Supplementary material for "HD-Fusion: Detailed Text-to-3D Generation Leveraging Multiple Noise Estimation"

Jinbo Wu, Xiaobo Gao, Xing Liu, Zhengyang Shen, Chen Zhao,
Haocheng Feng, Jingtuo Liu, Errui Ding
Department of Computer Vision Technology (VIS), Baidu Inc., China
{wujinbo01,gaoxiaobo,liuxing12,shenzhengyang01,zhaochen03,
fenghaocheng,liujingtuo,dingerrui}@baidu.com

This document provides additional explanations about the proposed method and additional experimental results.

A. Global and local consistency

We explain why a full-object prompt (e.g., Messi, full body) does not negatively affect local content (e.g., hands) generation when the proposed multiple noise estimation is applied. We found that pretrained Stable Diffusion automatically addresses this issue after the Color network reaches a certain stage of training. We therefore first train the network to a certain stage without the proposed multiple noise estimation, and then we activate it for continued training until completion in Stage 2 - Phase 2.

B. Computational cost with high-resolution renderings

We report the training times for each stage in Table 1. It is observed that the proposed multiple noise estimation mechanism yields a significant improvement (see Fig. 6 in the main text) in performance at an acceptable additional cost of 21 minutes. We compare the training time of our full pipeline with the SOTAs' in Table 2. Ours is faster than Magic3D's and is competitive with Fantasia3D's.

Table 1. Training on single GPU (A100). MNE denotes the proposed multiple noise estimation.

	Stage1	Stage2 w/ MNE	Stage2 w/o MNE
training time	$\sim$ 40 min	$\sim$ 65 min	$\sim$ 44 min

Table 2. Training time comparison on single GPU (A100).

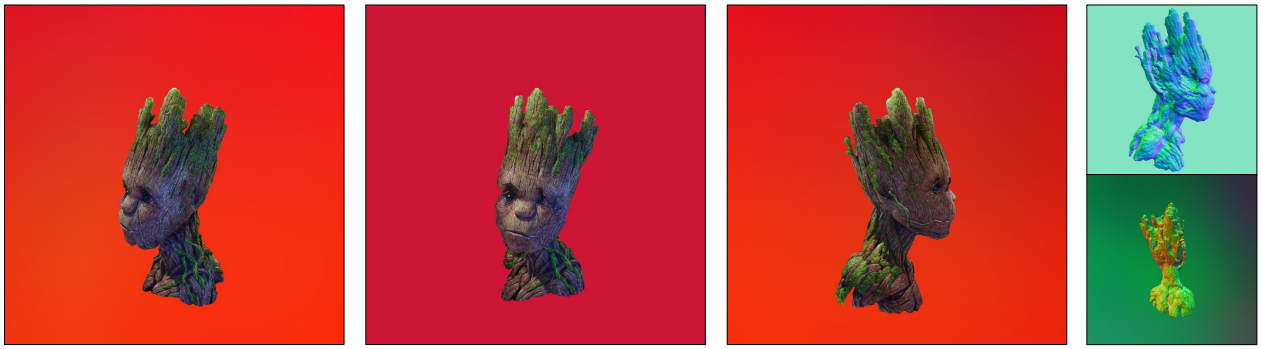
	Ours	Fantasia3D	Magic3D
training time	$\sim$ 1.75 h	$\sim$ 2 h	$\sim$ 5 h

C. More visual results

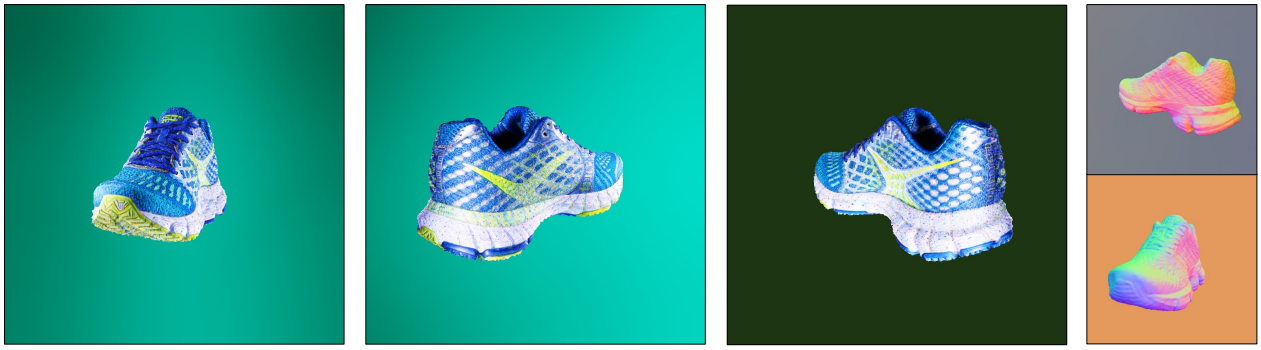
We provide more results in Figs. 1, 2, 3 on the following pages.

