phone	44.5	0.7	54.9	2.1	57.9	3.9	52.5	3.1	59.2	1.3
Pliers	99.6	0.0	99.5	0.1	99.5	0.1	99.6	0.0	99.6	1.0
Printer	20.5	0.4	20.4	5.6	19.7	6.4	23.8	0.8	27.1	0.0
Refrigerator	53.6	0.3	62.6	0.5	62.5	0.7	59.6	0.4	59.1	2.1
Remote	63.1	1.2	78.9	0.1	79.3	0.6	81.9	1.4	83.5	0.8
Safe	59.4	0.5	59.5	2.3	62.0	1.3	63.7	0.9	66.1	1.4
Scissors	69.4	0.4	73.6	1.4	74.1	0.6	73.2	1.2	75.2	0.6
Stapler	90.2	0.3	94.3	0.3	93.8	1.1	94.1	0.9	95.1	0.3
Storage furniture	57.0	0.3	58.2	0.4	58.6	0.6	59.1	0.5	60.1	0.4
Suitcase	70.1	0.4	70.5	2.8	66.9	5.4	73.0	0.9	77.3	1.3
Switch	63.2	0.6	72.6	1.4	70.6	1.2	75.1	1.5	76.1	1.2
Table	56.3	0.3	62.6	0.1	62.4	1.1	63.4	0.5	63.9	0.3
Toaster	54.6	0.4	46.6	0.6	60.9	2.2	62.1	3.6	66.1	6.2
Toilet	49.1	0.6	62.0	0.9	63.5	0.9	55.5	1.1	59.4	3.7
TrashCan	50.9	0.7	55.5	0.6	54.8	1.0	53.5	0.6	55.3	1.1
USB	54.7	0.8	64.0	0.6	63.8	0.2	60.2	1.0	63.8	1.8
WashingMachine	47.6	0.4	51.2	0.9	52.2	3.2	57.2	1.2	55.6	0.9
Window	84.6	0.1	79.7	0.3	79.2	0.9	85.2	0.1	84.3	0.8
Overall	65.3	0.4	70.5	1.3	71.0	1.7	71.3	1.0	72.8	1.1

Tab. A2 presents the results of the ablation study on all 45 categories of the PartNet-E dataset. Here, Base denotes a baseline where the aggregated 3D features are directly segmented using a simple MLP. AVE replaces the SE module and the process in Eq. (2) with average pooling and average unpooling operations.

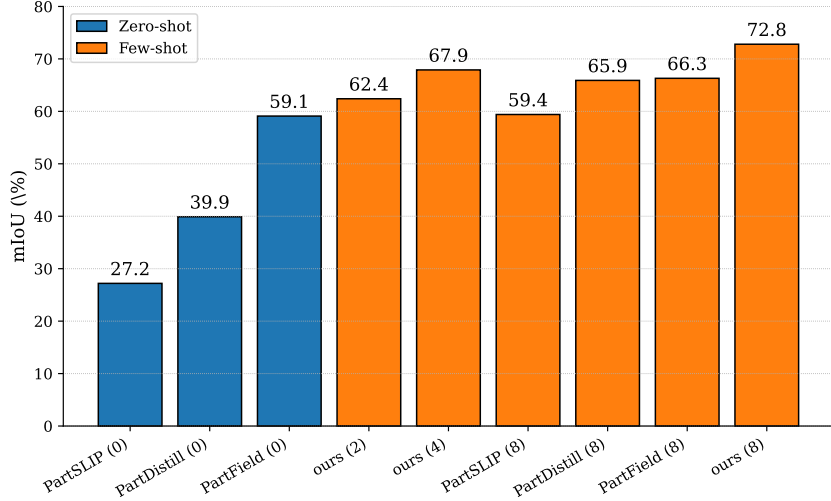


Figure A1: Evaluation under varying numbers of shots. We evaluated our method under different shot settings. In the figure, blue indicates zero-shot methods, while orange represents few-shot methods. The number in parentheses following each method name on the x-axis denotes the number of shots used during training.

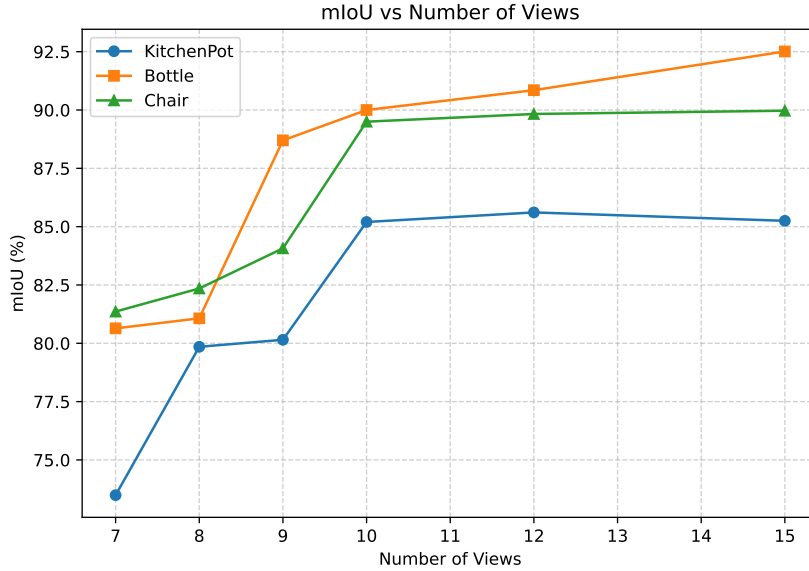


Figure A2: Evaluation under varying numbers of views.

