PATTERNVERSE

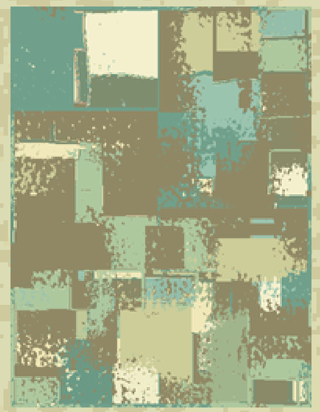
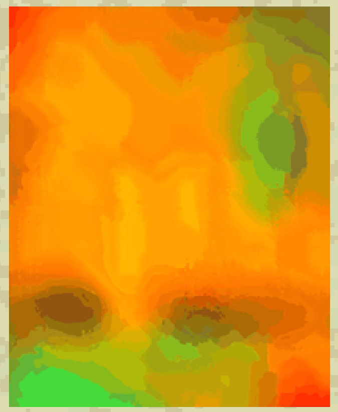
Presented By Team Quarters

INTRODUCTION

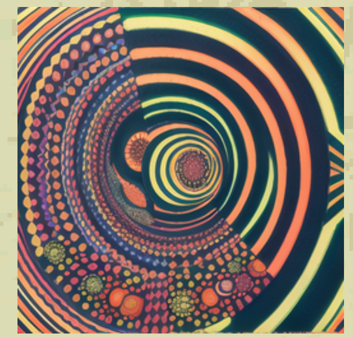
PROBLEM STATEMENT

- *Manual Design Inefficiency*
- *Lack of Automated Variation Tools*
- *Distinct Variation Generation Requirement*

Model


Model



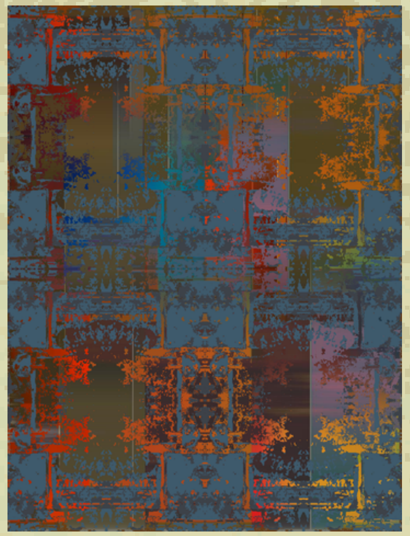
Training

Trained

Model

Trained

Model



Trained

Model

Our Methodology

- Research And Literature Review
- Variant Relationship Analysis
- Model Selection and Finetuning
- Hyperparameter Tuning and Optimization
- Model Inference and Variant Generation

1. Research & Literature Review

As per requirements of this challenge

Style Transfer Apply the artistic style of one image to another while preserving the content structure.	**GAN** Generator produces realistic data samples, Discriminator distinguishes real from generated data.
Encoder-Decoder Architecture Encoder compresses input data into a latent representation, Decoder reconstructs the desired output.	**Diffusion Models** Generative models that iteratively refine noisy inputs into high-quality samples.

Legend

Red	Unfeasible
Light green	Feasible
Dark green	Best Suited

1. Research & Literature Review

Diffusion Models:

Denoising Diffusion Probabilistic Models	A stochastic approach that progressively refines images from noise
Latent Diffusion Models	Compress data into a latent space for efficient computation.
Conditional Diffusion Models	Incorporate external conditioning signals, to guide the generation process for better control over output features.

