



# Concept-Nerf: When your imagination shapes Reality

Jinang Shah

Department of Electrical Engineering, Stanford



## Problem Statement

- Developing a 3D editing system using concepts learnt from the given images to enhance customization capabilities
- The high-level goal is to empower users with enhanced control over the 3D editing process, thereby fostering creativity and improving immersive experiences across various domains.
- Existing view-consistency-aware 3D editing methods, while promising, lack sufficient customization options beyond basic text prompts, resulting in limited editing capabilities and omission of high-level details in the output.
- Although concept decomposition techniques have demonstrated effectiveness in 2D image processing, their application in guiding 3D customization and editing processes remains unexplored.
- **We achieve our goal with following:**
  - Allowing customization with inputs (images and text)
  - Learning better embeddings of input through Textual Inversion
  - Using Textual Inversion embeddings to improve View-Consistency Aware NERF process for 3D editing

## Background

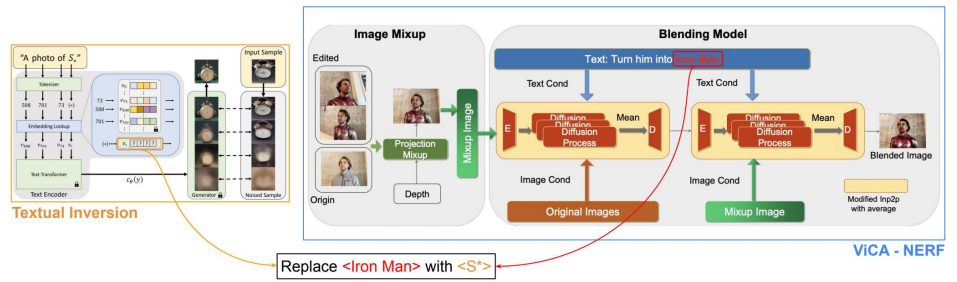
In the realm of 3D editing, several approaches have been explored to enhance the customization capabilities and realism of generated scenes. Notable methodologies include:

- View-Consistency-Aware 3D Editing of Neural Radiance Fields (ViCA-NeRF) [1], utilizes recent advancements in text-to-image models [2] and diffusion models [3] to perform text instructions guided 3D editing, similar to how Instruct Nerf-2-Nerf, first edits 2D key viewpoints to retrain 3D scene
- Both methods rely solely on text-guided 3D editing and does not allow for more images-based customization of the prompt
- Textual Inversion paper [4] learns a single embedding describing inherent concepts embedded and shared between given images using Latent Diffusion Models
- Concept Decomposition paper [5] goes further in learning new concepts from the concepts learnt through textual inversion.

## References

1. ViCA-NeRF: View-Consistency-Aware 3D Editing of Neural Radiance Fields
2. Learning Transferable Visual Models From Natural Language Supervision
3. High-resolution image synthesis with latent diffusion models
4. An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion
5. Concept Decomposition for Visual Exploration and Inspiration

## Approach



## Results

### Textual Inversion



### 3D Nerf Editing