C HyperParameter Analysis

The Number of Shots. We evaluate the method using varying numbers of shots, with detailed results presented in Fig. A1. As shown in the figure, our method already outperforms PartField—which relies on knowledge distillation from large-scale 3D data—in the 2-shot setting. Remarkably, with only 4 shots, our method achieves superior performance compared to all prior methods even under the 8-shot setting.

The Number of Views. We conducted experiments on Kitchen Pot, Bottle, and Chair under different numbers of rendered views, as shown in Fig. A2. When the number of views exceeds 10, the performance improvement becomes marginal as more views are added. In contrast, when the number of views falls below 10, the model performance drops sharply, underscoring the critical role of sufficient multi-view information for ensuring the model’s effectiveness.

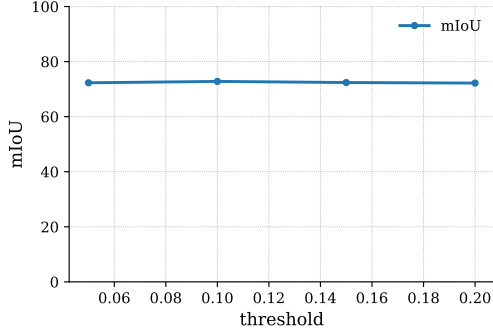


Figure A3: Effect of different overlap thresholds

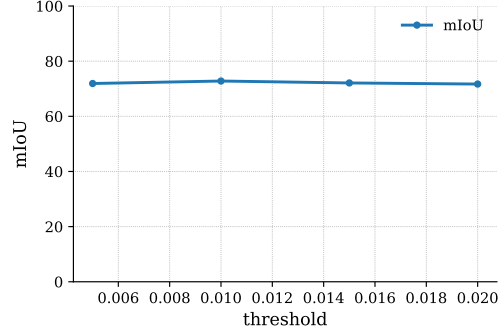


Figure A4: Effect of different adjacent thresholds

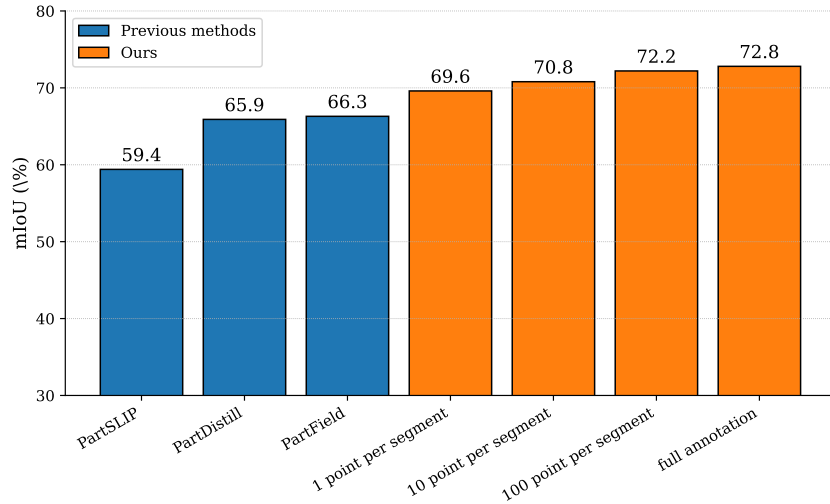


Figure A5: Evaluation under different numbers of annotation points. Under the 8-shot setting, we compared the label-efficient variants of our method with other methods trained using fully annotated data. In the figure, the blue bars represent the performance of methods trained with full labels under the 8-shot configuration, while the orange bars correspond to the performance of our method with varying numbers of labeled points.

Overlap and Adjacency Thresholds. When constructing the segment graph, we determine whether two segments from different viewpoints are overlapping based on the proportion of overlapping points between them, and whether this proportion exceeds a predefined threshold. We conduct experiments with different threshold values, as shown in Fig. A3. Results suggest that the model’s performance remains largely unaffected. We hypothesize that this robustness stems from the high precision of the segmentation produced by SAM, which allows accurate identification of overlapping relationships even when the threshold is set relatively low. Similarly, we conduct experiments with different minimum thresholds used to determine whether two segments are adjacent. As shown in Fig. A4, threshold variations within a reasonable range have no significant impact on the results. The above results were validated on the PartNet-E dataset.

Results for Label Efficient Annotations. Given that annotating 3D data is labor-intensive and challenging, we also explore applications in label-efficient learning. Leveraging the strong internal semantic consistency of segments obtained by SAM, we perform sparse annotations within each segment, randomly labeling 1, 10, or 100 points per segment. This evaluation was conducted on the PartNet-E dataset. As shown in Fig. A5, even when only a single point is annotated within each segment, our method still achieves state-of-the-art performance compared to previous few-shot approaches.

