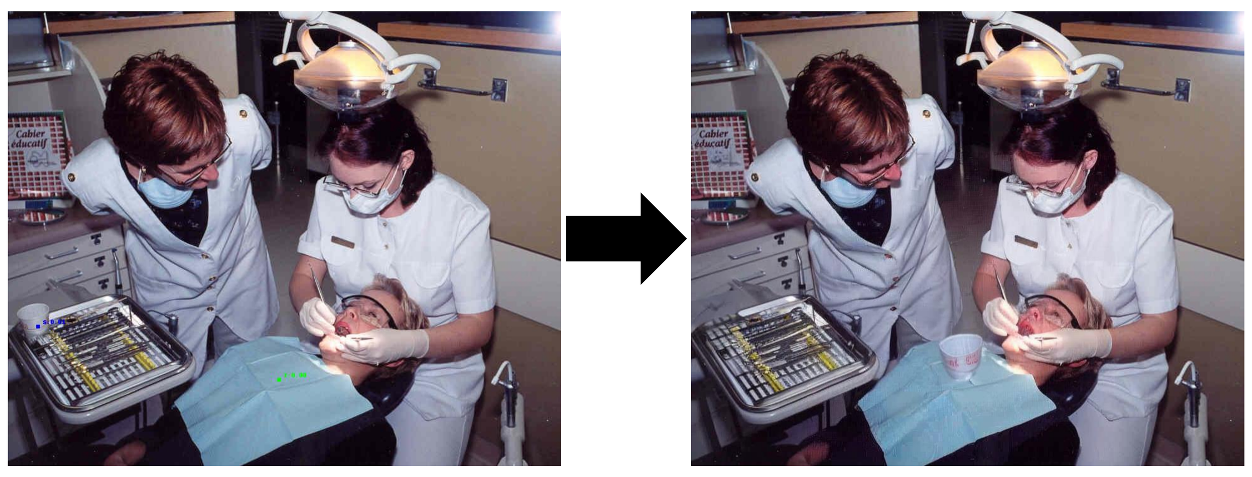
Depth Scaling for Diffusion Edits

Muhaimen Shamsi, Jadyn Wall
University of Toronto

Motivation

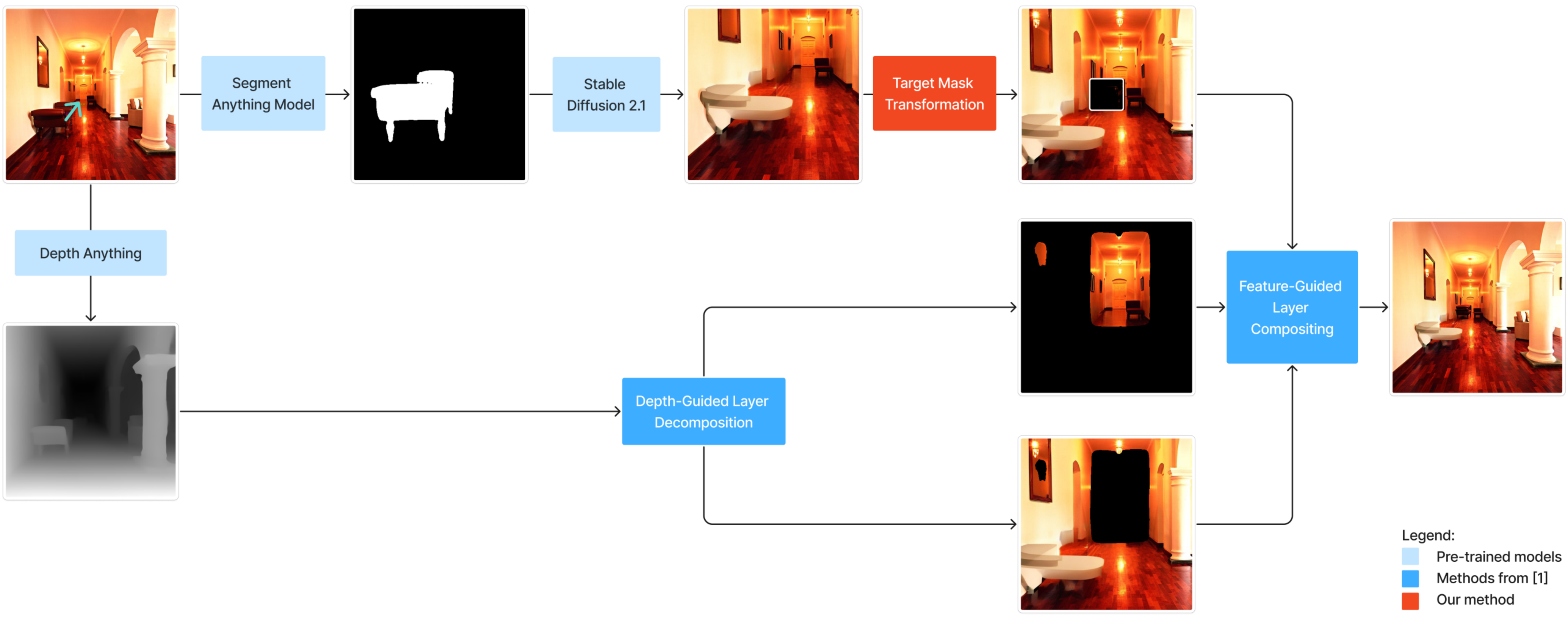
- Recent research in diffusion models demonstrates increasing control over object transformations such as rotation, scaling, and translation.
- We propose an addition to the existing Zero-Shot Depth-Aware Image Editing pipeline [1] that preserves the size of the object according to the difference in depth between the source and target points.



