Dehallu3D: Hallucination-Mitigated 3D Generation from Single Image via Cyclic View Consistency Refinement

Supplementary Material

1. Implementation details

Our experiments focus on both mesh visual quality and mesh geometric quality. Following previous work [5, 7, 8, 10], we perform our experiments on the Google Scanned Objects (GSO) dataset [2]. For a fair experimental comparison, all objects in the GSO dataset are rendered at a resolution of 512×512 with Blender Cycles rendering engine as the input images for all methods. For geometric quality evaluation, all generated meshes are normalized to the bounding box of $[-1, 1]$ to ensure alignment, where 30k points are sampled for each mesh for geometric evaluation. For visual quality evaluation, we select elevation angles from $\{0, 15, 30\}$ and 8 evenly distributed azimuth angles to render each mesh generated from different methods, resulting in 24 views per mesh. We employ PSNR, SSIM, LPIPS, and Clip-Similarity (Clip-Sim) as metrics to evaluate the visual quality. For geometric quality evaluation, we utilize Chamfer Distance (CD) and F-Score as metrics.

We utilize the pre-trained high-resolution multi-view image generator and high-resolution multi-view normal generator from Unique3D [8] with the same settings. We rely on supervision from orthogonal views in both mesh initialization and mesh coarse reconstruction. The initialized mesh is inferred from normal maps with a resolution of 256×256, and subsequently simplified to a mesh with 2,000 faces. To strike a balance between efficiency and generation quality, we set the parameter $V = 72$ in the Cyclic View Consistency Refinement (CVCR) module, resulting in an angle difference of $360^\circ/72 = 5^\circ$ between two adjacent views. We set the loss weights as $\lambda_1 = 1$ and $\lambda_2 = 10$ based on empirical evaluation.

In coarse mesh reconstruction and CVCR module, the mesh needs to be optimized through multiple iterations. The number of iterations for coarse mesh reconstruction is set to 200 steps, and the number of iterations for CVCR is set to 50 steps. This study focuses on optimizing the appearance and geometric quality of meshes. The enhanced multi-view color images are processed using the coloring algorithm from Unique3D [8]. For further details, refer to Unique3D [8]. For our designed mesh outlier metric, Outlier Risk Measures (ORM), we configure the confidence level $\xi = 95\%$ and set $\lambda = 10$.

2. Additional results

2.1. More visual comparison

We present more visual comparisons between our Dehallu3D and other 3D mesh reconstruction methods including SF3D [1], CRM [7], InstantMesh [5], TripoSR, and Wonder3D [4] in Figure 1. Due to the noticeable shortcomings of meshes generated by CRM, InstantMesh, TripoSR, and Wonder3D compared to Dehallu3D, we focus on comparing the differences between SF3D, Unique3D, and Dehallu3D to further highlight the effectiveness of Dehallu3D.

In the first row, it is evident that Unique3D exhibits two outliers, whereas Dehallu3D successfully eliminates these outliers. The quality of mesh generated by SF3D is obviously inferior to that of Dehallu3D and Unique3D. In the second row, the mesh generated by Unique3D also contains two outliers, the texture of the mesh generated by SF3D remains relatively blurry, and Dehallu3D generates a better mesh. In the third row, the mesh generated by Unique3D has one prominent outlier, the texture of the mesh generated by SF3D is also relatively blurry, while the mesh generated by Dehallu3D not only has better geometry but also exhibits clearer and more visible textures. In the fourth row, the mesh generated by Unique3D still contains outliers, and the mesh generated by SF3D shows a significant issue—in fact, SF3D fails to effectively generate a mesh corresponding to the input image. In contrast, Dehallu3D generates a relatively better mesh. In the fifth row, the mesh generated by Unique3D exhibits noticeable outliers, the texture of the mesh generated by SF3D is relatively blurry, and Dehallu3D generates a relatively better mesh. In the final row, the mesh generated by Unique3D has one prominent outlier, the texture of the mesh generated by SF3D is relatively blurry, while Dehallu3D effectively generates a better mesh.

Overall, although the textures of the meshes generated by Unique3D are relatively clear, they are prone to mesh outliers. The meshes generated by SF3D generally have blurry textures and may even fail to reconstruct effectively in some cases (e.g., in the fourth row). In contrast, Dehallu3D not only ensures clear textures and accurate geometry in the meshes but also effectively avoids potential mesh outliers during reconstruction, enabling the generation of high-fidelity and high-quality meshes.

