GeoSAM2: Unleashing the Power of SAM2 for 3D Part Segmentation

Supplementary Material

Figure 1. Examples of training data. Each color represents a different part.

In supplmentary material, we first further provide implementation details. We also provide additional discussion on ablation study. Finally, we provide additional visual results. We encourage readers to view our accompanying videos, which showcase the rotation of objects rendered with different colors (each color represents different parts) as presented in the paper.

0.1. Implementation Details

0.1.1 Lift Multi-View Masks to 3D

Given multi-view masks $\{M_i\}_{i=1}^N$ with corresponding camera pose $\{P_i\}_{i=1}^N$ and target mesh $\mathcal{M}$ (normalized to $[-0.5, 0.5]^3$), where $N = 12$, we first randomly sample 5 points on each face of mesh $\mathcal{M}$, and project them to each mask, obtaining depth map $\{D_i\}_{i=1}^N$. To verify the visibility of each point on each, we compare the depth map $\{D_i\}_{i=1}^N$ with the rendered depth map $\{D_ri\}_{i=1}^N$, then adopt those points with depth error $|D_i - D_ri|_{i=1}^N$ less than 0.001, which precisely reflect the visibility. After obtaining the labels of each point on masks $\{M_i\}_{i=1}^N$, we select the maximum label from those visible multi-view labels of points to get the point-level label.

0.1.2 Labeling Training Data

We built a large-scale dataset of 180k meshes from open-world repositories (Objaverse [1], TexVerse [3], Objaverse-XL [2]). We strictly isolated the test set by filtering out all PartObjaverse-Tiny/PartNetE object IDs via metadata matching to ensure zero overlap. Visualization examples are shown in Figure 1.

Annotation Pipeline. We employed a Hierarchical Semi-Automatic Pipeline to ensure scalability and quality:

Table 1. Quantitative evaluation of class-agnostic part segmentation on PartObjaverse-Tiny for ablation study. Instance-level labels; (mean IoU, %).

Method	Human	Animals	Daily	Build.&	Transp.	Plants	Food	Elec.	Avg.
Vanilla SAM2	67.04	64.34	64.37	54.88	52.05	75.78	67.16	65.46	62.59
w/o point map	81.17	83.87	77.68	67.24	63.85	81.66	81.91	78.08	75.56
w/o feature fusion	87.27	90.87	79.98	70.81	74.75	87.64	**83.19**	82.20	81.39
Ours	**88.99**	**91.50**	**86.04**	**74.57**	**77.60**	**88.92**	82.72	**84.95**	**84.06**

1) Geometric Initialization: Meshes are over-segmented using connected components and geometric heuristics (distance affinity, intersection analysis). 2) Human-in-the-Loop Refinement: Annotators verified the segmentation quality. Over-segmented regions were interactively merged via a custom interface to align with logical object parts.

0.1.3 Class Grouping of PartNetE

To ensure consistency with previous methods, we group the 45 categories in PartNetE following the categorization protocol used in prior work. The specific category groupings are as follows.

- **Electronics & Computing Devices:** Keyboard, Mouse, Laptop, Phone, Camera, USB, Display (monitor), Remote, Printer, Switch (if treated as a network or power switch)
- **Large Home Appliances:** Washing Machine, Dishwasher, Refrigerator, Oven, Microwave
- **Kitchen & Food-Related Items:** KitchenPot, Kettle, Toaster, CoffeeMachine, Faucet, Dispenser, Knife, Bottle, Bucket (often used in kitchen/cleaning contexts)
- **Furniture & Household Infrastructure:** Table, Chair, FoldingChair, StorageFurniture, Door, Window, Lamp, TrashCan, Safe (often a household or office fixture)
- **Tools, Office Supplies, & Miscellaneous:** Stapler, Scissors, Pen, Pliers, Lighter, Box, Cart (e.g., utility cart), Globe (decorative/educational), Suitcase (travel/personal), Eyeglasses (personal), Clock

0.2. Additional Ablation Study

0.2.1 Per-Category Results for Ablation Study

Due to space constraints in the main paper, we only reported the average performance across categories in our ablation study on different network designs. Here, we provide the full per-category results on PartObjaverse-Tiny and PartNetE to offer a more comprehensive view of the model's behavior across different object types, illustrated in 1 and 2.

Table 2. Quantitative evaluation of class-agnostic part segmentation on PartNetE for abltion study. Instance-level labels; (mean IoU, %).

