Using Generative Priors for 3D Scene Reconstruction

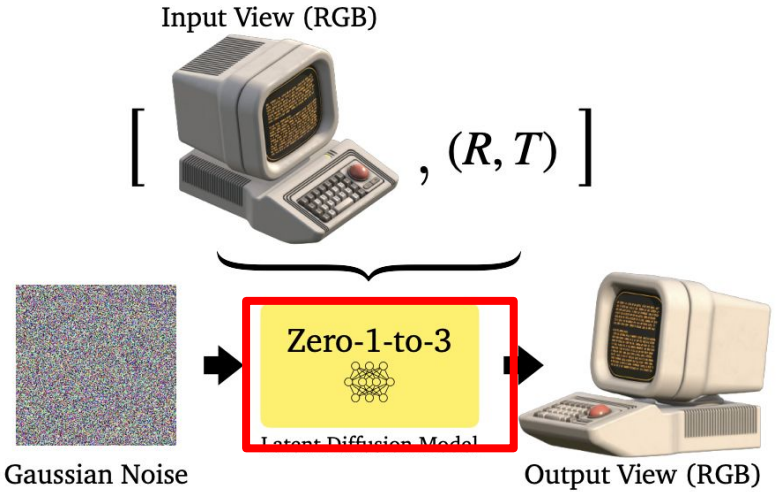
Alex Lyons

Motivation

- Need for 3D maps when exploring for planning
- Ideally only need 1 view
- Many modern methods fail with little data (especially 1 view)
  - NeRF
  - Gaussian Splatting
- How to overcome having few viewpoints?

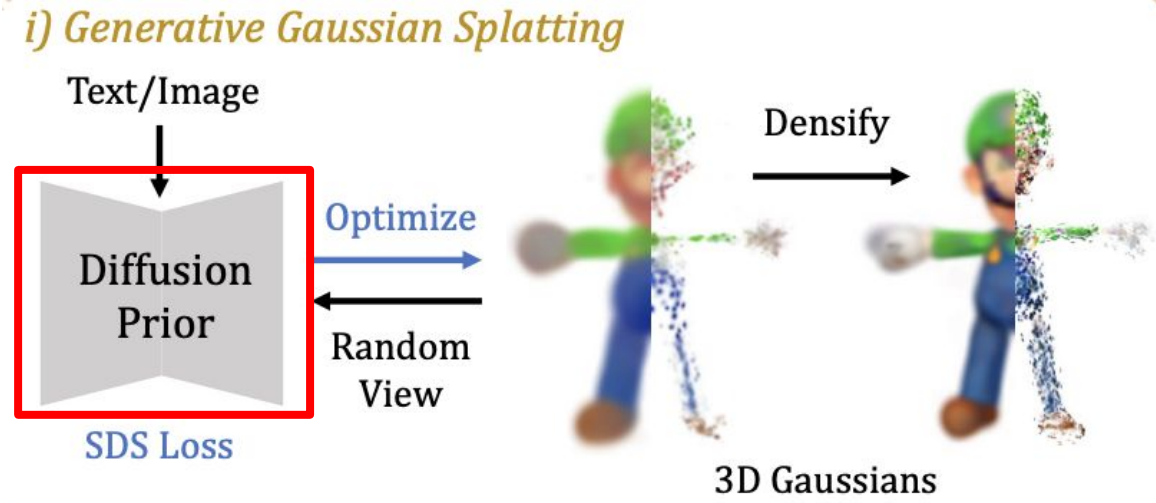
Zero-Shot Generation: Zero123

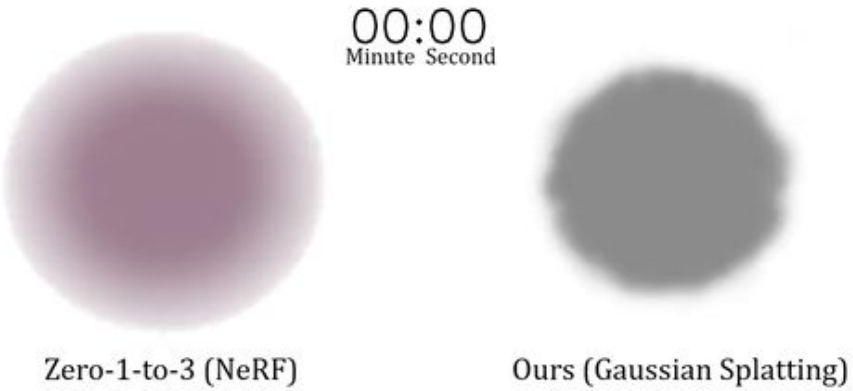
- Given image and transformation, generate new view
- Uses Stable Diffusion priors