- Applications include data augmentation for robotic simulations and virtual product previews.

New Technique

- Given an object initially observed at depth z that we want re-positioned to a target depth z', we compute the physically motivated scale factor $s = z/z'$ and apply it uniformly to the cropped object patch and the ControlNet hint tensor.



$$\text{Object scale factor} = \frac{\text{Source object depth}}{\text{Target object depth}}$$

Related Work

- The limitation shared by existing approaches, such as FreeFine [2] and SceneDiffusion [3], is that object movement occurs purely within the 2D image plane, meaning relocated objects are not rescaled according to their new depth positions, leading to inconsistencies in perspective and spatial realism.
- Zero-Shot Depth-Aware Image Editing with Diffusion Models (ZSDAD) [1] enables depth-conditioned object insertion, allowing new objects to be placed at specific scene depths. However, it does not support intra-image object transformations, such as moving an existing object to a new position while maintaining scale.
- DeGLaD & FeatGLaC, proposed in ZSDAD [1], are methods for scene decomposition and reconstruction, respectively. DeGLaD separates a scene into foreground and background components for 3D object insertion, while FeatGLaC recomposes the layered scene representation.

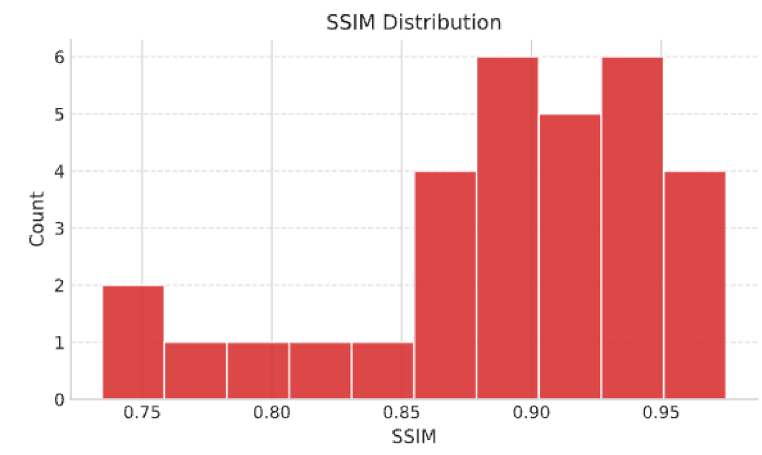
References

[1] Parihar et al., *Zero-Shot Depth-Aware Image Editing with Diffusion Models*, ICCV, 2025
[2] Pandey et al., *Diffusion Handles: Enabling 3D Edits for Diffusion Models by Lifting Activations to 3D, CVPR*, 2024
[3] Zhu et al., *Training-Free Geometric Image Editing on Diffusion Models, ICCV, 2025*

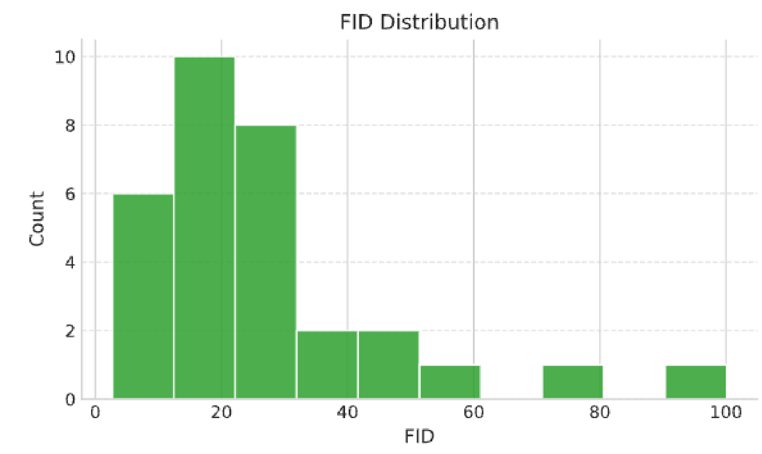
Experimental Results

Original Image

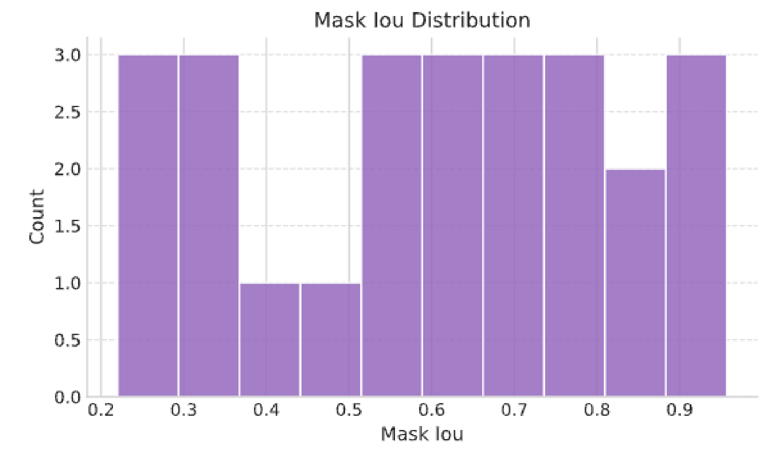
Object Repositioning Results



- SSIM (n=33): 0.876 ± 0.075 (min 0.662, max 0.975) — generally high structural fidelity; a few outliers dip below 0.7.



- FID (n=33): 27.7 ± 22.7 (min 2.7, max 100.1) — quality varies widely; several samples near photographic realism, others degrade noticeably.



- Mask IoU (n=25): 0.591 ± 0.227 (min 0.220, max 0.957) — moderate placement alignment on average, with some cases exhibiting large mask mismatch.