Method	Electro. & Comput.	Home Appl.	Kitchen & Food	Furnit. & Househo.	Tools, Offi., & Misc.	Avg.
Vanilla SAM2	64.33	69.26	71.73	65.10	65.41	66.55
w/o point map	68.47	69.11	77.74	68.74	73.19	71.26
w/o feature fusion	**70.18**	**71.29**	78.00	68.08	75.05	72.25
Ours	69.93	71.23	**79.73**	**71.97**	**79.41**	**74.42**

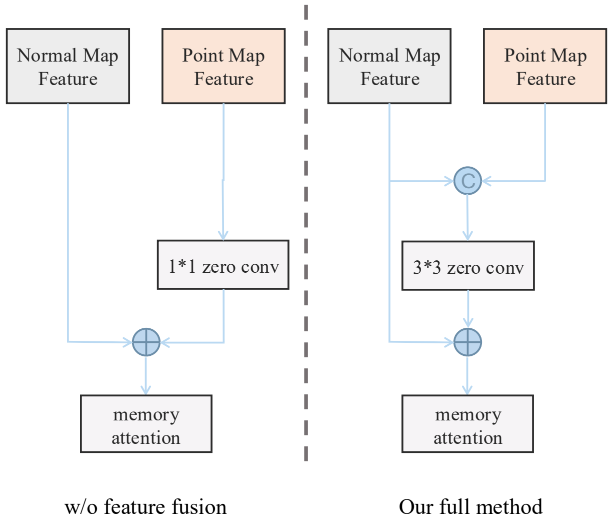

Figure 2. Illustrations of w/o feature fusion and our full method in ablation study.

0.2.2 Ablation Study on Feature Fusion

To provide a clearer understanding of the design choices in our ablation study on the feature fusion module, we include detailed visualizations of the different network variants used. Figure 2 highlight the architectural differences between each setting, helping to better illustrate how various fusion strategies affect performance.

0.3. Additional Visualization

0.3.1 Additional Comparison Visualization

To further illustrate the superiority of our approach and its robust 3D segmentation performance, we present additional qualitative comparisons on PartObjaverse-Tiny, as shown in Figure 3 and Figure 4.

0.3.2 Additional Hierachical segmentation Visualization

We include additional qualitative examples of our hierarchical segmentation results in this subsection, as shown in Figure 5, Figure 6, and Figure 7. These visualizations further demonstrate the effectiveness of our method in producing semantically meaningful and consistent part hierarchies across diverse 3D objects. The results are rendered from multiple viewpoints to better illustrate the multi-level structure and fine-grained part delineation.