2.2. More generated meshes

We show additional meshes generated by Dehallu3D in Figure 2. As demonstrated in the figure, our method consis-

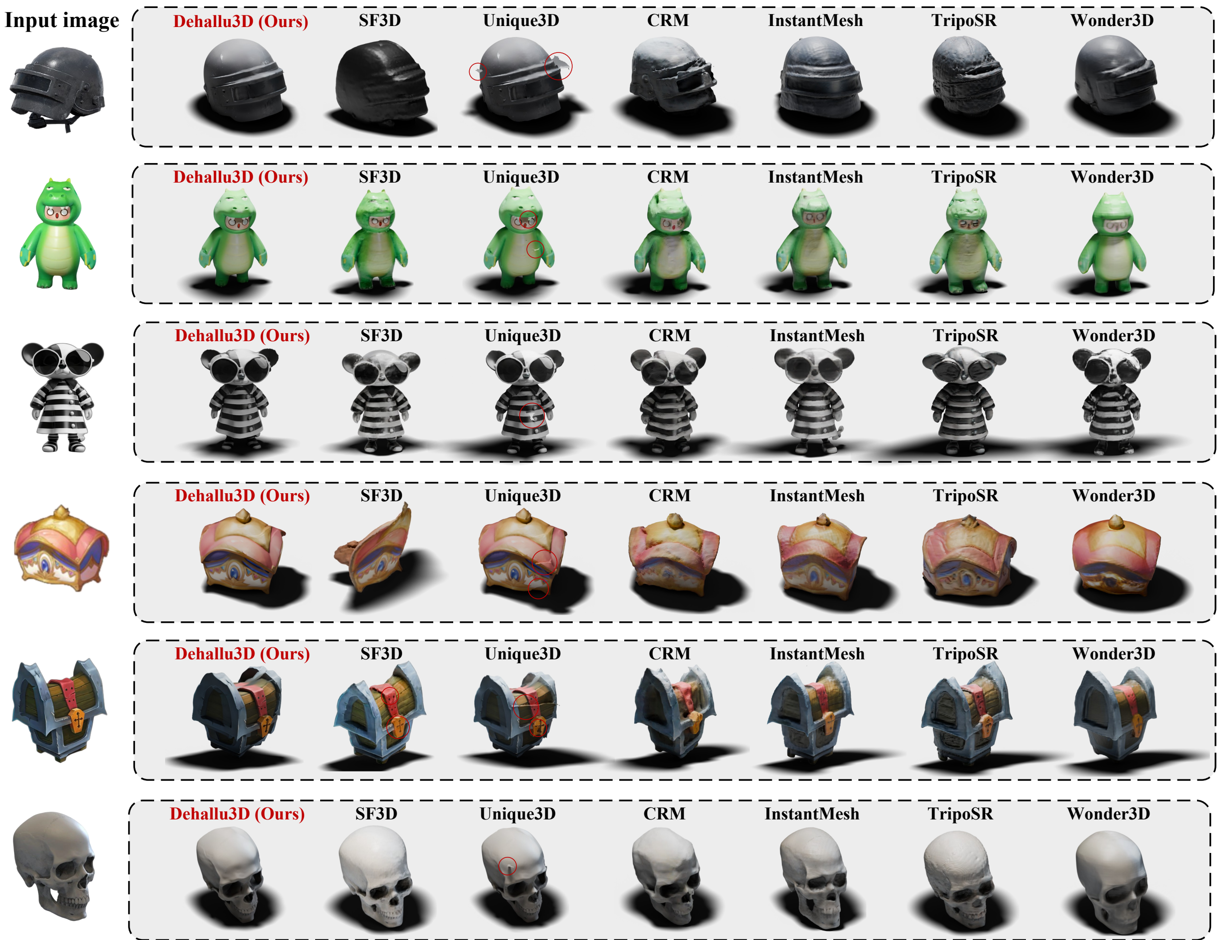

Figure 1. More qualitative comparison results.

