Supplementary materials for "Fantasia3D: Disentangling Geometry and Appearance for High-quality Text-to-3D Content Creation"

Rui Chen*, Yongwei Chen*, Ningxin Jiao, Kui Jia†
South China University of Technology

In this section, we present the supplementary materials for Fantasia3D. We first show the additional generated results of our method in Section A. Subsequently, we provide more comparison results with DreamFusion [2] and Magic3D [1] in Section B. We finally demonstrate the additional implementation details in Section C.

A. Additional results of our method

In Fig. 1, we present more results of zero-shot generation produced by Fantasia3D. Besides the photorealistic appearance, physically based reflectance effects can also be observed in our results when changing the observation angle, which are illustrated in our attached video.

B. Additional comparison

More qualitative comparisons. We conduct the additional qualitative comparisons among DreamFusion [2], Magic3D [1], and Fantasia3D in Fig. 2. Comparing our method with Magic3D, we observe that our results are more photorealistic with competitive geometries. We consistently outperform DreamFusion in both appearance and geometry generation.

User studies. We randomly selected 148 users to evaluate 10 meshes with the textures generated by three different methods based on the same text prompt. Each user was asked to respond to three questions: (Q1)"How closely does the result match the text description?" (Q2) "How realistic is the result?" (Q3) "How detailed is the result?" Tab. 1 displays the mean scores and standard deviations for our method and two baselines, with the highest score intentionally highlighted in bold. Our proposed method, Fantasia3D, outperforms the two baselines in all three questions.

C. Additional implementation details

We select camera positions (r, θ, φ) in the spherical coordinate system, where r, θ, and φ denote the radius, elevation, and azimuth angle, respectively. To be specific, for a batch of $s \times l$ images, we partition the azimuth angle $\varphi \in [0, 2\pi]$ into l intervals of $[\frac{2d\pi}{l}, \frac{2(d+1)\pi}{l}]$ and uniformly sample s azimuth angles in each interval. This results in the constraint for the k-th viewpoint:

$$\begin{cases} r_k \in [r_{\min}, r_{\max}] \\ \theta_k \in [\theta_{\min}, \theta_{\max}] \\ \varphi_k \in \left[\frac{2d\pi}{l}, \frac{2(d+1)\pi}{l}\right) \end{cases}, \tag{1}$$

where $r_{\min}$, $r_{\max}$, $\theta_{\min}$, $\theta_{\max}$ are hyperparameters that decide the range of r and θ, and $d = k \bmod l$. In most cases, we set $r_{\min} = r_{\max} = 3$, $\theta_{\min} = -\frac{\pi}{18}$ and $\theta_{\max} = \frac{\pi}{4}$, and sample the field of view from $\mathcal{U}(\frac{\pi}{7}, \frac{\pi}{4})$.

For the DMTET initialization, we utilize the Open3D library [3] to calculate the signed distance function (SDF) for each point in close proximity to the mesh surface. In our experiment setup, the initialization procedure is not necessarily required to achieve convergence and can be terminated prematurely. The initialization process, which includes sampling the SDF and performing a coarse fitting, typically takes approximately a minute and a half.

For the score distillation sampling, we sample $t \sim \mathcal{U}(0.02, 0.5)$ and set $w(t) = \sigma^2\sqrt{(1-\sigma^2)}$.

*Equal contribution.
†Corresponding author.

Table 1. Mean opinion scores (1∼ 5) for Q1∼Q3

	(Q1)	(Q2)	(Q3)
DreamFusion [2]	3.89 (± 0.82)	3.77 (± 0.79)	3.69 (± 0.75)
Magic3D [1]	3.88 (± 0.81)	3.78 (± 0.82)	3.96 (± 0.76)
Ours	**4.03** (± 0.79)	**4.38** (± 0.77)	**4.44** (± 0.72)