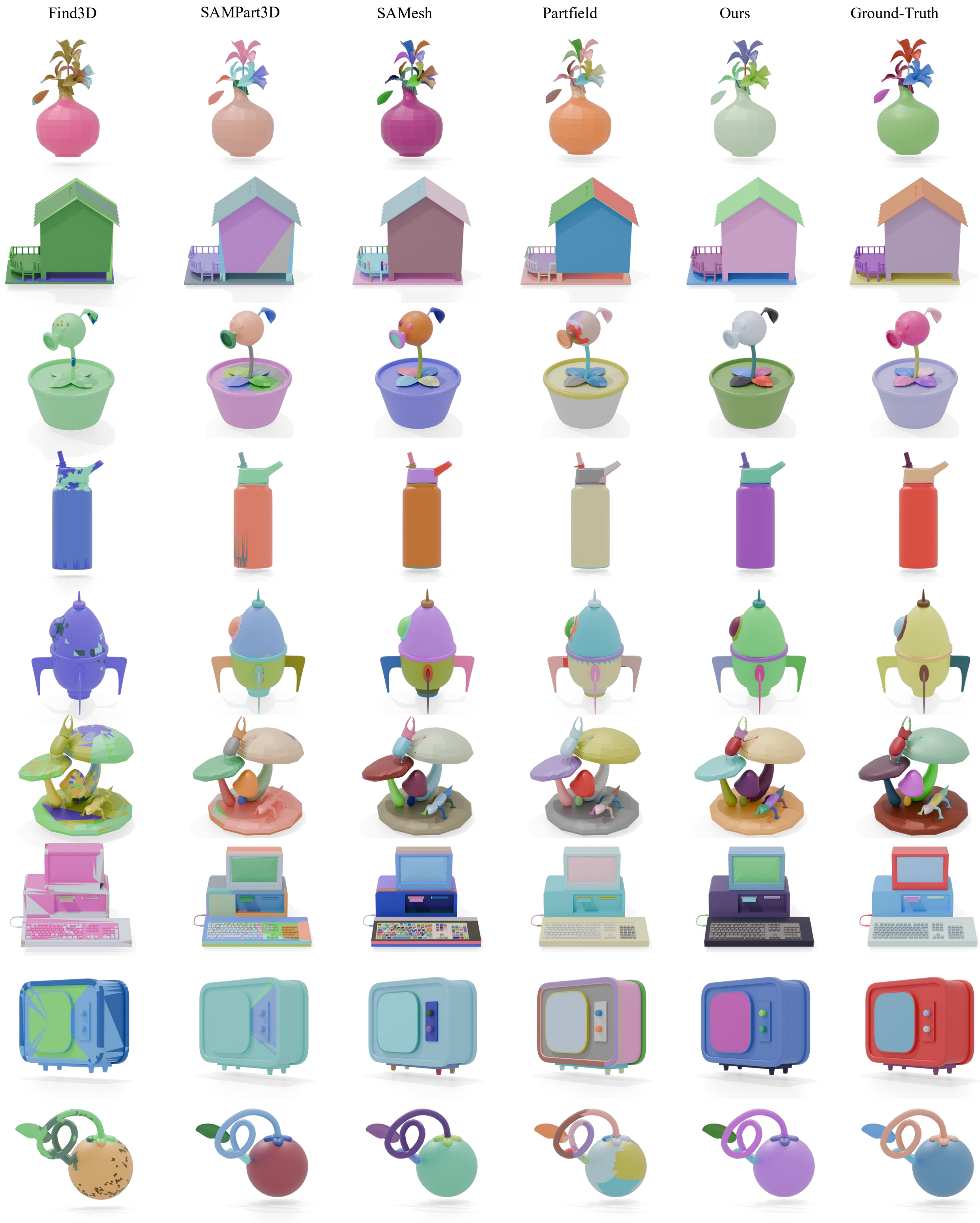

Figure 3. More comparison results on PartObjaverse-Tiny

.

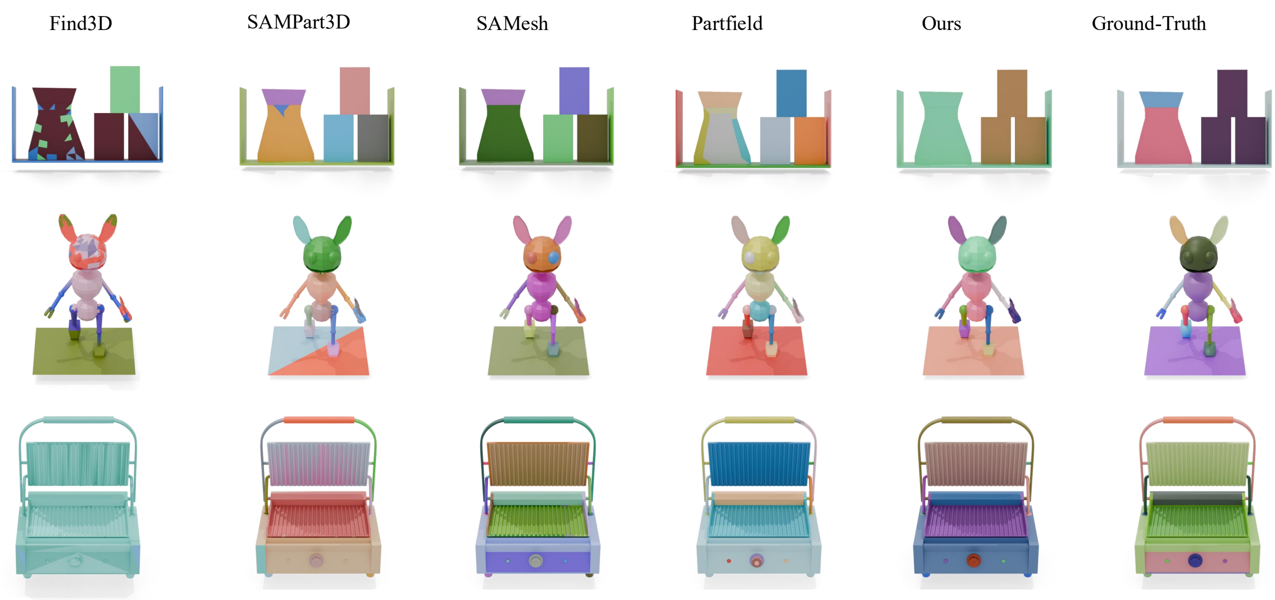

Figure 4. More comparison results on PartObjaverse-Tiny
.