Methods	View Alignment↑	Geometric Fidelity↑	Absence of Hallucinations↑	ORM↓
Wonder3D [4]	6.1	5.8	6.9	0.5452
TripoSR [6]	6.3	6.1	6.8	0.5519
InstantMesh [9]	6.9	6.6	7.0	0.5449
CRM [7]	7.1	6.5	6.8	0.5541
Unique3D [8]	7.3	7.0	5.6	0.7225
SF3D [1]	7.0	6.2	7.1	0.5410
Dehallu3D (Ours)	**8.4**	**8.0**	**7.2**	**0.5387**

Table 1. User study results compared with our proposed ORM metric. Human perceptual scores (scale: 1–10) are averaged from 30 participants on 10 representative samples.

tently ensures high-quality mesh generation and robustly avoids the introduction of geometric outliers or structural artifacts.

2.3. User study

2.3.1. Settings

To conduct user studies, we rendered 360-degree videos of 3D meshes generated from various popular single-view images across different methods. Ten representative samples were randomly selected and shown to each participant in the form of rotating turntable animations, allowing them to observe the reconstructed geometry and visual consistency comprehensively. A total of 30 volunteers participated in the study.

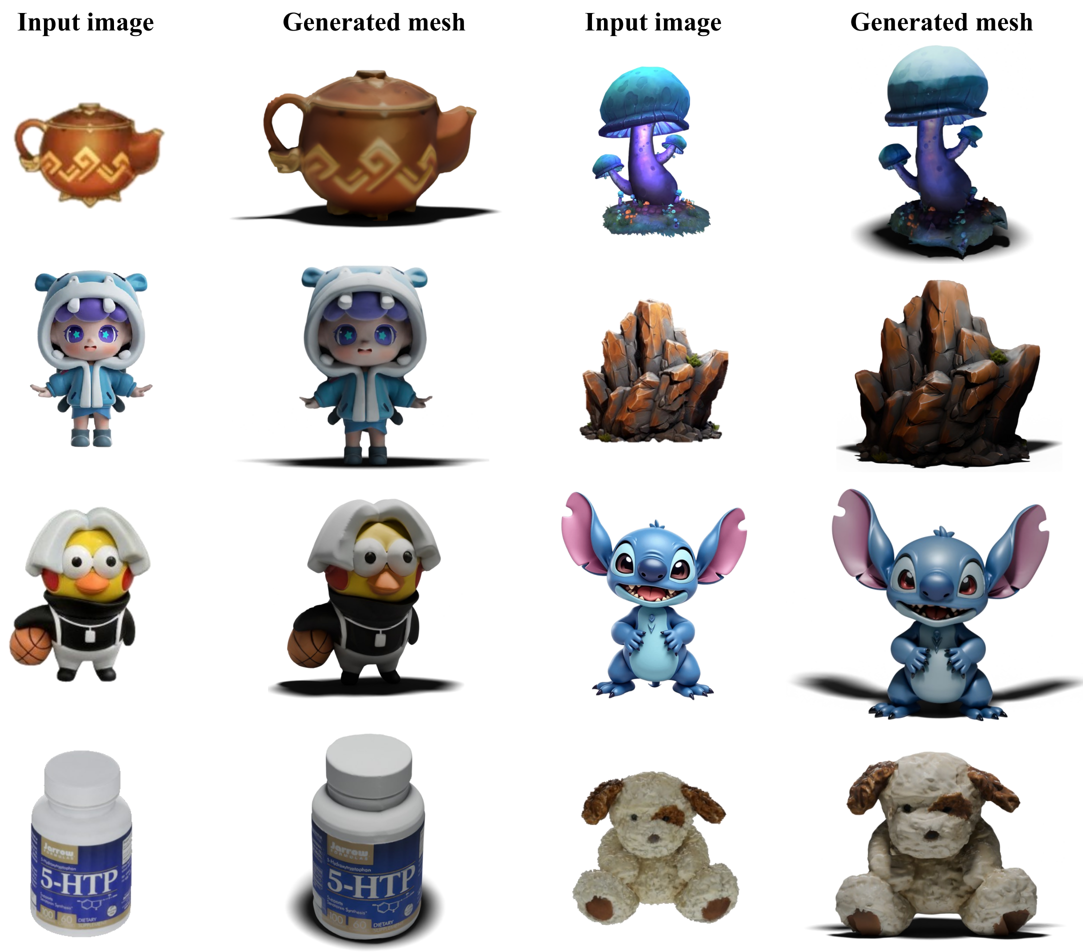

Figure 2. More meshes generated by Dehallu3D.

