Sym-to-real Multi-view Synthesis

Overview

Calibrated multi-view data is difficult to come by in the real world. Yet, such data is essential for 3D reconstruction, which is foundational to many downstream computer vision tasks. One plausible approach is to leverage synthetic data, where generating multiple consistent views is cheap. However, it is unclear how would the models trained on synthetic data generalise to real-world data. It is the goal of this project to explore these limitations.

In this project, we will learn a model which synthesises a novel camera view conditioned on a single view from the user input. To train such a model, we will leverage synthetic data [5].

There are two aspects important to the implementation of this task. The first one is a shared 3D representation between the user-provided input view and the generated novel view. We will leverage Gaussian splatting [1], owing to its efficiency and differentiability. The second aspect concerns the sim-to-real gap: a model trained on synthetic data may not generalise well to real data. Using a large diffusion model as the image prior [4], we will experiment with a test-time refinement process to adapt the underlying 3D representation to realistic data (e.g. using score distillation sampling [3]).

Goals

The outcome of this project is a model which can synthesise a realistic novel view from a single input image. Fig. 1 presents a conceptual overview of the framework for this project.

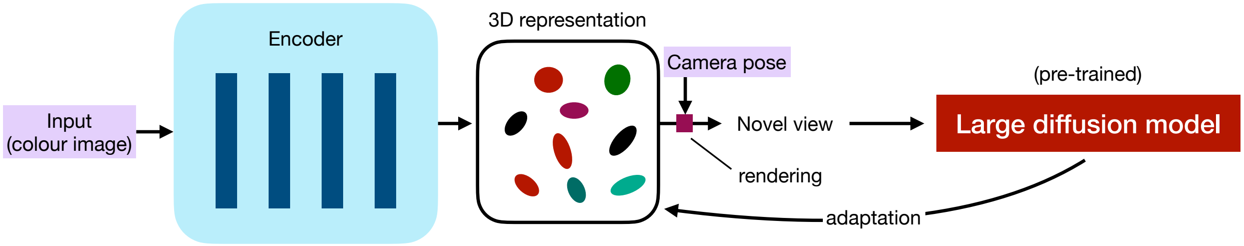

Figure 1: The goal of the project is to leverage synthetic data to learn an encoder network creating a differentiable 3D representation (Gaussian splatting) from a single image. Using a pre-trained diffusion model, we will adapt this model to real-world data.

Tasks

1 Getting familiar with the baseline[1] (~ 30h);

2 Prototyping the encoder network (~ 30h);

3 Setting up the dataset and model training; concept refinement (~ 120h);

4 Adapting the model to real-world data with diffusion models (~ 60h);

5 Writing a project report and preparing a presentation (~ 60h).

Contact

Nikita Araslanov
Email: `nikita.araslanov@tum.de`

References

[1] Bernhard Kerbl et al. “3D Gaussian Splatting for Real-Time Radiance Field Rendering”. In: *SIGGRAPH* (2023).

[2] Jonathon Luiten et al. “Dynamic 3D Gaussians: Tracking by persistent dynamic view synthesis”. In: *3DV*. 2023.

[3] Ben Poole et al. “DreamFusion: Text-to-3d using 2D diffusion”. In: *arXiv:2209.14988 [cs.CV]* (2022).

[4] Robin Rombach et al. “High-resolution image synthesis with latent diffusion models”. In: *CVPR*. 2022.

[5] Hao Yang et al. “ContraNeRF: Generalizable neural radiance fields for synthetic-to-real novel view synthesis via contrastive learning”. In: *CVPR*. 2023.

[6] Xiaoyu Zhou et al. “DrivingGaussian: Composite gaussian splatting for surrounding dynamic autonomous driving scenes”. In: *arXiv:2312.07920 [cs.CV]* (2023).

---

[1] https://github.com/ContraNeRF/ContraNeRF