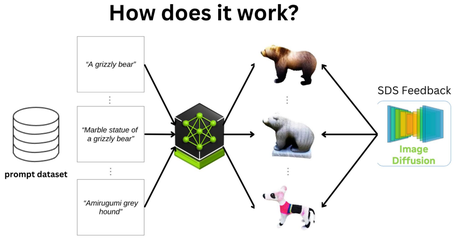
LATTE3D: Large-scale Amortized Text-To-Enhanced3D Synthesis

Kevin Xie, Jonathan Lorraine, Tianshi Cao, Jun Gao,
James Lucas, Antonio Torralba, Sanja Fidler, Xiaohui Zeng

Introduction

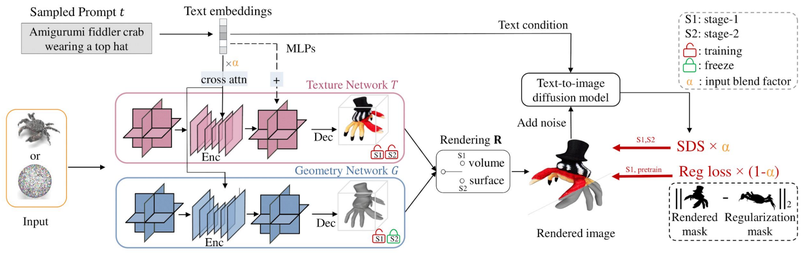
- We generate high-quality textured meshes from text robustly in **400ms** by combining:
  - Amortized optimization [1] for speed,
  - A surface rendering stage [2] for quality,
  - and various 3D priors for robustness

- TLDR: We train a network on a large set of text prompts to minimize an augmented SDS objective

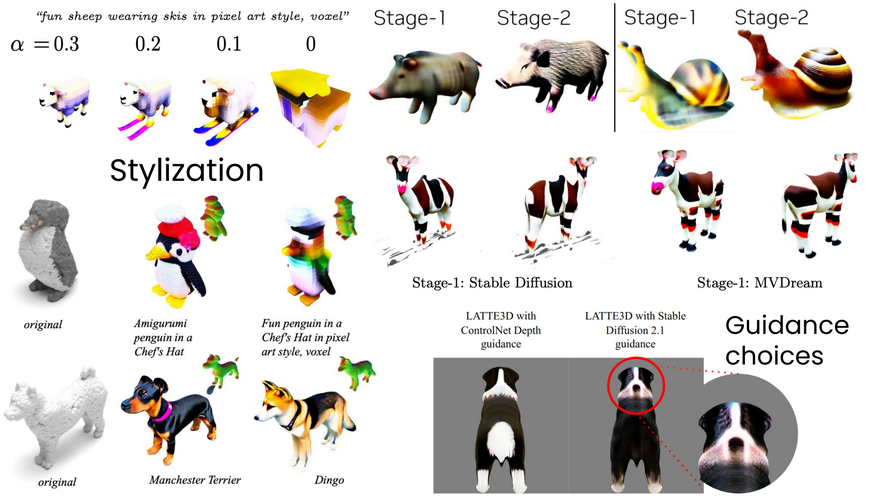


Our Method - Overview

- First, train texture+geometry with volumetric
- Second, train texture only with surface rendering
- We enhance robustness with 3D priors by:
  - Using 3D aware guidance, like MVDream [3]
  - Penalizing difference from a source 3D object
    - Also allows stylization



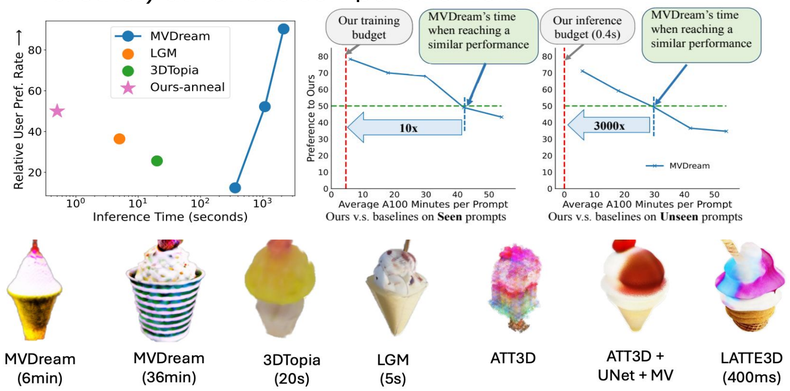
Our Method - More details



Experimental Results

Experimental Results

- We assess performance with user-preference rate (ours vs. other) at various compute horizons



- Ablations and test-time fine-tuning results in paper

Our View of the Future

- Limitations to improve upon:
  - Enhance generalization – ex., training on more prompts
  - Reduce artifacts (Janus, baked light) – ex., other guidances
- Use case: Enable agentic VLMs to procedurally build complex environments wielding fast text-to-3D tools

Links

- Webpage: research.nvidia.com/labs/toronto-ai/LATTE3D/
- [1] Lorraine, Jonathan, et al. "ATT3D: Amortized Text-To-3D object synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.
- [2] Lin, Chen-Hsuan, et al. "Magic3D: High-resolution text-to-3D content creation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
- [3] Shi, Yichun, et al. "MVDream: Multi-view Diffusion for 3D Generation." The Twelfth International Conference on Learning Representations.