Table A3: Performance comparison on coarse- and fine-grained part segmentation on the 3DCom-PaT++ dataset.

Method	Coarse			Fine		
	Airplane	Bicycle	Chair	Airplane	Bicycle	Chair
PartSLIP++ (zero-shot)	11.56	60.52	16.59	9.48	2.60	4.87
Ours (1-shot)	93.31	91.21	41.41	45.34	26.26	27.21
Ours (8-shot)	94.64	95.46	69.26	65.38	54.17	42.61

D Experiment on More Complex Datasets

Our method can be extended to multi-granularity once the annotation is available. For example, 3DComPat++ [A4] provides coarse-grained annotations (at most 5 part categories) and fine-grained annotations (10-20 part categories). We selected three categories—airplane, bicycle, and chair—for evaluation. Airplanes have 14 fine-grained and 3 coarse part labels; bicycles have 26 fine-grained ones and 3 coarse ones; chairs have 22 fine-grained labels and 4 coarse labels. We extend to multi-grained segmentation by tuning separate networks. In the Tab. A3, our method demonstrates significant advantages across different levels of part annotation granularity.

E Robustness to Noise and Real-World Data

To assess the robustness of the proposed method to noisy real-world data, experiments were conducted by injecting Gaussian noise into the normalized point cloud space. As shown in Tab. A4. Specifically, noise with standard deviations of 0.005, 0.01, 0.02, 0.05 was added to evaluate the degradation trend in segmentation performance.

Table A4: Performance under different noise levels (mIoU, %).

Category	0	0.005	0.01	0.02	0.05
KitchenPot	85.2	81.83	80.29	75.01	66.68
Bottle	90.0	83.98	78.99	74.88	69.04
Chair	89.5	87.03	82.34	78.02	67.58

F Comparison with Semantic Mask-Based Approaches

The comparison with representative semantic mask-based methods, including OpenMask3D [A5], SATR [A1], and Find3D [A2], is presented in Table A5. As shown, our method consistently achieves a significant performance advantage across all evaluation metrics. This improvement stems from two main factors. First, OpenMask3D relies on Mask3D [A3] as its 3D feature extractor, which is pre-trained on the indoor scene dataset ScanNet, leading to poor generalization to the more diverse and geometrically complex 3D shapes in PartNetE. Second, as highlighted in the Find3D paper, PartNetE poses substantial challenges due to numerous small parts with minimal geometric or color distinctiveness (e.g., buttons on uniformly colored surfaces), which degrades the performance of both Find3D and SATR. In contrast, our method excels in identifying and segmenting such small, indistinct components, which underscores its robustness in fine-grained part segmentation.

G Graph Construction from Multi-View Segments

The graph in our framework is constructed directly from the segments generated by applying SAM to multi-view images. Each segment corresponds to a node, and edges are established based on adjacency and overlap relationships among these segments. To quantitatively analyze the graph scale, we report the average and maximum numbers of nodes, adjacent edges, and overlapping edges per object on the PartNet-E dataset in Table A6.

Table A5: Comparison with semantic mask-based approaches on PartNet-E.

Setting	Method	mIoU (%)
Zero-shot	OpenMask3D	12.54
	Find3D	16.38
	SATR	19.42
Few-shot	Ours (1-shot)	53.36
	Ours (8-shots)	72.80

Table A6: Statistics of the constructed graphs on PartNet-E.

Metric	Nodes	Adjacent Edges	Overlapping Edges
Mean	136.53	310.58	2386.84
Max	1121	1872	21024

As shown in Table A6, the constructed graphs, with an average of a few hundred nodes and several thousand edges, are significantly smaller and more compact than the raw point clouds of each object in PartNet-E, which typically contain hundreds of thousands of points. This compact representation enables efficient structural reasoning without compromising geometric detail.

H More Qualitative Results

In Figs. A6–A8, we present additional segmentation results on PartNet-E, focusing on categories that contain small parts, such as cameras and remote controls with button parts, doors with handle parts, and lamps with bulb parts. As illustrated in the figures, our method achieves higher accuracy in segmenting these fine-grained components.

I More Results on Feature Similarity

In Figs. A9–A12, we present additional visualizations of feature similarity across different parts. For each shape, we select an anchor point (indicated by the red arrow in the first column) and compute the Euclidean distance between the anchor and all other points—both within the same shape (first column) and across different shapes (remaining columns). The distances are then normalized for visualization. Points with higher similarity (i.e., smaller distances) are shown in blue, while less similar points appear in gray. As illustrated in the results, features from the same part category exhibit strong consistency both within a single shape and across different shapes, whereas features from different part categories demonstrate clear dissimilarity.