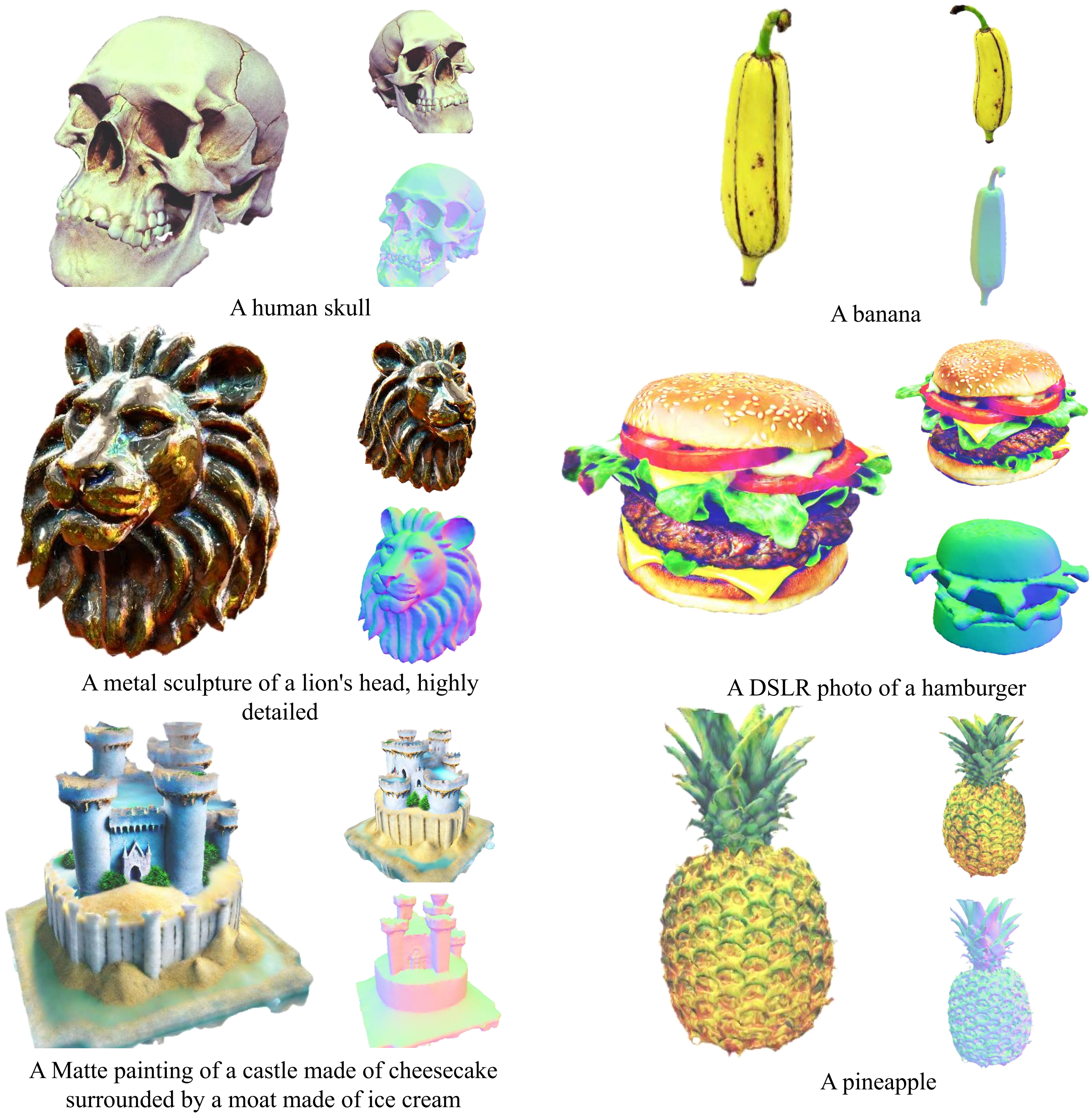

Figure 1. **Results of our method.** This figure showcases the generation results obtained from solely text prompts.