DreamGaussian

- Gaussian splatting with generated viewpoints
- Use Zero123 as a prior
- Gaussian splatting greatly improves speed over NeRF





My Method

1. Generate objects using DreamGaussian
2. Combine the objects into one scene

Object Generation Pipeline

Input Scene

SceneNet synthetic dataset

Object Selection & Segmentation

Using SceneNet segmentations, selected by size

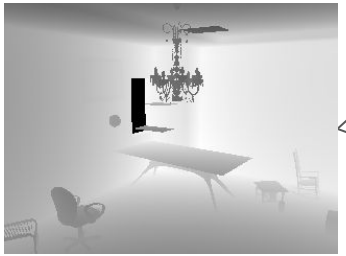
Input Cropped Objects

DreamGaussian

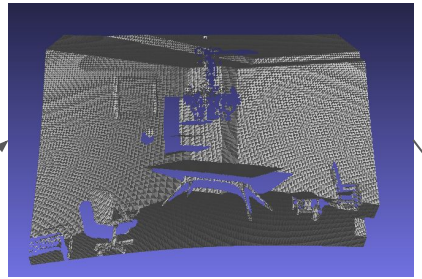
Google Colab

Outputted Meshes

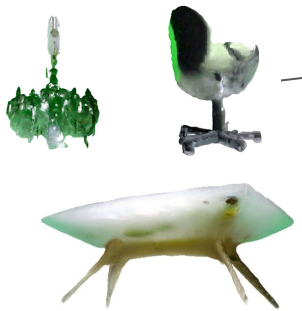
Scene Reconstruction Pipeline



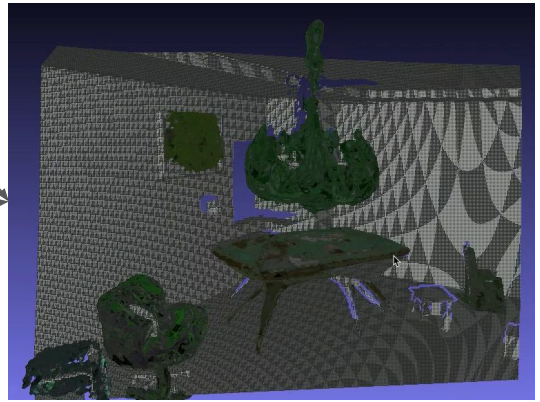
Depth map
from dataset



Construct point cloud
with intrinsics

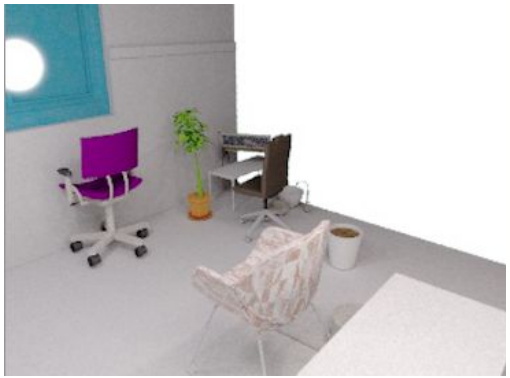


Estimate object size &
location with bounding boxes



Construct final scene
with new objects

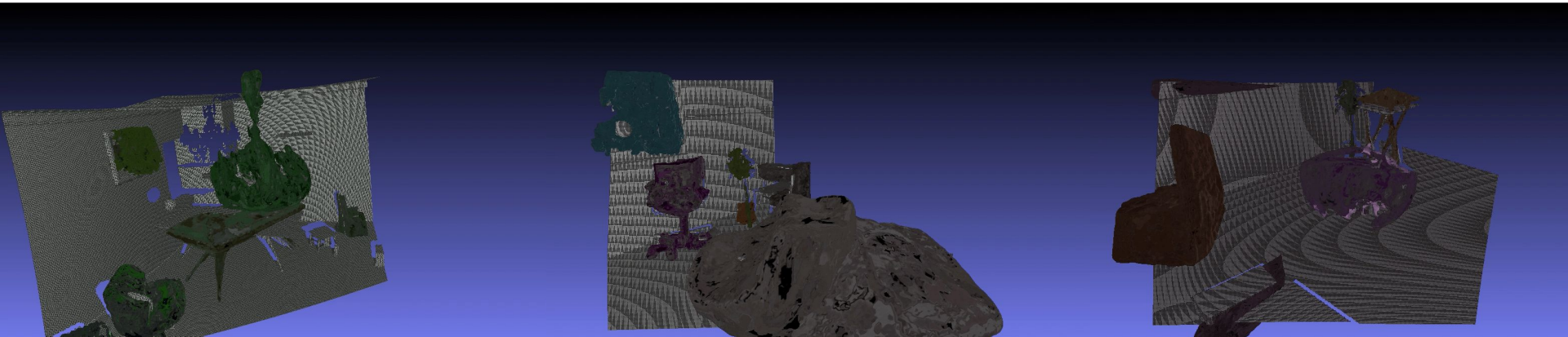
Qualitative Results: Object Reconstruction



Quantitative Results and Failures: Object Reconstruction

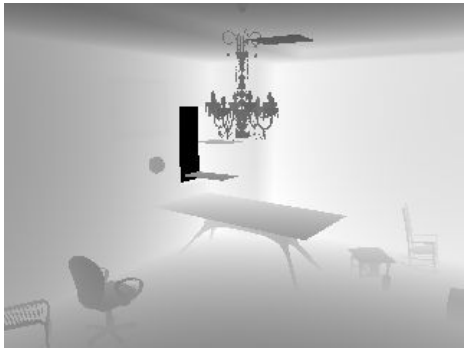
- Mean CLIP Similarity Scores (Closer to 1 is better):
  - Comparing original segmentation against:
    - Same-view render: 0.8451
    - Left view render (novel): 0.7435
    - Back view render (novel): 0.8128
    - Right view render (novel): 0.7841
- Failures:
  - Segmentations very out of distribution of DreamGaussian training data
    - Pixelated
    - Unusual angles

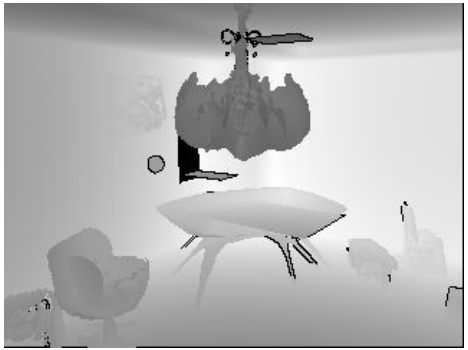
Qualitative Results: Scene Reconstruction