Stable Diffusion

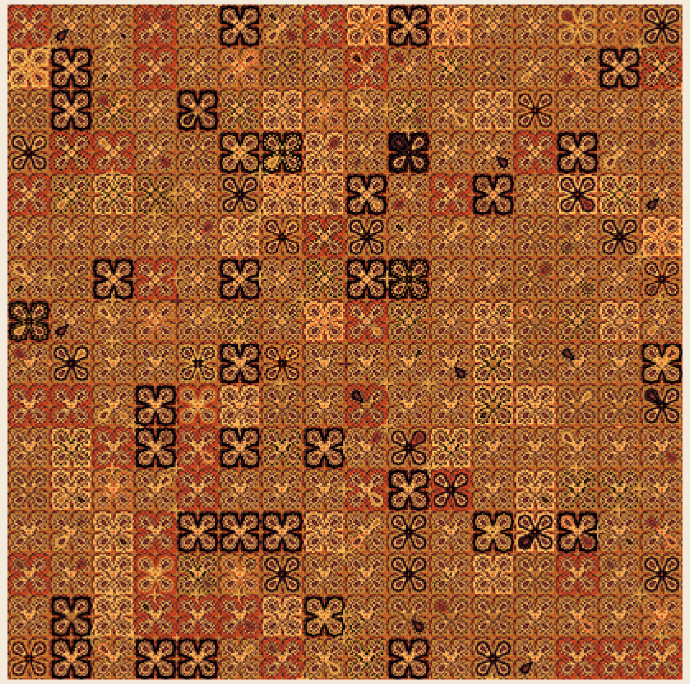
2. Variant Relationship Analysis

File Summary:

Total Images: 34686
Corrupted Files: 0
Unique Resolutions: 14
Unique Color Modes: 3

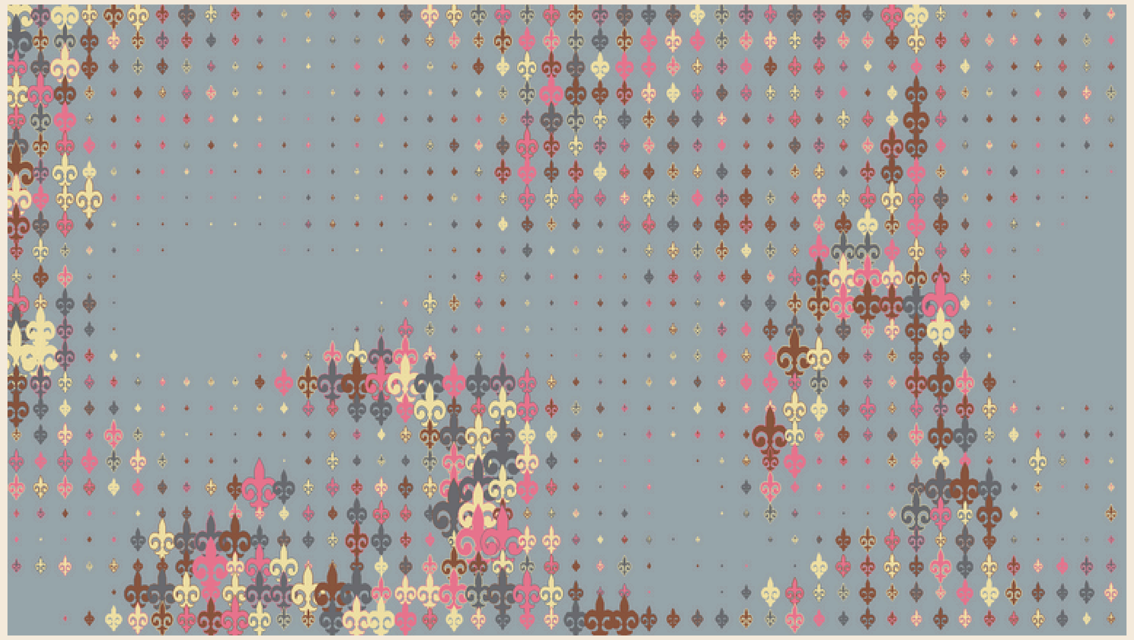
Color Modes:

P: 28949 Images
RGB: 5685 Images
RGBA: 52 Images

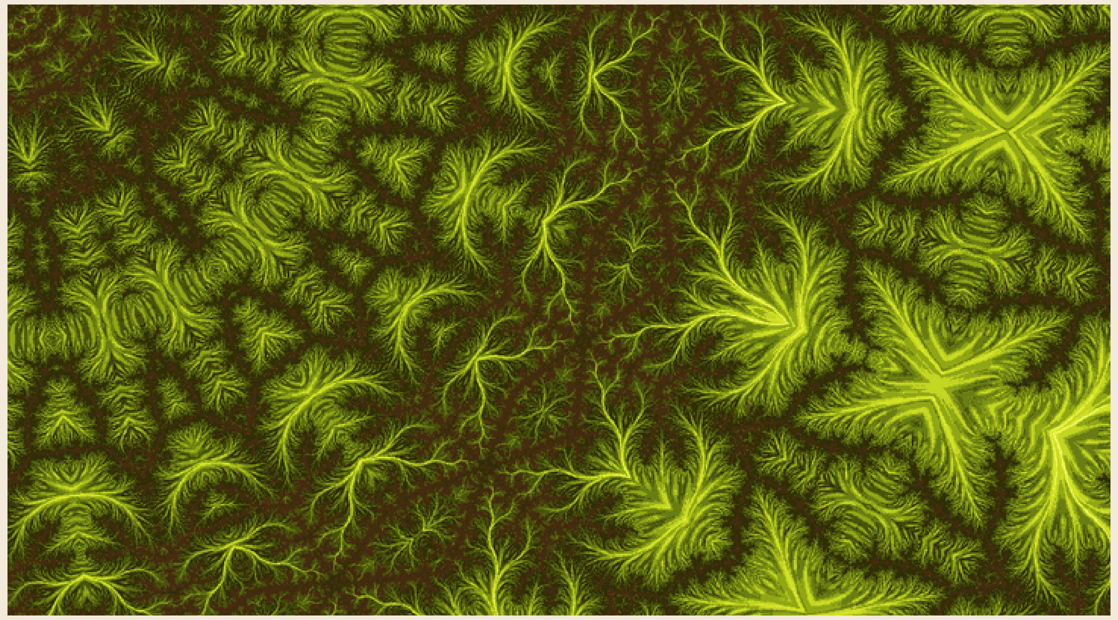


Palette

RGBA



RGB



We converted all modes into RGB

2. Variant Relationship Analysis

Resolutions:

- 910x512: 2,297 images
- 512x512: 5,075 images
- 932x512: 120 images
- 885x512: 152 images
- 877x512: 30 images
- 875x512: 60 images
- 1920x1080: 50 images
- 384x512: 25,855 images
- 900x1200: 190 images
- 409x512: 447 images
- 382x512: 300 images
- 888x1184: 99 images
- 960x1280: 6 images
- 1120x1120: 5 images

Seeing these resolutions, We decided to train our model in 512 X 512 dimensions to capture essential details

3. Model Selection & Finetuning

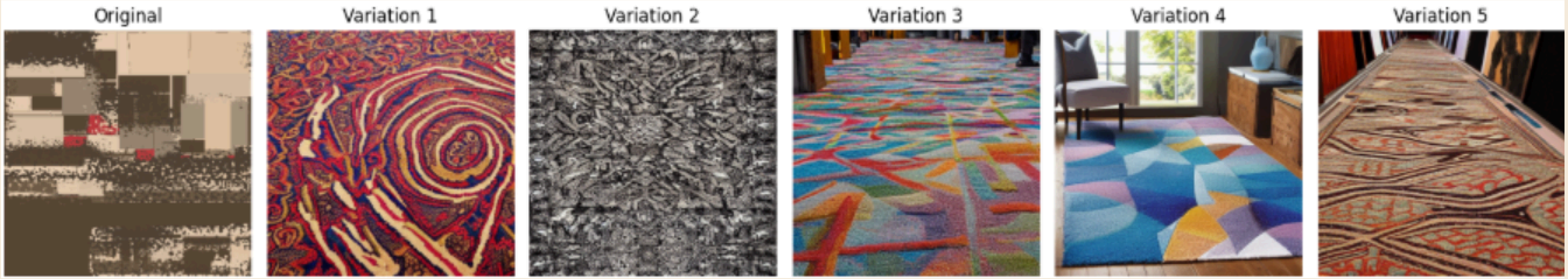
1. Initially, we tested various Stable Diffusion pipelines:
   - Standard Stable Diffusion Pipeline
   - Stable Diffusion Upscale Pipeline
   - Stable Diffusion Inpainting Pipeline
   - Stable Diffusion Img2Img Pipeline
   - Stable Diffusion ControlNet Pipeline in Guess Mode
   - Stable Diffusion ControlNet Img2Img Pipeline
   - Stable Diffusion Image Variation Pipeline