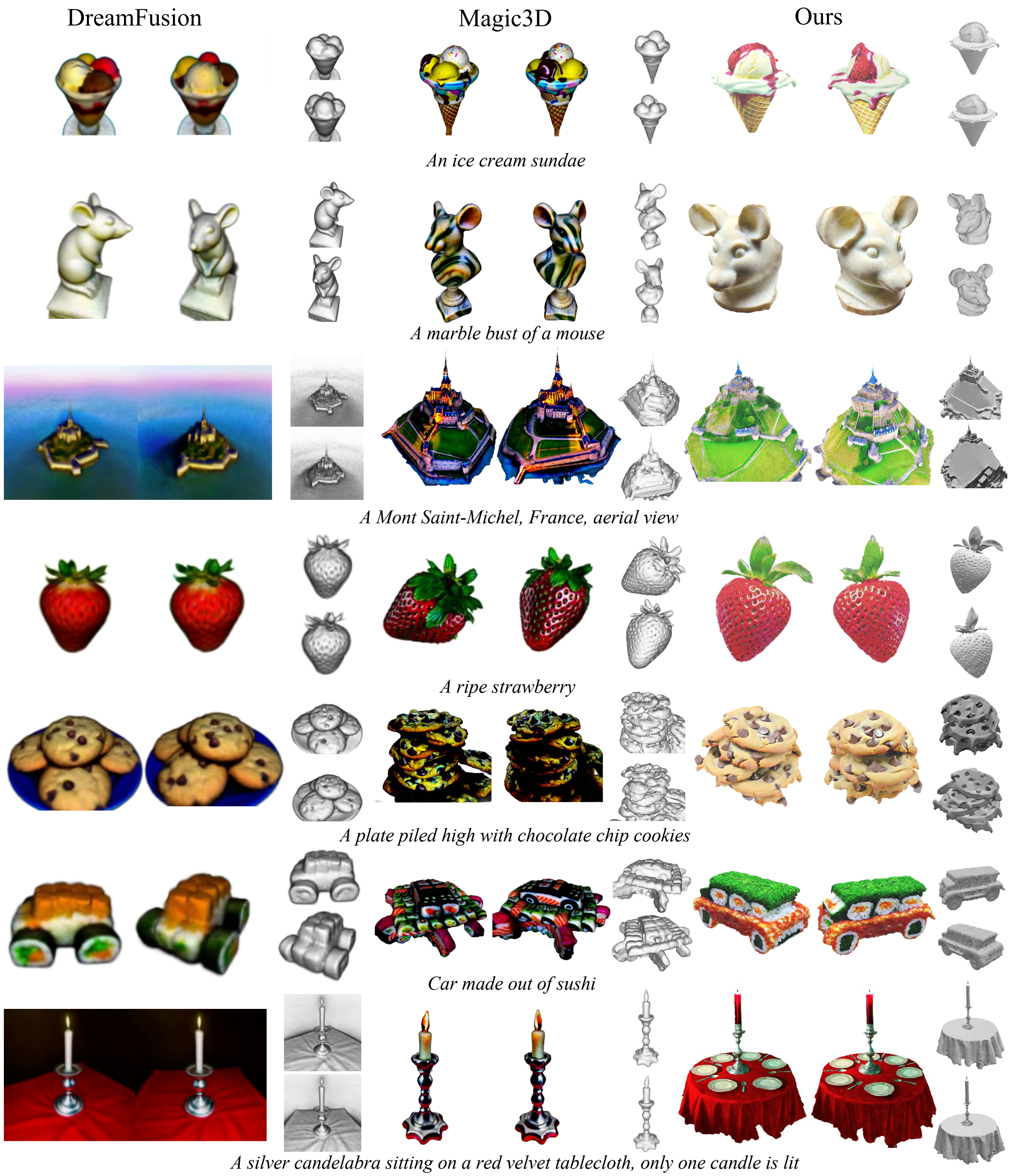

Figure 2. **Comparison of zero-shot generation.** Since DreamFusion and Magic3D do not have released codes, their results are obtained by downloading from their project pages. Our results are more photorealistic with competitive geometries compared with Magic3d. We consistently outperform DreamFusion in both appearance and geometry generation.

References

[1] Chen-Hsuan Lin, Jun Gao, Luming Tang, Towaki Takikawa, Xiaohui Zeng, Xun Huang, Karsten Kreis, Sanja Fidler, Ming-Yu Liu, and Tsung-Yi Lin. Magic3D: High-resolution text-to-3D content creation. *arXiv preprint arXiv:2211.10440*, 2022.

[2] Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. *International Conference on Learning Representations (ICLR)*, 2023.

[3] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv preprint arXiv:1801.09847*, 2018.