Figure 5. More hierarchical segmentation visualization. 3D models are segmented hierarchically, starting with the finest level of detail at the top, and then proceeding with varying granularities of segmentation for the hands, legs, and head.

.

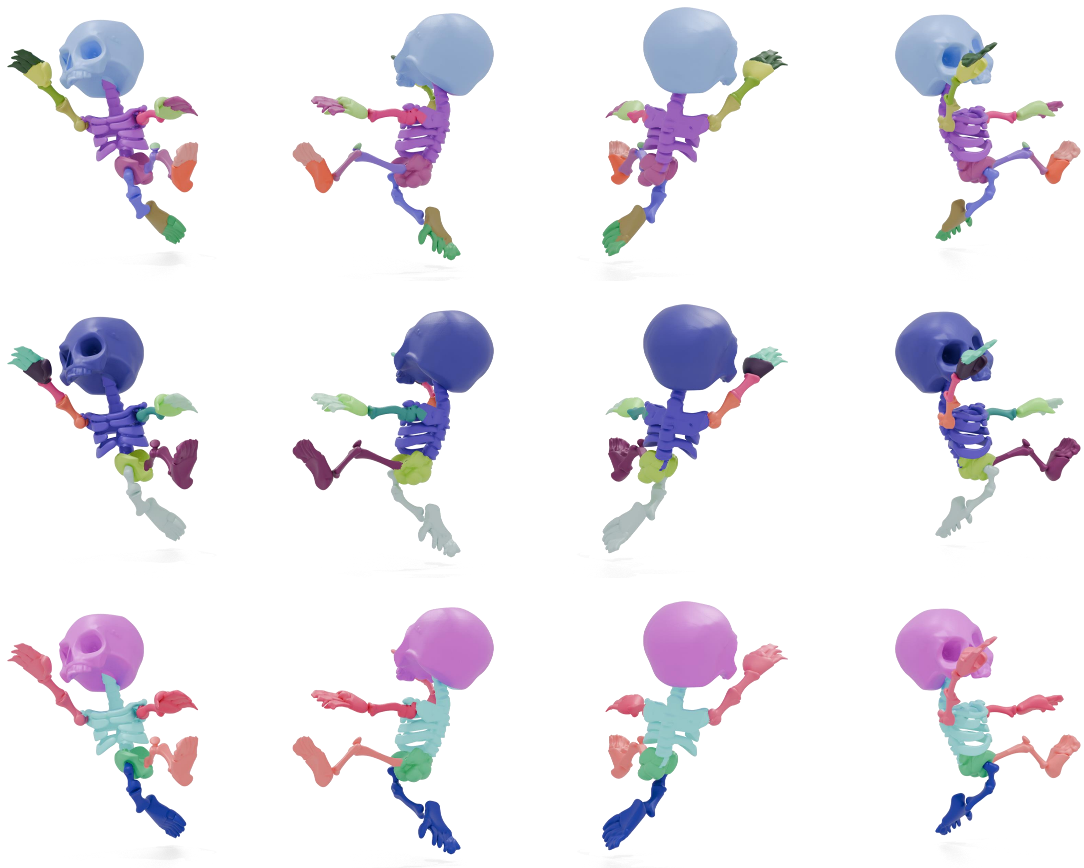

Figure 6. More hierarchical segmentation visualization. 3D models are segmented hierarchically, starting with the finest level of detail at the top, and then proceeding with varying granularities of segmentation for the hands, legs, and head.

.

Figure 7. More hierarchical segmentation visualization. 3D models are segmented hierarchically, starting with the finest level of detail at the top, and then proceeding with varying granularities of segmentation for the hands, legs, and head.

.

References

[1] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *CVPR*, 2023. 1

[2] Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, et al. Objaverse-xl: A universe of 10m+ 3d objects. In *NeurIPS*, 2024. 1

[3] Yibo Zhang, Li Zhang, Rui Ma, and Nan Cao. Texverse: A universe of 3d objects with high-resolution textures. *arXiv preprint arXiv:2508.10868*, 2025. 1