Gaussians-to-Life: Text-Driven Animation of 3D Gaussian Splatting Scenes

Thomas Wimmer[1,2] Michael Oechsle[3] Michael Niemeyer[3] Federico Tombari[1,3]

[1] [2] [3]

Paper Website

3DV 2025

Motivation

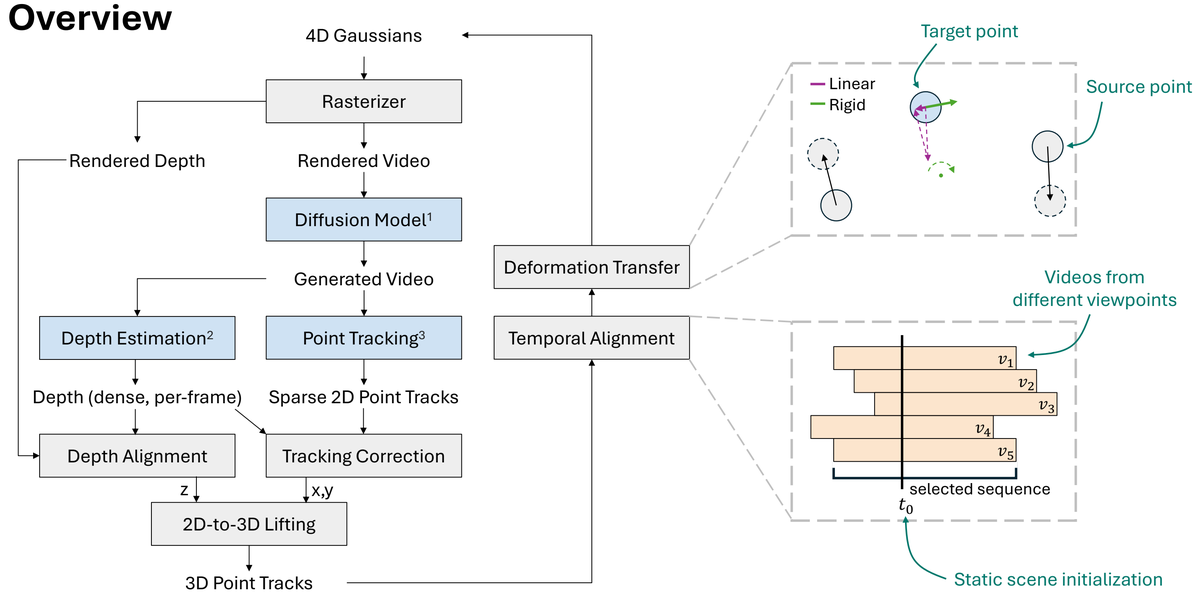
- Text-to-4D and video diffusion models **do not provide sufficient control** for real-world applications in AR/VR or film-making.
- Instead, **animate given 3D scenes** using video diffusion models as guidance!

- Make use of several pre-trained 2D models to lift motion into 3D without training or expensive runtime optimization as in Score Distillation Sampling.

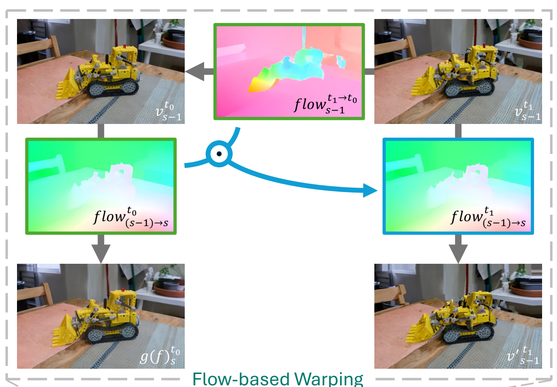
→ Animate scenes in **under 10 minutes** on a **24GB** GPU!

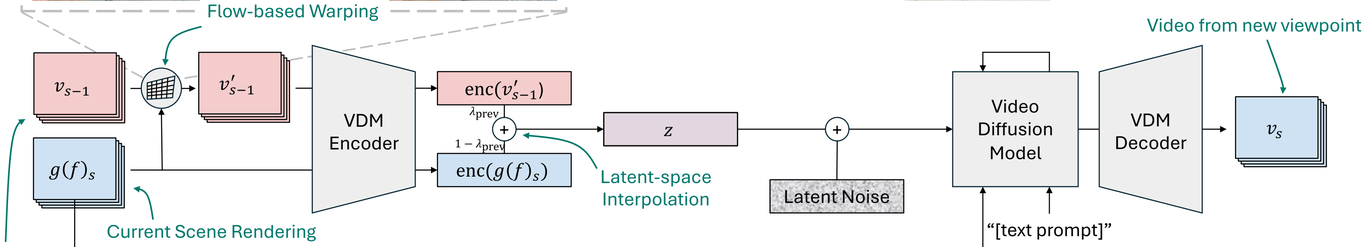
Results

Method Overview



Improving 3D Consistency of Generated Videos





References:
[1] Xing, Jinbo, et al. "Dynamicrafter: Animating open-domain images with video diffusion priors." *ECCV*, 2024.
[2] Piccinelli, Luigi, et al. "UniDepth: Universal monocular metric depth estimation." *CVPR*, 2024.
[3] Karaev, Nikita, et al. "Cotracker: It is better to track together." *ECCV*, 2024.

Acknowledgments: Thomas Wimmer is partially supported by the Konrad Zuse School of Excellence in Learning and Intelligent Systems (ELIZA) through the DAAD program Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the German Federal Ministry of Education and Research.