

EI-Part: Explode for Completion and Implode for Refinement

Supplementary Material

1. Network Details

FluxDev-1 [2] is employed as the backbone of MVSegNet, while InSegNet adopts the large texturing architecture from UniTex [4].

In the **Explode for Completion** stage, the DiT network utilizes the *Sparse Latent Flow* architecture from Trellis [7], which comprises 24 Transformer blocks. Since we adopt the Flux-Kontext [1] training strategy, a lightweight positional embedding is applied to distinguish noise latent tokens from conditional latent tokens. The resulting embeddings are added to the token features before being input into the DiT blocks.

For the **Implode for Refinement** stage, we utilize the Sparc3D [3] VAE with a latent resolution of 64^3 . To avoid degradation in reconstruction quality, the downsampling operation in the original Stage-II Trellis [7] DiT network is removed, while the internal structure of each DiT block remains unchanged. Positional embeddings are consistently applied to both noise and conditional tokens to encode spatial positional information before being processed by the DiT blocks.

2. Training Details

The training data pairs for the two-stage DiT are constructed as follows:

- **Incomplete part meshes:** These are obtained by first converting the complete mesh into a watertight representation, then discarding the internal structures while preserving the face correspondence information. The resulting meshes are used as the data for incomplete part meshes.
- **Complete part meshes:** These are directly extracted from the original GLB files. Each individual part is processed to ensure watertightness and is then used as the ground-truth (GT) part mesh data.

In the first stage, the VAE is fine-tuned on the curated part dataset, and all mesh data are encoded into the latent space for DiT training. In the second stage, the original Sparc3D [3] VAE is employed to encode meshes into latents, where the latents of incomplete sub-meshes serve as conditions, and the active voxel positions are guided by the GT part meshes.

3. More Comparison

Comparison with Baseline Methods. Fig. 1 presents additional qualitative results compared to state-of-the-art (SOTA) methods. As illustrated, BANG [11] and PartPacker [6] struggle to capture fine geometric details that correspond

with the input model. This issue stems from their practice of embedding all components into a single or dual set of latent tokens, which limits the representational capacity of individual parts and ultimately leads to inaccuracies in geometric detail. HoloPart [9] neglects the interrelationships between parts, leading to unnatural completion outcomes that resemble merely filling a hole, thus compromising structural coherence. While PartCrafter [5] and X-Part [8] utilize a cross-attention mechanism to enhance structural coherence, their reliance on fixed-length part tokens proves inflexible for components of varying sizes. This limitation results in inefficient resource allocation, where larger components may be underrepresented while resources are wasted on smaller parts. OmniPart [10] employs voxel-based representations to adaptively allocate spatial resolution to different parts; however, it supports completion only within active voxels, restricting the available spatial extents for completion. This constraint can lead to implausible part geometries and reduced geometric fidelity. In contrast, our method excels in generating high-quality 3D shapes with well-defined parts, demonstrating significant performance improvements in structural coherence, geometric plausibility, geometric fidelity, and efficiency.

Ablation Study on Explode and Implode. Tab. 1 presents the quantitative comparisons between our method and alternative strategies. Our proposed method outperforms all the other options, achieving the highest scores.