Appendix A References

- [A1] Ahmed Abdelreheem et al. “SATR: Zero-Shot Semantic Segmentation of 3D Shapes”. In: *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*. IEEE, 2023, pp. 15120–15133.
- [A2] Ziqi Ma, Yisong Yue, and Georgia Gkioxari. “Find Any Part in 3D”. In: *CoRR* abs/2411.13550 (2024).
- [A3] Jonas Schult et al. “Mask3D: Mask Transformer for 3D Semantic Instance Segmentation”. In: *IEEE International Conference on Robotics and Automation, ICRA 2023, London, UK, May 29 - June 2, 2023*. IEEE, 2023, pp. 8216–8223.
- [A4] Habib Slim et al. “3DCoMPaT⁺⁺: An improved Large-scale 3D Vision Dataset for Compositional Recognition”. In: *CoRR* abs/2310.18511 (2023).
- [A5] Ayça Takmaz et al. “OpenMask3D: Open-Vocabulary 3D Instance Segmentation”. In: *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems*. Ed. by Alice Oh et al. 2023.

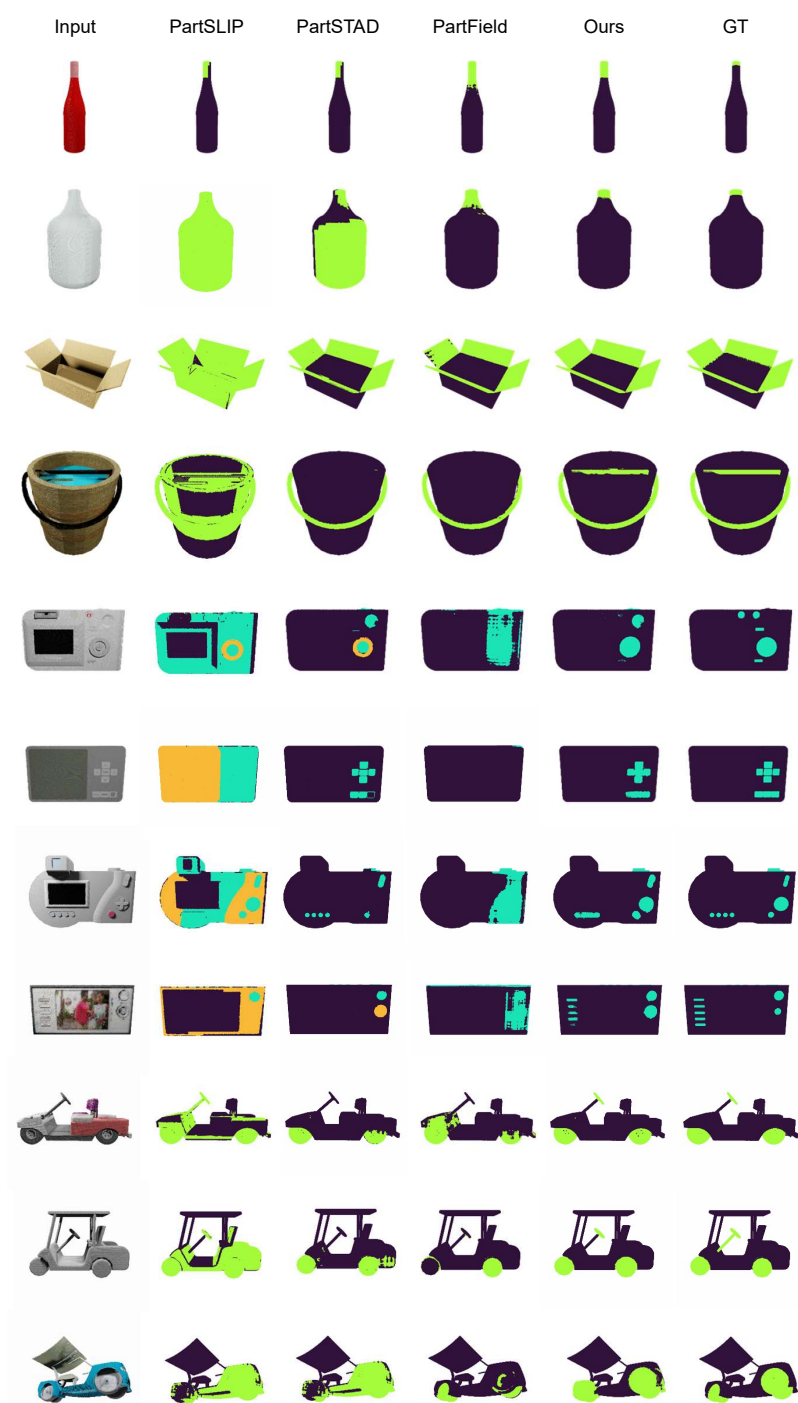
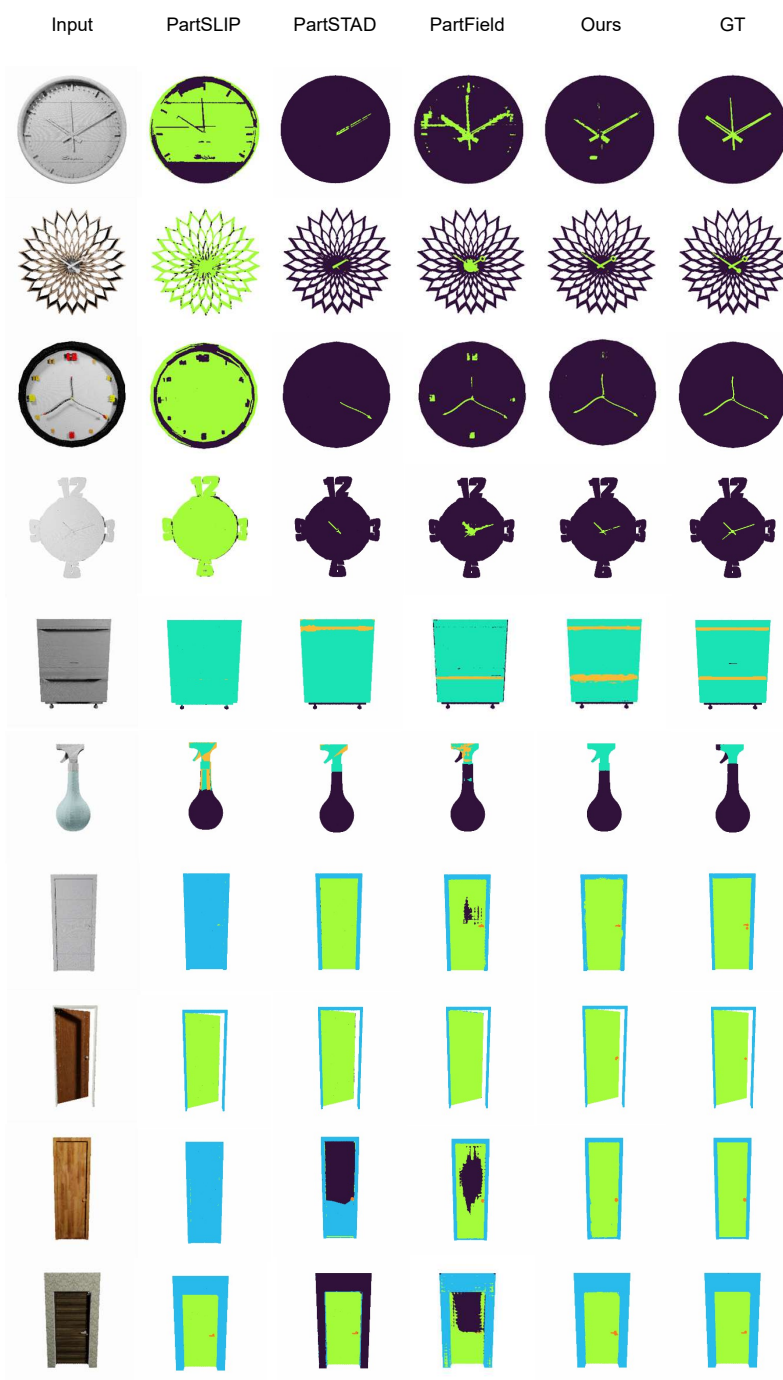
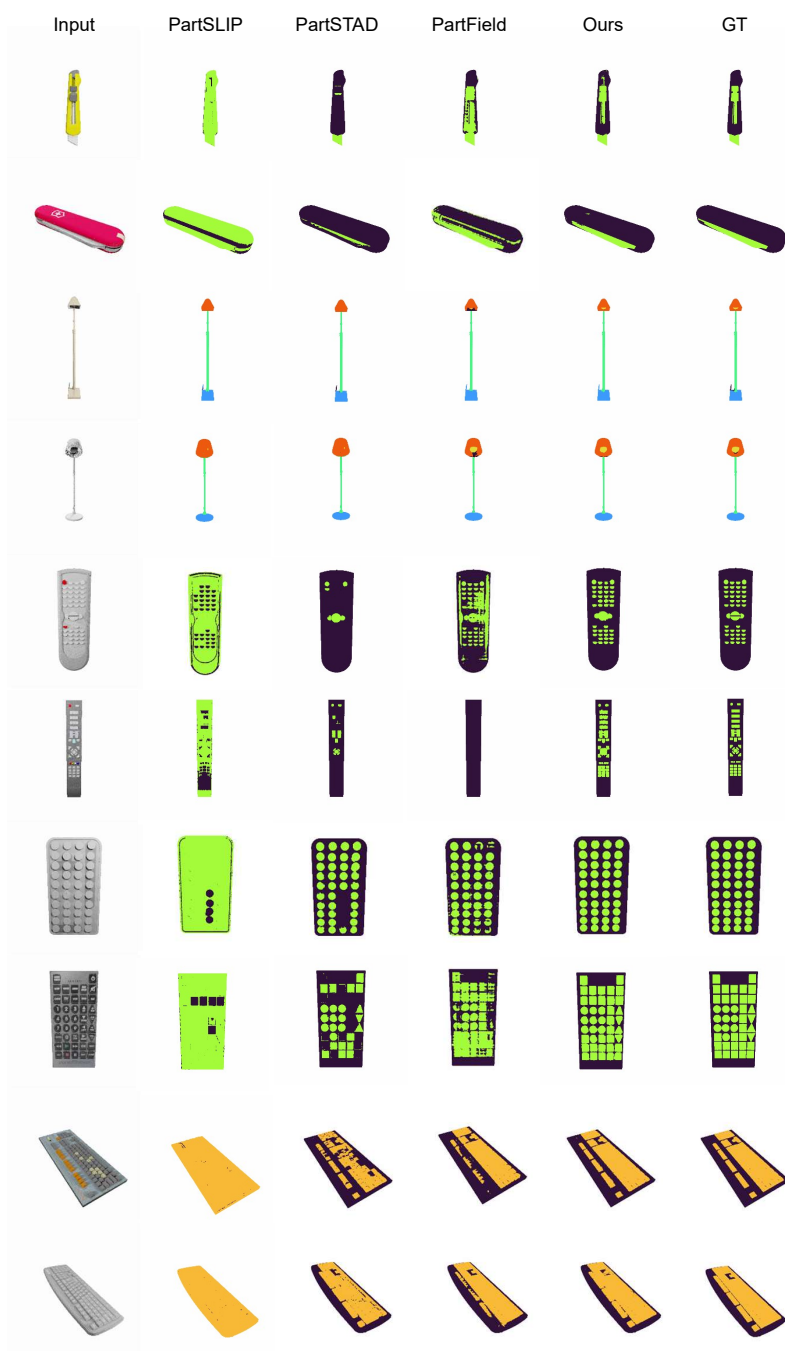