Quantitative Results and Failures: Scene Reconstruction

- Average (across scenes) mean square error between depth map and re-rendered depth map: 940





- Failures:
  - Distortions in point cloud
    - Even after using intrinsics, surfaces curved
  - Object scaling and placement
    - Occlusions added error to depth placement
    - Point cloud distortions added error to calculated object scaling factor

Future Work/Addressing Failures

- Object Generation:
  - Fine-tune DreamGaussian on segmentations
  - Use pre-selected classes for chosen objects
- Scene reconstruction (More difficult):
  - Further investigation on depth distortion (Most difficult challenge)
    - Manually un-distort
  - Use object point clouds for scaling generated objects
- General:
  - Test on more scenes

References

Jain, A., Tancik, M., & Abbeel, P. (2021). Putting nerf on a diet: Semantically consistent few-shot view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 5885-5894).

Liu, R., Wu, R., Van Hoorick, B., Tokmakov, P., Zakharov, S., & Vondrick, C. (2023). Zero-1-to-3: Zero-shot one image to 3d object. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9298-9309).

Tang, J., Ren, J., Zhou, H., Liu, Z., & Zeng, G. (2023). Dreamgaussian: Generative gaussian splatting for efficient 3d content creation. *arXiv preprint arXiv:2309.16653*.

McCormac, J., Handa, A., Leutenegger, S., & Davison, A. J. (2016). Scenenet rgb-d: 5m photorealistic images of synthetic indoor trajectories with ground truth. *arXiv preprint arXiv:1612.05079*.

Questions?