2. Selection

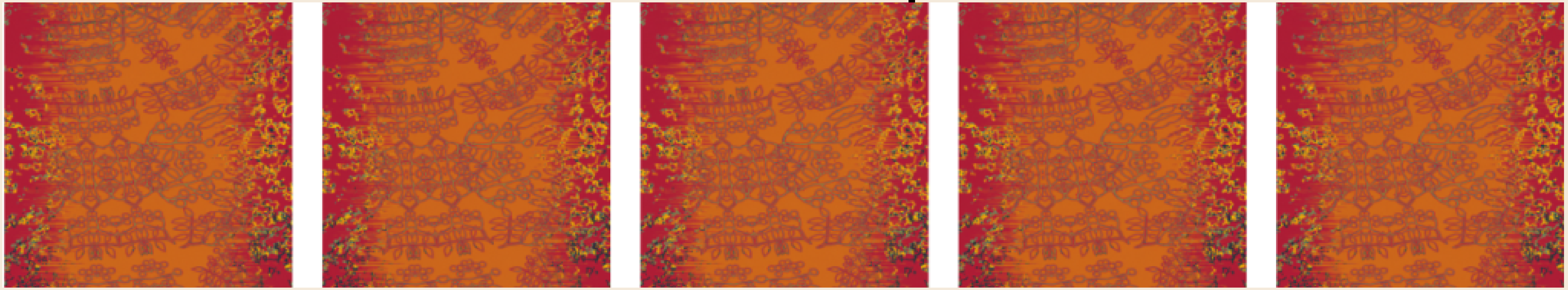
3. Finetuning

3.1 Pipeline Testing

Prompt: "An artistic carpet"