Figure A6: Qualitative comparison of part segmentation results





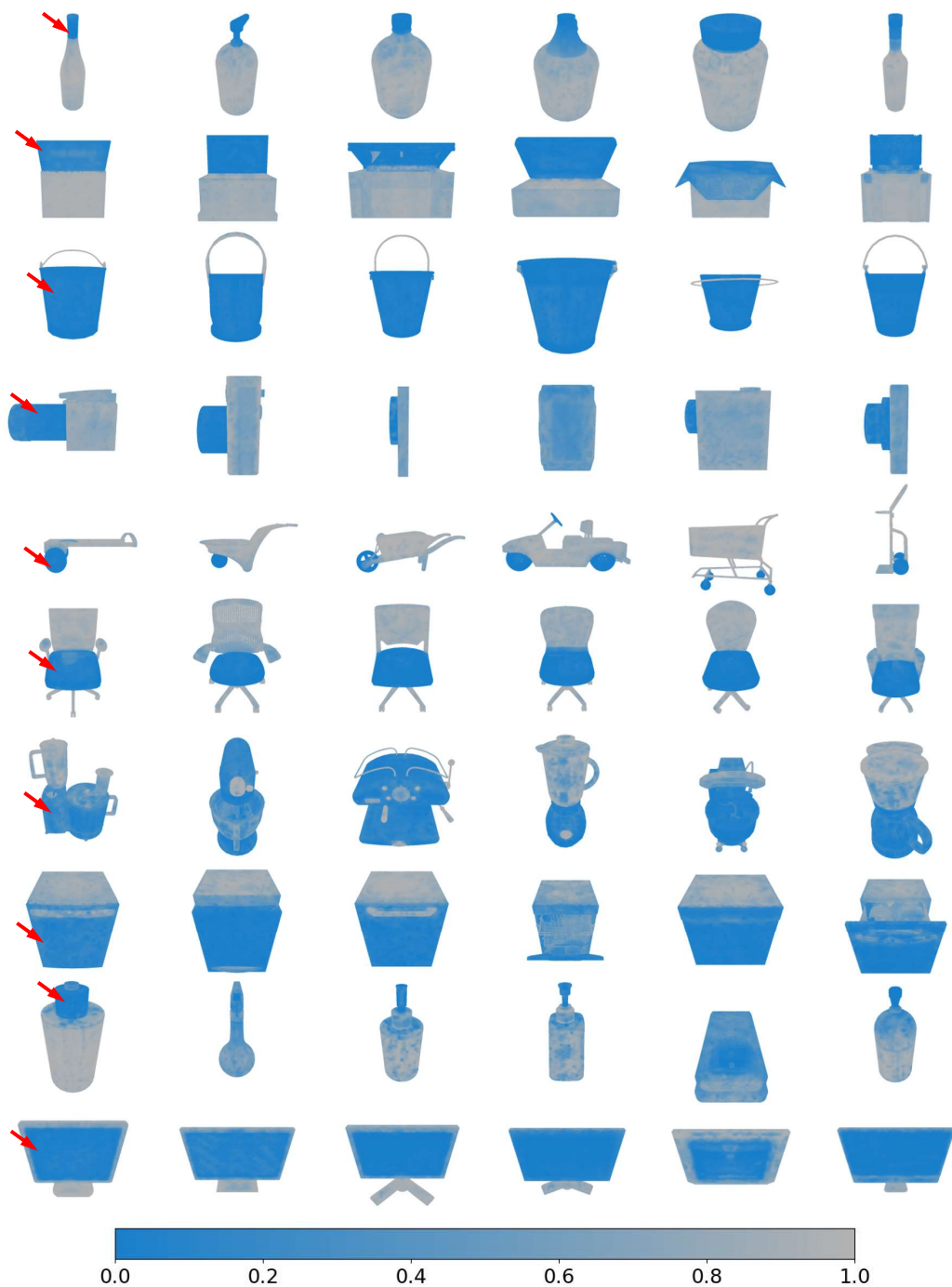


Figure A9: Feature similarities between an anchor point (indicated by red arrows) and points within the same shape (leftmost column) as well as points from different shapes (the five columns on the right). Color indicates the distance between features: smaller distances are shown in blue, while larger distances are shown in gray.