Prompt: A butterfly over a tree stump

Prompt: A snail on a leaf



Prompt: A head of I am groot



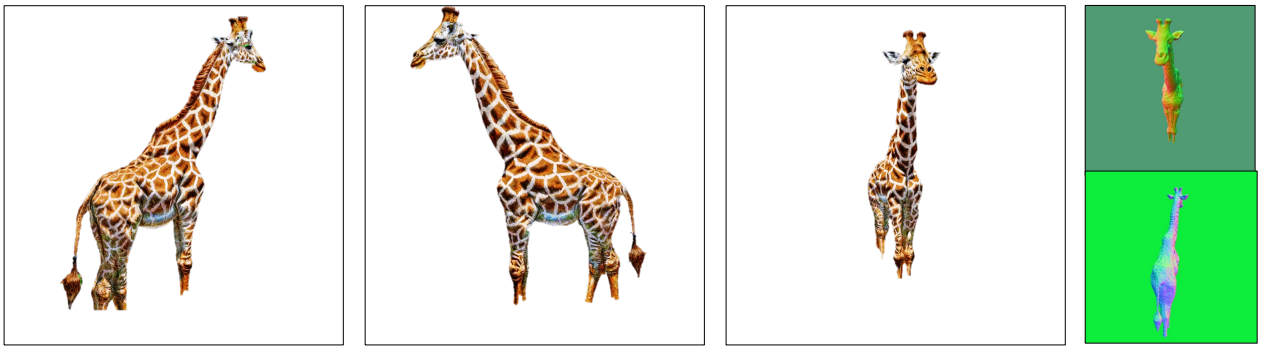
Prompt: A sport shoe

Figure 1. More results. We show them in multiple figures to see them clearly.

Prompt: An Obama figure, full body

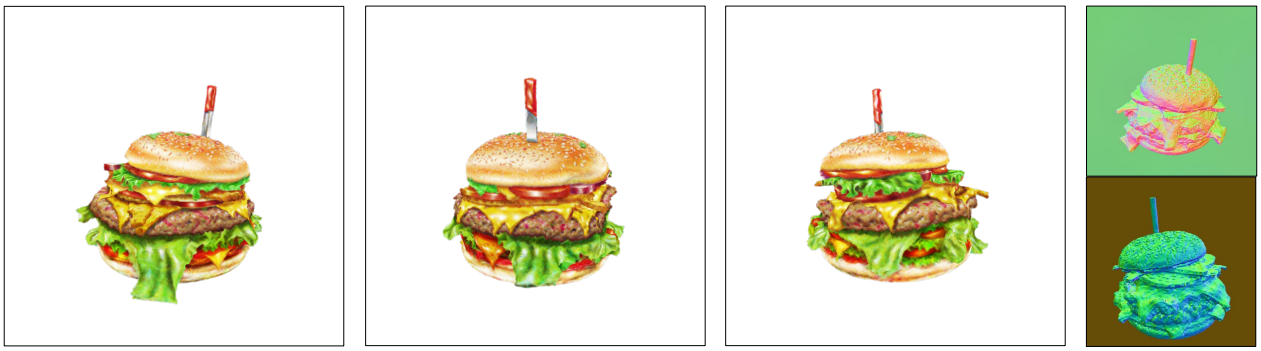
Prompt: Wuddy in toy story, full body

Prompt: Luffy, full body



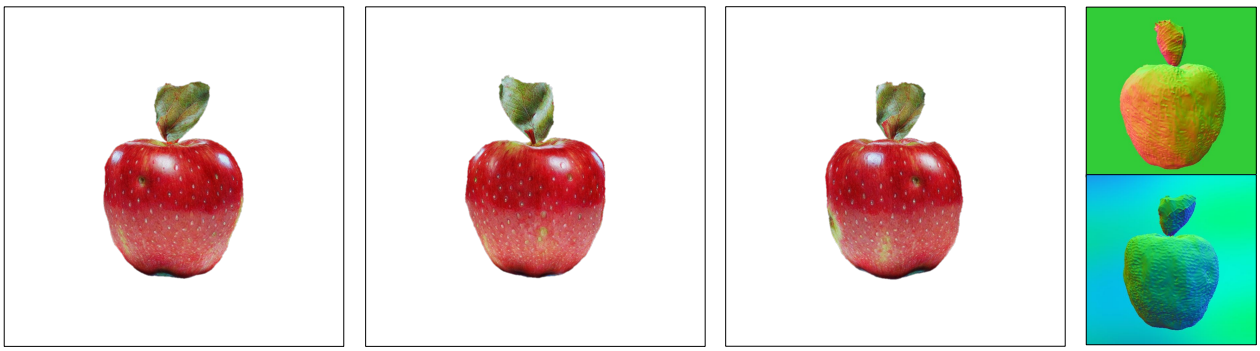
Prompt: A giraffe, full body

Figure 2. More results. We show them in multiple figures to see them clearly.



Prompt: A hamburger

Prompt: Bagel filled with cream cheese and lox



Prompt: A red apple

Prompt: An ice cream sundae

Figure 3. More results. We show them in multiple figures to see them clearly.