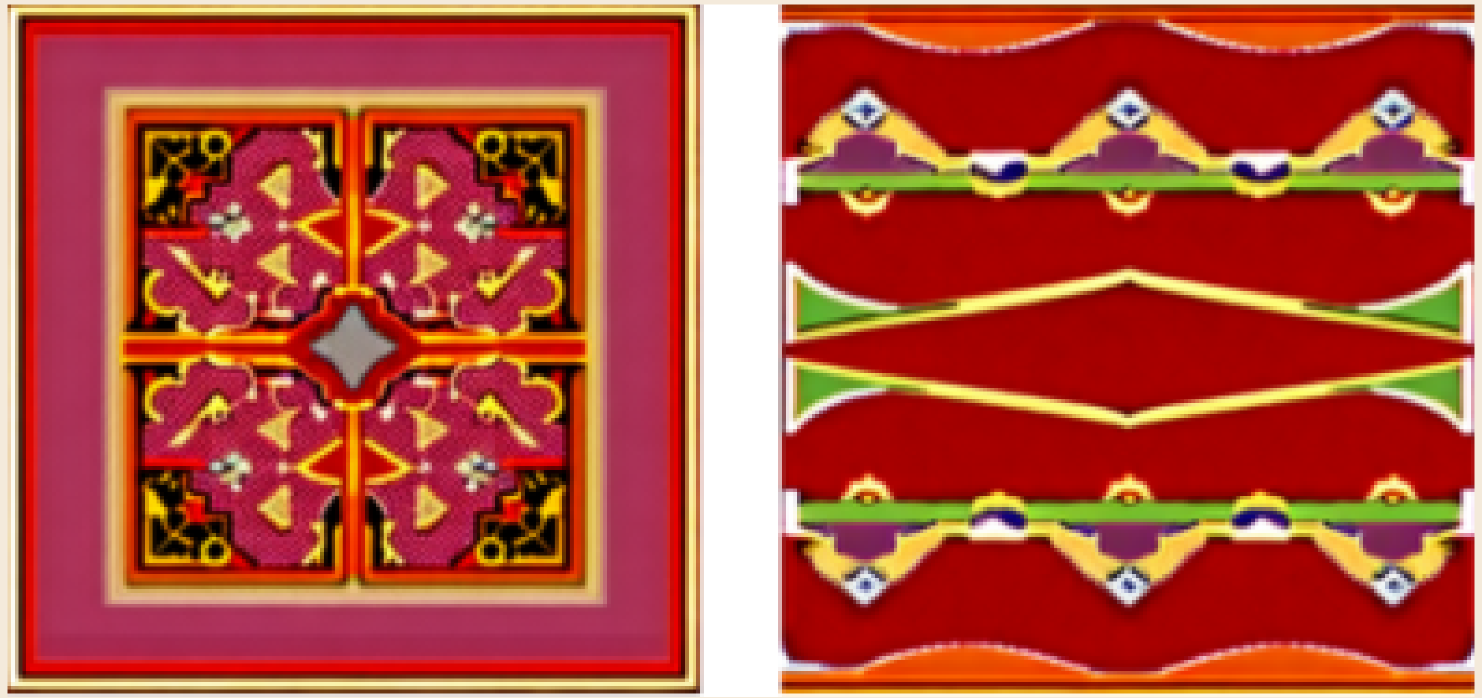
StableDiffusionPipeline



StableDiffusionUpscalePipeline
(for low resolution details enhancement)

3.1 Pipeline Testing

Input



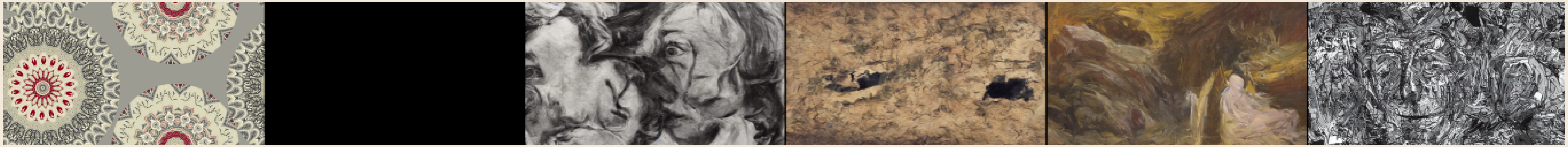
Output

StableDiffusionInPaintingPipeline

StableDiffusionImg2ImgPipeline

3.1 Pipeline Testing

StableDiffusionControlNetPipeline in Guess mode



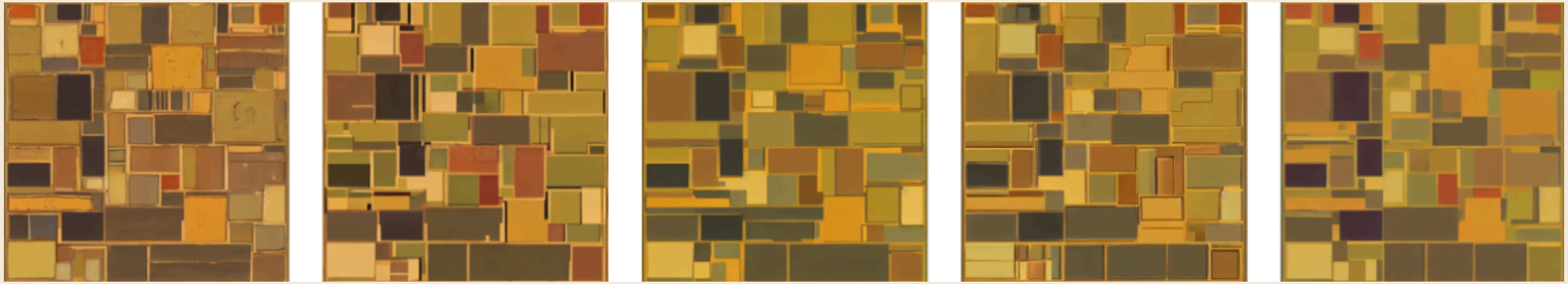
If the image to be inferred can't produce a Canny edge map or compute depth, we can't generate variations (Figure 2nd).

