

# FYP INTERIM REPORT

**Project ID:** CCDS24-0163

**Title:** Automated Image Generation -- Enhancing Multi-Conditional Training-Free Image Generation with Interaction-Aware Gaussian Kernels

**Name:** Hill Seah Wen Qi

**Matriculation Number:** U2121346H

**School:** Nanyang Technological University – College of Computing and Data Science

**Date:** 27/01/2025

**Abstract**

Multi-conditional image generation aims to synthesize images that satisfy diverse conditions, such as textual descriptions, segmentation masks, and landmark constraints. While training-free approaches to image generation often excel in single-condition scenarios by leveraging off-the-shelf, open-source pre-trained networks to estimate the distance between an intermediate image and the condition, they struggle with multi-conditional tasks. This limitation arises from the inability of training-free methods to effectively handle interactions and dependencies between multiple conditions. To address this challenge, this project investigates strategies for effective multi-conditional image generation. It introduces a novel framework that integrates a naive weighted sums of distance functions with interaction models that capture complex, non-linear interdependencies between conditions, to approximate the energy function and guide the iterative denoising process. Experimental results demonstrate that the proposed method outperforms existing techniques in generating coherent, condition-consistent images across a variety of conditions, showcasing its effectiveness in addressing the challenges of multi-conditional image generation.

**Acknowledgements**

The writer uses this section to thank all those he or she is indebted for guidance, financial or

any other assistance rendered during the course of the project.

**Table of Contents:**

1. **Introduction**
   1. **Problem Statement (Focus on generally, single cond heavy focused)**
   2. **Objective (want to create multi cond + effective multi cond)**
2. **Body**
   1. **Preliminaries (all the est formulas)**
   2. **Related Work**
   3. **Models (Naïve Multi Cond denoising formula, m1, m2, m3, …)**
   4. **Evaluation**
3. **Conclusion**
   1. **Limitations**
   2. **Recommendations**
4. **References**
   1. **Biblo**
   2. **Appendix**
   3. **Glossary**

**Introduction**

**Problem Statement**

< CRUX: training-free image generation often works well for single conditions but struggle with multi conditions due to the inability for training-free approaches to handle interactions between multi conditions well >

< Training-free over training required methods>

<Training-free cannot handle multi conditions as well (FIND OUT MORE) + citations>

Multi-conditional image generation has emerged as a critical area in computer vision, where the goal is to generate images that simultaneously satisfy multiple user-defined constraints. For example, generating a facial image that aligns with textual descriptions, adheres to geometric landmark constraints, and respects segmentation masks. For single-condition image generation, training-free approaches often leverage off-the-shelf, open-source pre-trained networks to estimate the distance between an intermediate image and the condition, guiding iterative denoising processes (Freedom citation).

<Insert how training-free approaches struggle with mutli conditions + reasons + citations>

Its inability to account for interactions between conditions results in:

* Conflicting conditions being treated in isolation.
* Loss of coherence in generated outputs.
* Limited adaptability to complex, real-world multi-modal scenarios.
* <list more>

TODO:

* Seriously read the report + research done previously
* Identify the crux of the report
* See how to organize body \*see research portion headers (decide the main flow + idea you want to present)
* Redo Intro (PS) and abstract