2.3.2. Perceptual metrics

For each sample, they were asked to rate the generated meshes from all methods under three perceptual dimensions on a 1–10 scale (higher is better):

- **View Alignment**: Measures cross-view structural consistency, indicating whether different views of the same object align properly without distortions or mismatches.
- **Geometric Fidelity**: Evaluates the accuracy and completeness of the 3D shape, especially for preserving fine details and clean surfaces.
- **Absence of Hallucinations**: Assesses the structural integrity of the mesh, specifically focusing on the lack of geometric outliers, artifacts, or structural flaws (e.g., odd holes, unnatural protrusions) that deviate from a plausible object shape.

2.3.3. Results

The evaluation results are summarized in Table 1. Our method achieves the highest ratings across all four criteria, showing significant improvements in both geometric and perceptual quality over compared methods. Notably, participants consistently favored our results for their cross-view smoothness, sharper geometry, and better semantic alignment.

Furthermore, we compare these subjective scores to the objective ORM metric (lower is better), presented in the final column. Our method achieves the lowest (best) ORM value. This objective ranking shows a strong correlation with the human perceptual scores, particularly with our top rank in the **Absence of Hallucinations** metric. This consistency robustly validates that our proposed ORM metric effectively captures human perception of geometric outliers.

2.4. Ethical Considerations and Broader Impact

The advancement of high-fidelity 3D generation, as demonstrated by Dehallu3D, provides significant value for digital content creation. However, it also necessitates a critical reflection on data privacy and security. Since our method relies on single-view images as input, the protection of source image integrity becomes vital to prevent potential misuse. Future research could explore image privacy protection [3] to strike a balance between high-quality 3D generation and robust privacy preservation. By addressing these security challenges, we aim to ensure that 3D generation technologies are developed and deployed in a responsible and secure manner.

References

[1] Mark Boss, Zixuan Huang, Aaryaman Vasishta, and Varun Jampani. Sf3d: Stable fast 3d mesh reconstruction with uv-unwrapping and illumination disentanglement. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 16240–16250, 2025. 1, 2

[2] Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B McHugh, and Vincent Vanhoucke. Google scanned objects: A high-quality dataset of 3d scanned household items. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2553–2560. Ieee, 2022. 1

[3] Xia Du, Jiajie Zhu, Jizhe Zhou, Chi-man Pun, Zheng Lin, Cong Wu, Zhe Chen, and Jun Luo. Dp-trae: A dual-phase merging transferable reversible adversarial example for image privacy protection. *IEEE Transactions on Dependable and Secure Computing*, 2025. 3

[4] Xiaoxiao Long, Yuan-Chen Guo, Cheng Lin, Yuan Liu, Zhiyang Dou, Lingjie Liu, Yuexin Ma, Song-Hai Zhang, Marc Habermann, Christian Theobalt, et al. Wonder3d: Single image to 3d using cross-domain diffusion. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9970–9980. IEEE, 2024. 1, 2

[5] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 1

[6] Dmitry Tochilkin, David Pankratz, Zexiang Liu, Zixuan Huang, Adam Letts, Yangguang Li, Ding Liang, Christian Laforte, Varun Jampani, and Yan-Pei Cao. Triposr: Fast 3d object reconstruction from a single image. *arXiv preprint arXiv:2403.02151*, 2024. 2

[7] Zhengyi Wang, Yikai Wang, Yifei Chen, Chendong Xiang, Shuo Chen, Dajiang Yu, Chongxuan Li, Hang Su, and Jun Zhu. Crm: Single image to 3d textured mesh with convolutional reconstruction model. In *European conference on computer vision*, pages 57–74. Springer, 2024. 1, 2

[8] Kailu Wu, Fangfu Liu, Zhihan Cai, Runjie Yan, Hanyang Wang, Yating Hu, Yueqi Duan, and Kaisheng Ma. Unique3d: High-quality and efficient 3d mesh generation from a single image. *Advances in Neural Information Processing Systems*, 37:125116–125141, 2024. 1, 2

[9] Jiale Xu, Weihao Cheng, Yiming Gao, Xintao Wang, Shenghua Gao, and Ying Shan. Instantmesh: Efficient 3d mesh generation from a single image with sparse-view large reconstruction models. *arXiv preprint arXiv:2404.07191*, 2024. 2

[10] Qiao Yu, Xianzhi Li, Yuan Tang, Xu Han, Long Hu, Yixue Hao, and Min Chen. Fancy123: One image to high-quality 3d mesh generation via plug-and-play deformation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 595–604, 2025. 1