3.1 Pipeline Testing

StableDiffusionControlNetImg2ImgPipeline

Input

Output



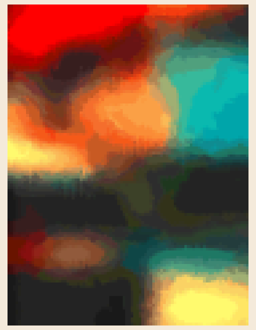
Tries to preserve the structure more rather than generate variations

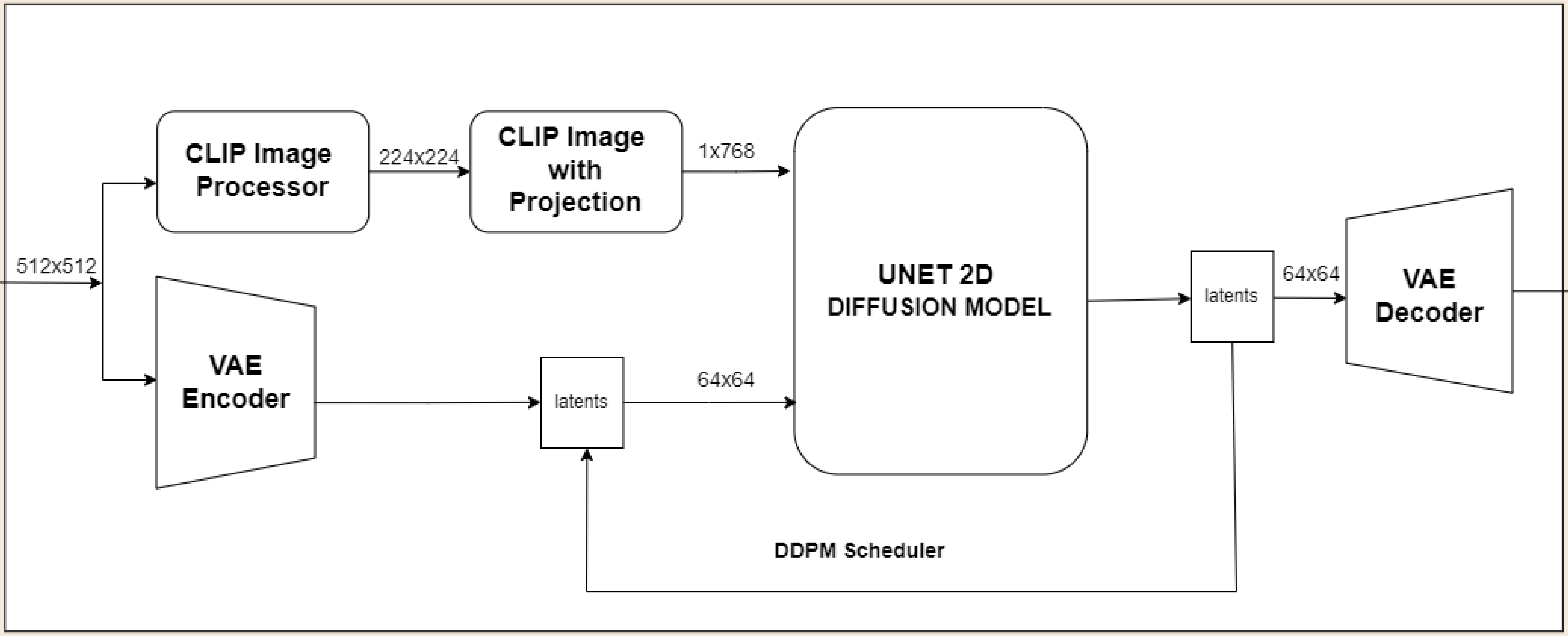
3.1 Pipeline Testing

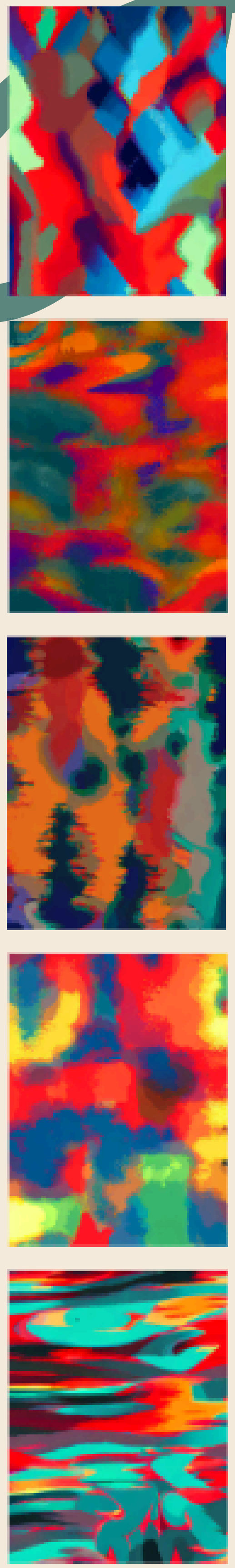
Stable Diffusion Image Variation Pipeline

This learns the different structure and features of the given image, and that's why we choose this pipeline.

