RichDreamer: A Generalizable Normal-Depth Diffusion Model for Detail Richness in Text-to-3D

Lingteng Qiu[1,3*], Guanying Chen[2,1*], Xiaodong Gu[3*],
Qi Zuo[3], Mutian Xu[1], Yushuang Wu[2,1,3], Weihao Yuan[3], Zilong Dong[3], Liefeng Bo[3], Xiaoguang Han[1,2]

[1]SSE, CUHKSZ   [2]FNii, CUHKSZ   [3]Alibaba Group

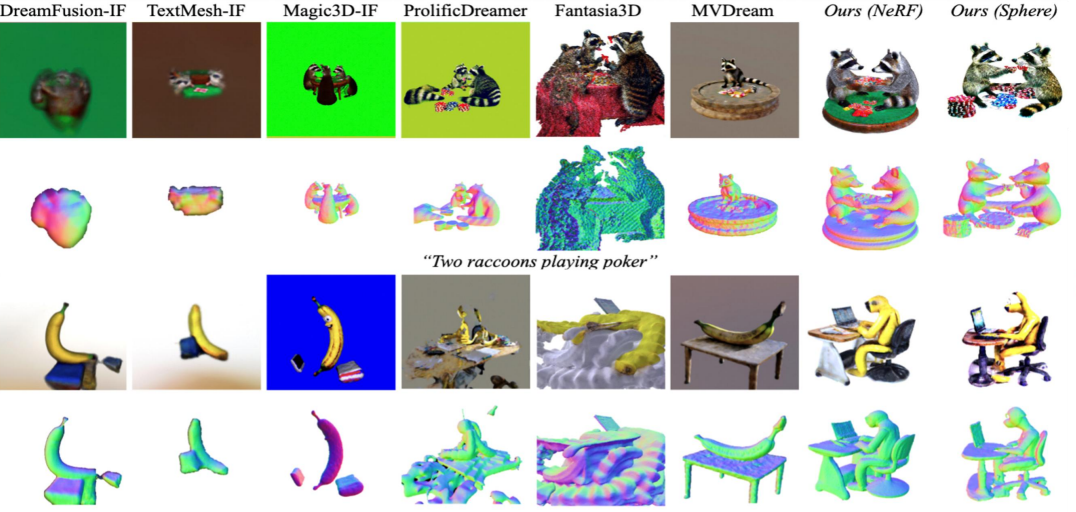CVPR JUNE 17-21, 2024 SEATTLE, WA Highlight Webpage
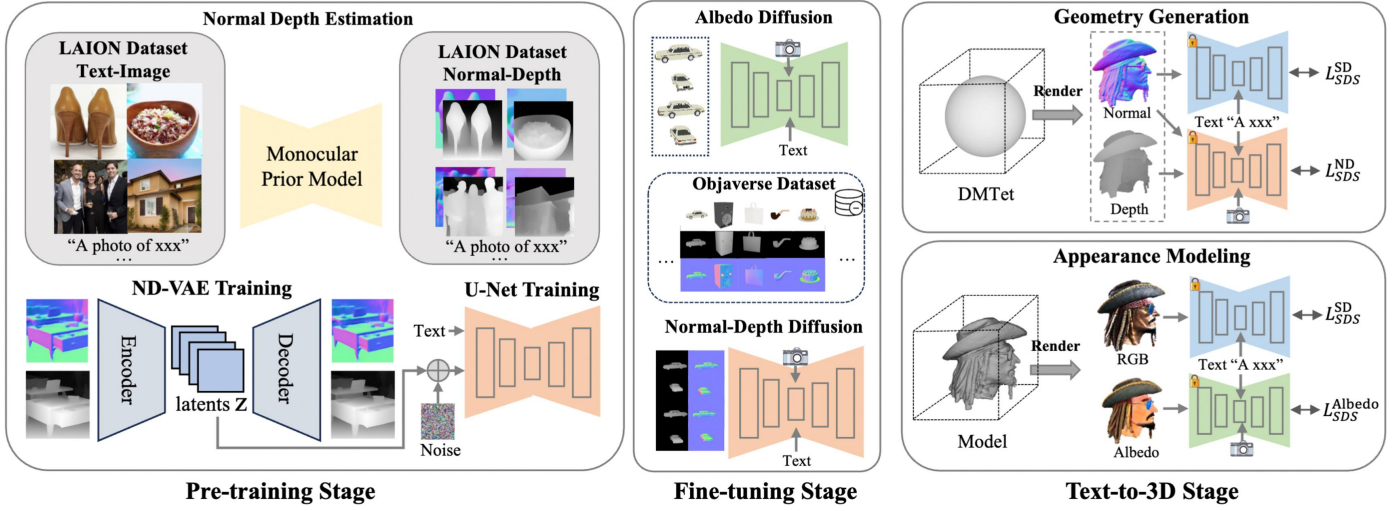
Introduction