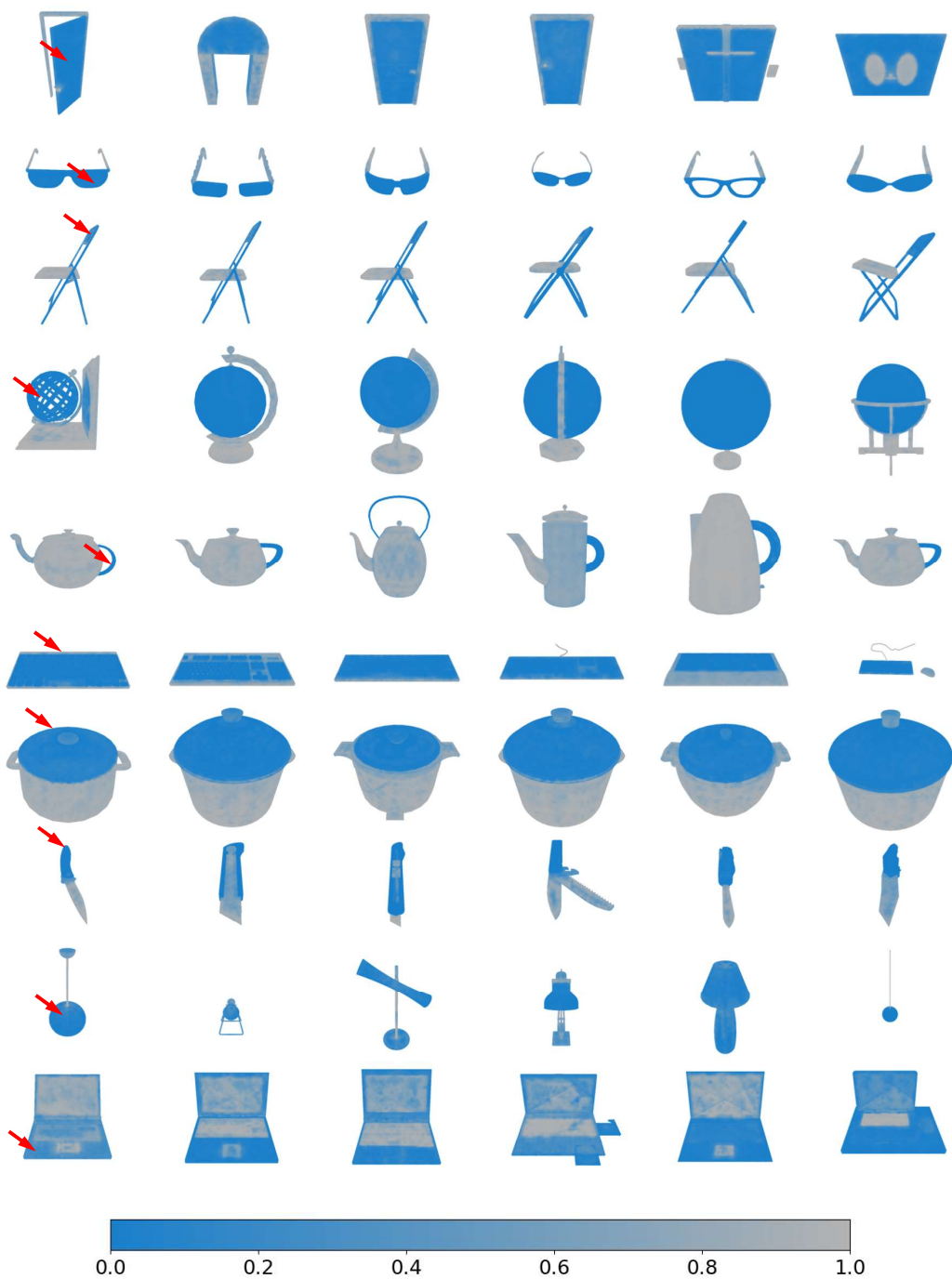


Figure A10: Feature similarities between anchor points and intra-/inter-shape points. See caption above for details.