4. Supplementary Rebuttal Visualizations and Metrics

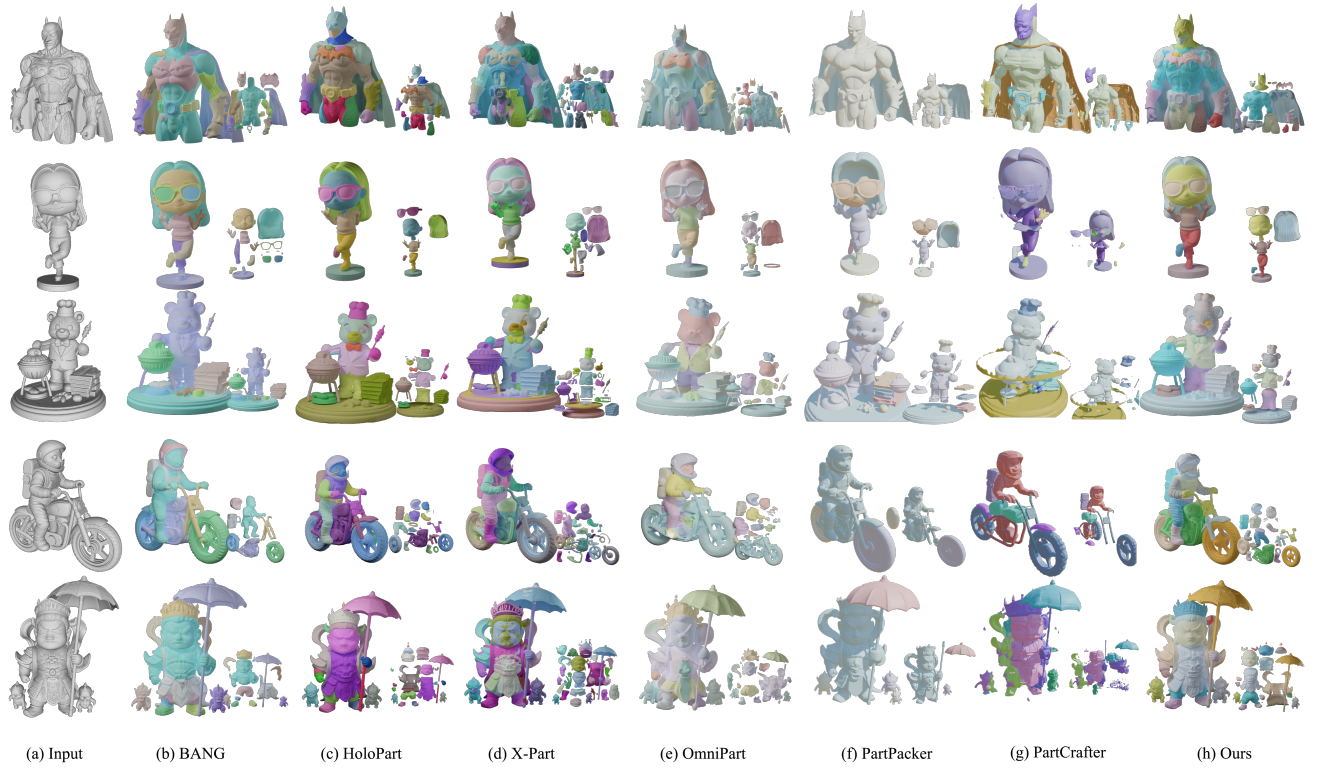


Figure 1. Qualitative comparison between ours and the state of the arts. For each row, the input model (a) is shown alongside the parts generated by (b) BANG [11], (c) HoloPart [9], (d) X-Part [8], (e) OmniPart [10], (f) PartPacker [6], (g) PartCrafter [5], and (h) our method. Our method demonstrates exceptional proficiency in generating high-quality 3D shapes with well-defined parts, achieving significant improvements in structural coherence, geometric plausibility, fidelity, and efficiency.

Table 1. Quantitative comparison across different settings at both the part and overall object levels. Since the part-level ground truth (GT) is derived from the complete results, the absence of the Explode operation results in lower part completion quality, underscoring its significance. Meanwhile, the explode–implode refinement slightly compromises overall mesh accuracy, revealing a trade-off between enhancing part-level details and maintaining global reconstruction precision.

Setting	Part-level						Overall-object					
	Voxel IOU \uparrow	CD \downarrow	Voxel F-Score 0.01 \uparrow	F-Score 0.1 \uparrow	F-Score 0.05 \uparrow	F-Score 0.01 \uparrow	Voxel IOU \uparrow	CD \downarrow	Voxel F-Score 0.01 \uparrow	F-Score 0.1 \uparrow	F-Score 0.05 \uparrow	F-Score 0.01 \uparrow
w/o Explode	0.7647	0.0153	0.8883	0.9975	0.9866	0.7599	0.9310	0.0092	0.9642	1.0000	1.0000	0.9680
w/o Implode	0.7992	0.0168	0.8664	0.9998	0.9937	0.8678	0.8036	0.0171	0.8896	0.9769	0.9615	0.8243
Whole	-	-	-	-	-	-	0.9005	0.0152	0.9091	0.9975	0.9869	0.8691

Table 2. Part-level quantitative comparison and inference cost. Compared with HoloPart, X-Part, OmniPart, and OmniPart-Mask, our method achieves the best performance on all reported geometric metrics, including Voxel IoU, Chamfer Distance, and F-scores at multiple thresholds. The OmniPart-Mask variant improves mask granularity but remains constrained by limited completion capacity within active voxels, which is consistent with the artifacts observed in qualitative results. Inference cost is also reported to show the efficiency-accuracy tradeoff of all methods.

