

Project Proposal - B09611007 Po-Lin Chen

Project Topic: Generating Realistic Object-Specific Visuals using Deep Learning

Objective: To implement textual inversion to represent a specific object and generate realistic images of it in various contexts.

Research Method:

- 1. Object Representation with Few-Shot Learning:** Create a unique token (*) for representing a specific object using 4-6 reference images as ground truth for training. Following the approach described by Gal et al. (2022), the framework integrates a text encoder with an image generator based on the stable diffusion model.
- 2. Image Generation and Evaluation:** Use prompts, such as "A photo of * in a greenhouse," to generate images that depict the object within specified environmental contexts. These images will serve as the basis for performance evaluation.
- 3. Comparison and Optimization:** Conduct comparative analyses of the generated results under various conditions, including:
 - Training hyperparameters
 - Image generation settings
 - Prompting strategies

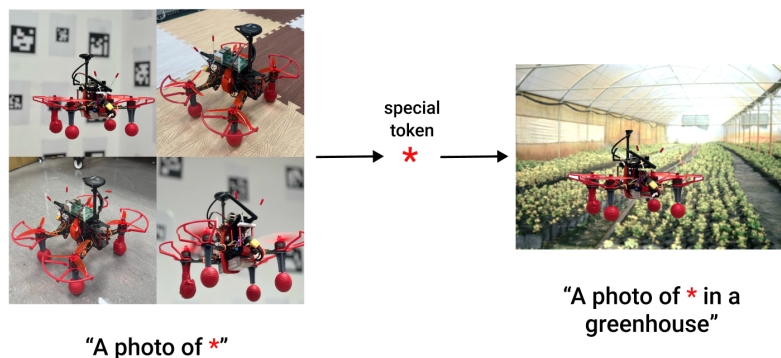


Figure 1. Project overview

References :

Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A. H., Chechik, G., & Cohen-Or, D. (2022). An image is worth one word: Personalizing text-to-image generation using textual inversion. arXiv preprint arXiv:2208.01618