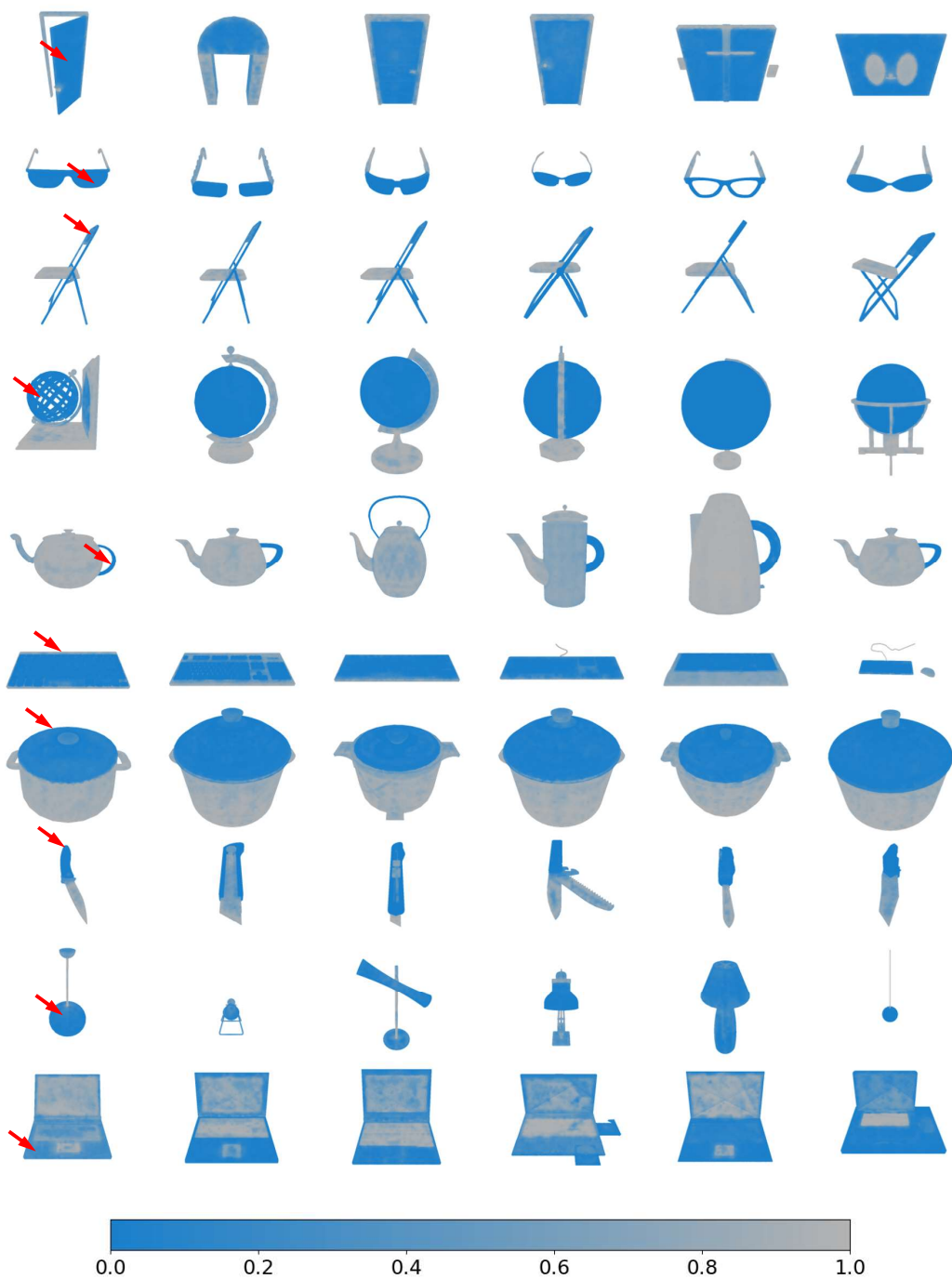


Figure A11: Feature similarities between anchor points and intra-/inter-shape points. See caption above for details.

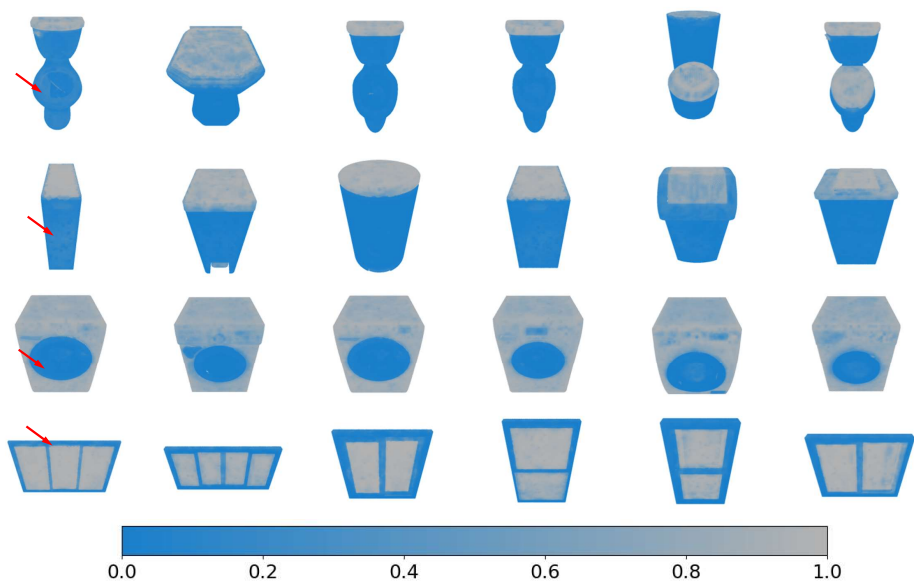


Figure A12: Feature similarities between anchor points and intra-/inter-shape points. See caption above for details.