Method	Voxel IOU \uparrow	CD \downarrow	Voxel F-Score@0.01 \uparrow	F-Score@0.1 \uparrow	F-Score@0.05 \uparrow	F-Score@0.01 \uparrow	Time (s) \downarrow
HoloPart	0.5232	0.1212	0.6417	0.8870	0.8634	0.5400	300
X-Part	0.4456	0.1845	0.5371	0.7133	0.6466	0.5200	200
OmniPart	0.4453	0.1846	0.5368	0.7134	0.6465	0.5142	80
OmniPart-Mask	0.4300	0.1733	0.4940	0.7024	0.6175	0.5000	80
Ours	0.7950	0.0859	0.8207	0.9075	0.8908	0.8259	252

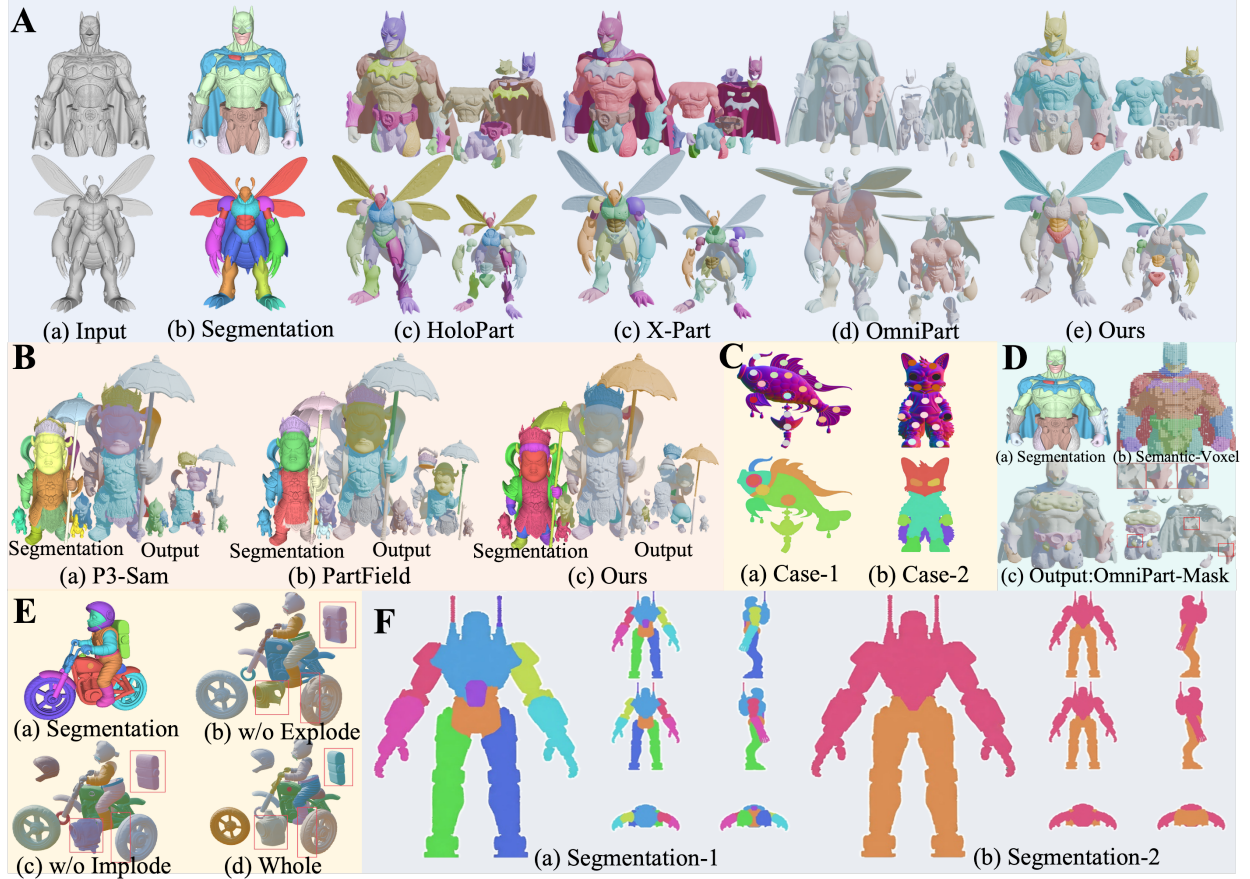


Figure 2. Additional qualitative results for part segmentation and completion (zoom in for better visualization). A) Comparisons with baseline methods under the same segmentation setting, showing stronger structural coherence and geometric fidelity. B) Results under different segmentation methods, demonstrating robustness to the segmentation source. C) SAM segmentation results on normal maps; SAM achieves comparable IoU on textured and normal renderings on PartObjaverse-Tiny (0.685 vs. 0.670), supporting the use of frontal normal-map segmentation. D) Demo of OmniPart-Mask, where completion capacity is limited and artifacts such as holes are visible. E) Part-exploded visualizations that explicitly show how spatial reallocation supports part completion before refinement. F) Multiview segmentations conditioned on different frontal masks, showing controllable and consistent multiview part decomposition.

References

- [1] Stephen Batifol, Andreas Blattmann, Frederic Boesel, Saksham Consul, Cyril Diagne, Tim Dockhorn, Jack English, Zion English, Patrick Esser, Sumith Kulal, et al. Flux. 1 kontext: Flow matching for in-context image generation and editing in latent space. *arXiv e-prints*, pages arXiv–2506, 2025. 1
- [2] Black Forest Labs. Flux. <https://github.com/black-forest-labs/flux>, 2024. 1
- [3] Zhihao Li, Yufei Wang, Heliang Zheng, Yihao Luo, and Bihan Wen. Sparc3d: Sparse representation and construction for high-resolution 3d shapes modeling. *arXiv preprint arXiv:2505.14521*, 2025. 1
