# Midterm Project: AI-Generated Drawings with CLIP Scoring

Zhiyu Cheng

March 17, 2025

# Introduction to the Kaggle Competition

- This competition challenges participants to generate images from text descriptions.
- The goal is to create images that align with the given prompts and are evaluated using the CLIP model.
- CLIP scores determine how well an image represents the given text.

# Text Rendering Ban and AI Model Approach

**What is Text Rendering?**

- Text rendering involves embedding the text description directly into the generated image.
- This method is banned in the competition to encourage real image generation.

**AI Model Approach (Stable Diffusion)**

- I use a **Stable Diffusion** text-to-image AI model.
- It generates images based on textual descriptions without embedding the text.
- The model learns from large-scale datasets to create realistic and contextually accurate images.

# CLIP Scoring: Evaluating Image Quality

- CLIP (Contrastive Language–Image Pretraining) scores how well an image aligns with a given text description.
- A higher CLIP score indicates a closer match between the generated image and the prompt.
- Typically, a **CLIP score above 55** (replace with final results) is considered a solid drawing.

# Why Use Color/Style Variations?

- Different colors and styles help highlight important features in an image.
- Certain variations (e.g., contrast adjustment, sketch styles) align better with CLIP's learned representations.
- Applying these variations improves CLIP scores and enhances visual clarity.

# Image Comparison: Prompt 1

**Prompt:** *"A futuristic cityscape with neon lights at night."*

**Without Variations**

**With Variations**

# Image Comparison: Prompt 2

**Prompt:** *"A watercolor painting of a sunset over the ocean."*

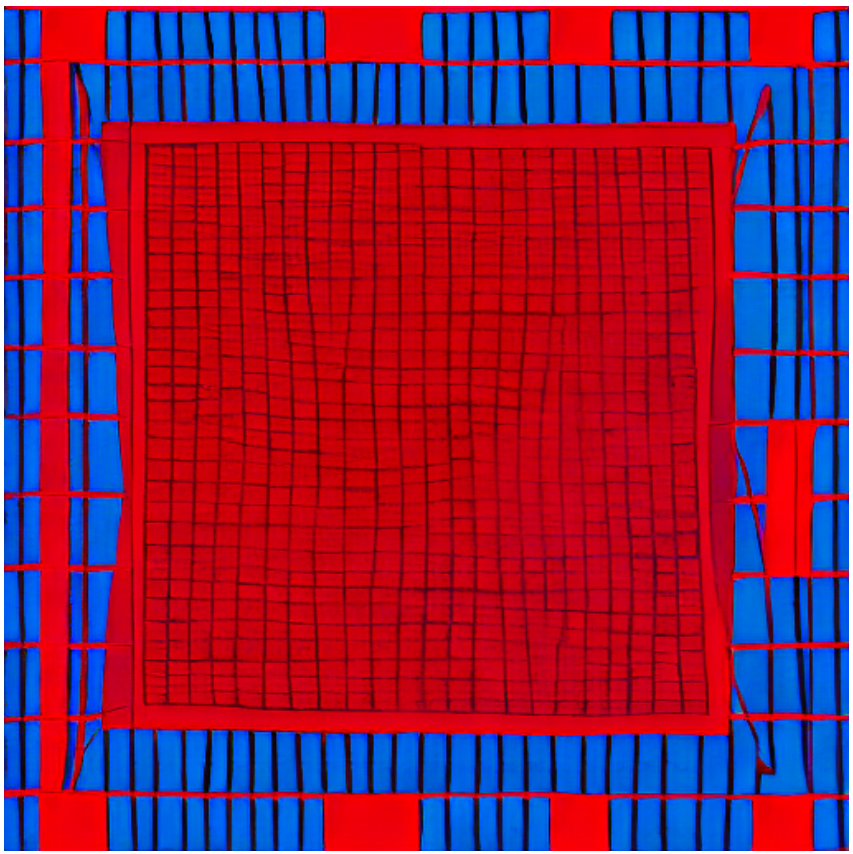
**Without Variations**

**With Variations**

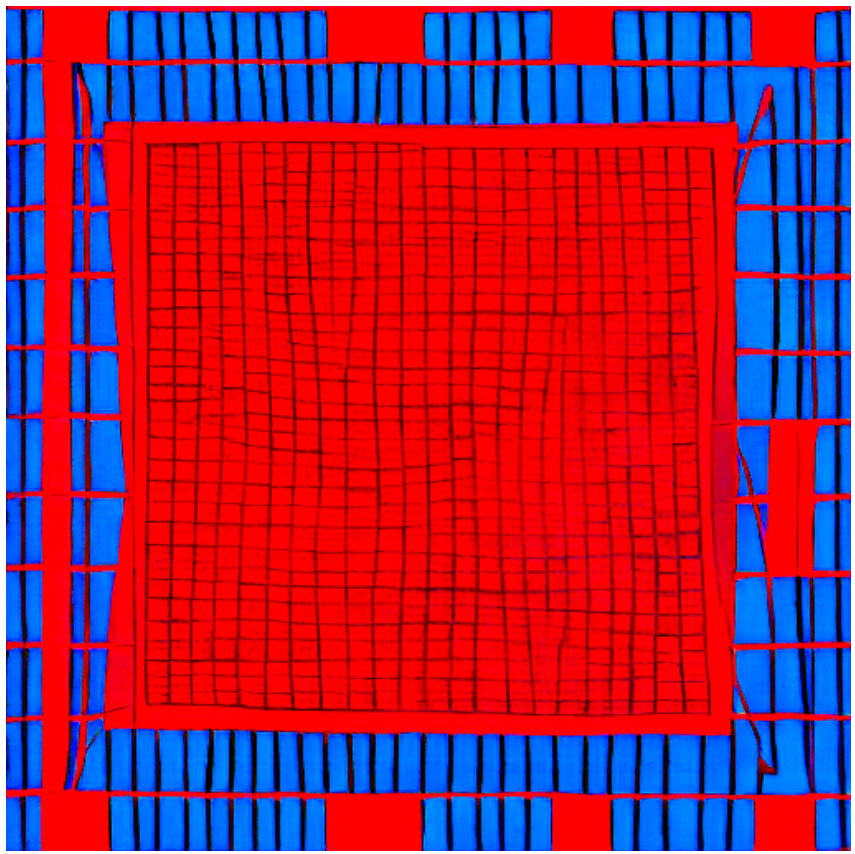
# Image Comparison: Prompt 3

**Prompt:** *"A pencil sketch of a mountain range with fog."*

**Without Variations**



**With Variations**



# Image Comparison: Prompt 4

**Prompt:** *"A detailed oil painting of a historical battle scene."*

**Without Variations**

**With Variations**

# Image Comparison: Prompt 5

**Prompt:** *"A cyberpunk-styled futuristic marketplace."*

**Without Variations**

**With Variations**

# CLIP Score Differences

| Image | CLIP Score (Without Variation) | CLIP Score (With Variation) |
|---|---|---|
| Example 1 | 29.9684 | 29.8788 |
| Example 2 | 33.7195 | 34.2642 |
| Example 3 | 32.3884 | 32.6722 |
| Example 4 | 32.4716 | 31.7947 |
| Example 5 | 34.3191 | 33.0808 |
| **Average Score** | **32.5734** | **32.3382** |

# Conclusion

- The **average CLIP score does not change significantly** after applying variations.
- However, for **some images, the score increases significantly**, suggesting strong improvements in alignment.
- Images with color and style variations appear brighter than those without modifications.