To bridge the gap between 2D diffusion and 3D generation posed by geometric and illumination complexities, our paper introduces a **Normal-Depth** diffusion model, trained with LAION dataset together with the generalizable image-to-depth and normal priors, which surpasses current methods with its robustness to lighting by leveraging an **Albedo** diffusion model and achieves richness in detail.
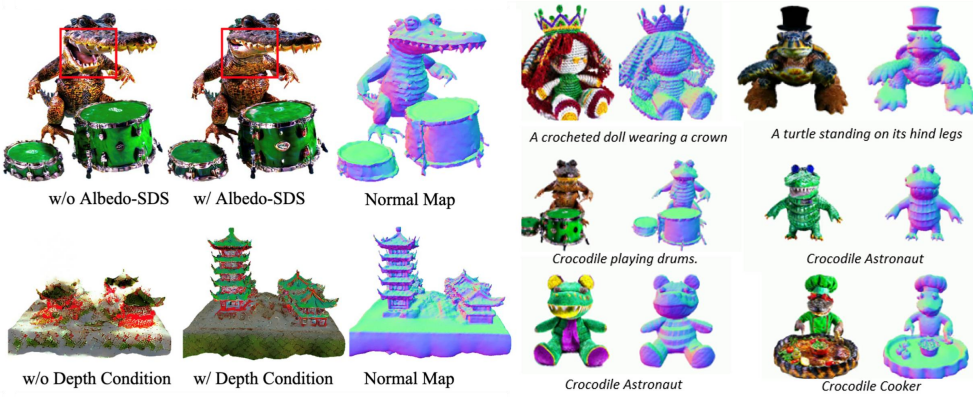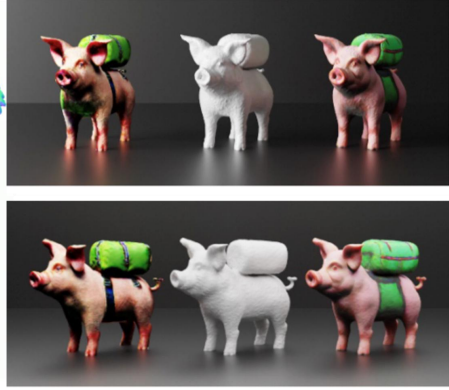
Methodology

Overview of the proposed **RichDreamer**. We introduce a generalizable **Normal-Depth** diffusion model that is trained on the **LAION-2B** dataset with normal and depth predicted by Midas and Normal-Bae , followed by fine-tuning on the synthetic dataset. Our model can be incorporated with the DMTet and NeRF representations to enhance the geometry generation. To alleviate the ambiguity in appearance modeling, we propose an **albedo** diffusion model to impose data-drive prior on the albedo component.

Experiments

"Two raccoons playing poker"

"A wide angle DSLR photo of a humanoid banana sitting at a desk doing homework"

Comparison results Between ours and existing methods

Ablation results for the albedo diffusion model.

Qualitative results demonstrate that our method can generate realistic object.

Relighting results. From top to bottom: w/o albedo diffusion and w/ albedo diffusion.