- [4] Yixun Liang, Kunming Luo, Xiao Chen, Rui Chen, Hongyu Yan, Weiyu Li, Jiarui Liu, and Ping Tan. Unitex: Universal high fidelity generative texturing for 3d shapes. *arXiv preprint arXiv:2505.23253*, 2025. 1
- [5] Yuchen Lin, Chenguo Lin, Panwang Pan, Honglei Yan, Yiqiang Feng, Yadong Mu, and Katerina Fragkiadaki. Partcrafter: Structured 3d mesh generation via compositional latent diffusion transformers. *arXiv preprint arXiv:2506.05573*, 2025. 1, 2
- [6] Jiaxiang Tang, Ruijie Lu, Zhaoshuo Li, Zekun Hao, Xuan Li, Fangyin Wei, Shuran Song, Gang Zeng, Ming-Yu Liu, and Tsung-Yi Lin. Efficient part-level 3d object generation via dual volume packing. *arXiv preprint arXiv:2506.09980*, 2025. 1, 2
- [7] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3d latents for scalable and versatile 3d generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 21469–21480, 2025. 1
- [8] Xinhao Yan, Jiachen Xu, Yang Li, Changfeng Ma, Yunhan Yang, Chunshi Wang, Zibo Zhao, Zeqiang Lai, Yunfei Zhao, Zhuo Chen, et al. X-part: high fidelity and structure coherent shape decomposition. *arXiv preprint arXiv:2509.08643*, 2025. 1, 2
- [9] Yunhan Yang, Yuan-Chen Guo, Yukun Huang, Zi-Xin Zou, Zhipeng Yu, Yangguang Li, Yan-Pei Cao, and Xihui Liu. Holopart: Generative 3d part amodal segmentation. *arXiv preprint arXiv:2504.07943*, 2025. 1, 2
- [10] Yunhan Yang, Yufan Zhou, Yuan-Chen Guo, Zi-Xin Zou, Yukun Huang, Ying-Tian Liu, Hao Xu, Ding Liang, Yan-Pei Cao, and Xihui Liu. Omnipart: Part-aware 3d generation with semantic decoupling and structural cohesion. *arXiv preprint arXiv:2507.06165*, 2025. 1, 2
- [11] Longwen Zhang, Qixuan Zhang, Haoran Jiang, Yinuo Bai, Wei Yang, Lan Xu, and Jingyi Yu. Bang: Dividing 3d assets via generative exploded dynamics. *ACM Transactions on Graphics (TOG)*, 44(4):1–21, 2025. 1, 2