3.2 Selection of the Pipeline: Image Variation







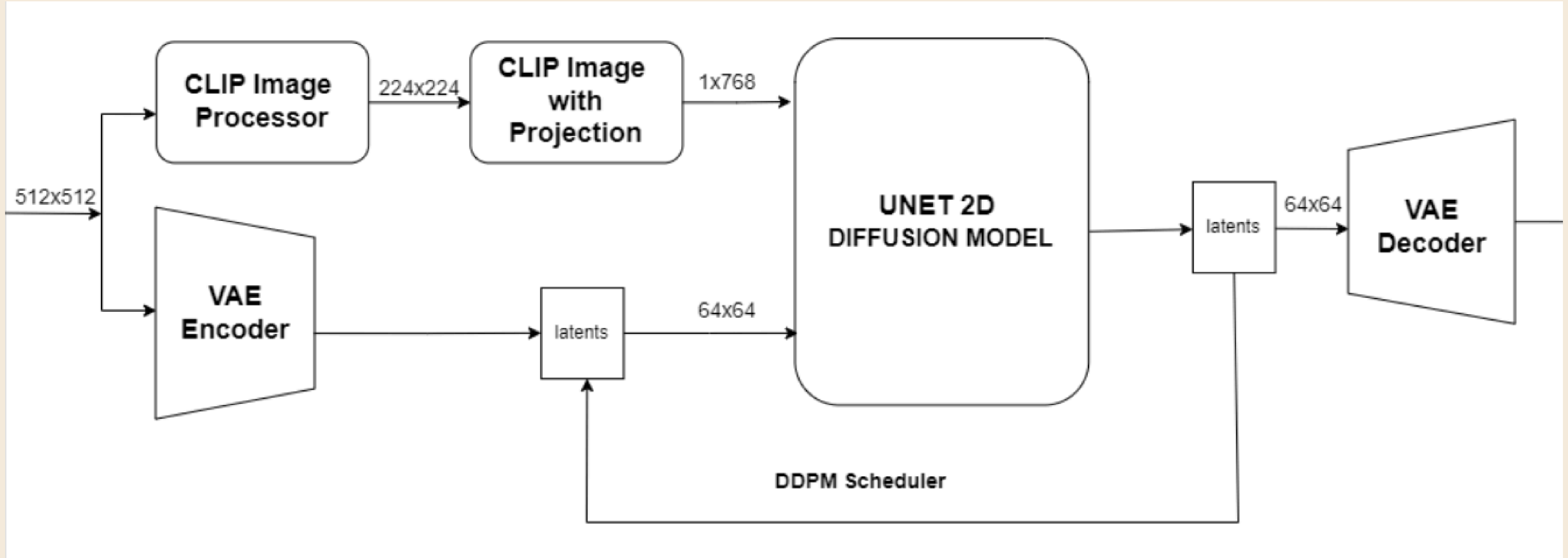
Components

CLIP Image Processor
CLIP Image With Projection

VAE Encoder
VAE Decoder

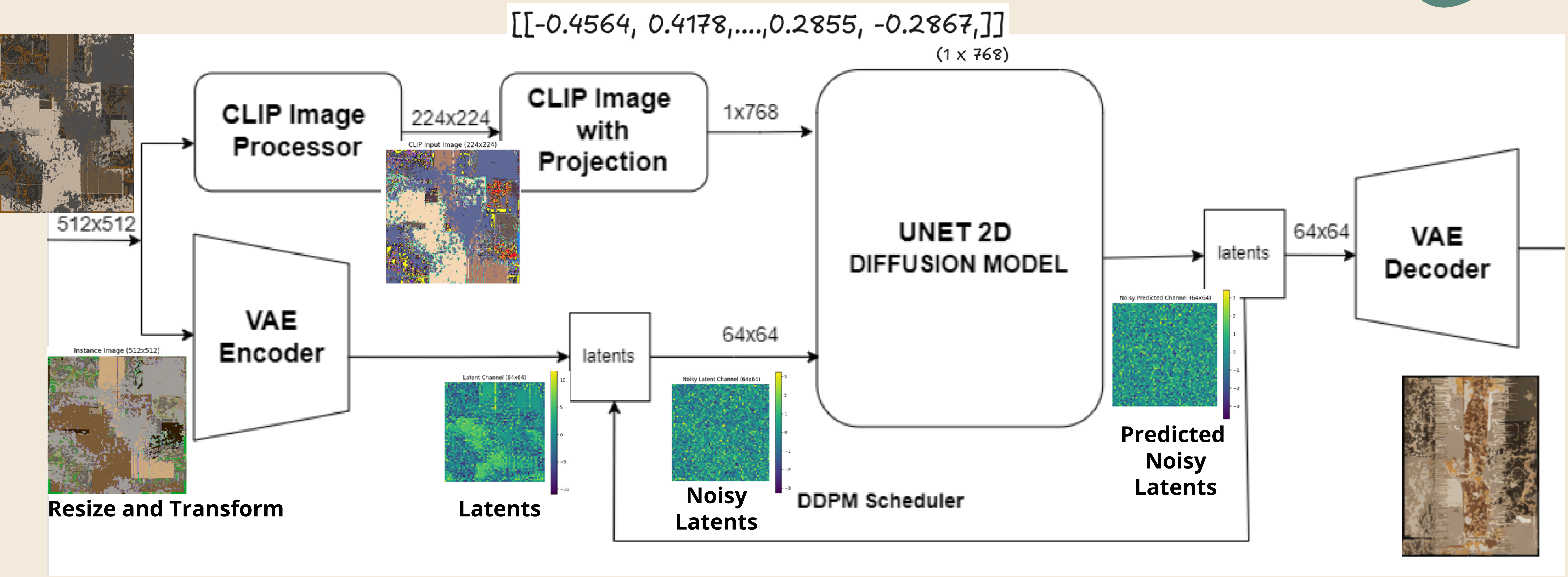
UNET 2D Diffusion Model
DDPM Scheduler

3.2 Selection of the Pipeline: Image Variation



CLIP Image Processor: Feature extractor
CLIP Image With Projection: Image Embeddings
VAE Encoder: Converts to latents understanding high level features
UNET 2D Diffusion Model: Denoising Noisy Features guided by embeddings
DDPM Scheduler: Controls the Noise Schedule
VAE Decoder: Generates Image from latents

3.2 Pipeline Workflow Visualization



3.3 Fine-Tuning

1. Model Components and Setup:

UNet: The primary trainable model for denoising.

VAE and CLIP Encoder: Pre-trained frozen components

2. Dataset Preparation:

Custom Image Dataset preprocessed with resizing (512, 512), normalizing and CLIP embeddings

3. Training Objective

Trained to minimize MSE between predicted and actual noise in the latent space.

4. Hyperparameter Tuning & Optimization

4.1 Hyperparameter Tuning

Category	Details	Value	Rationale
Learning Rate	Initial learning rate	2e-06	Ensures stable fine-tuning.
Batch Size	Training batch size	1	Limited by GPU memory
Number of Epochs	Total training epochs	80	Ensures sufficient adaptation to the dataset.
Gradient Accumulation	Simulate larger batch size	1	Increase to reduce memory usage.

4. Hyperparameter Tuning & Optimization

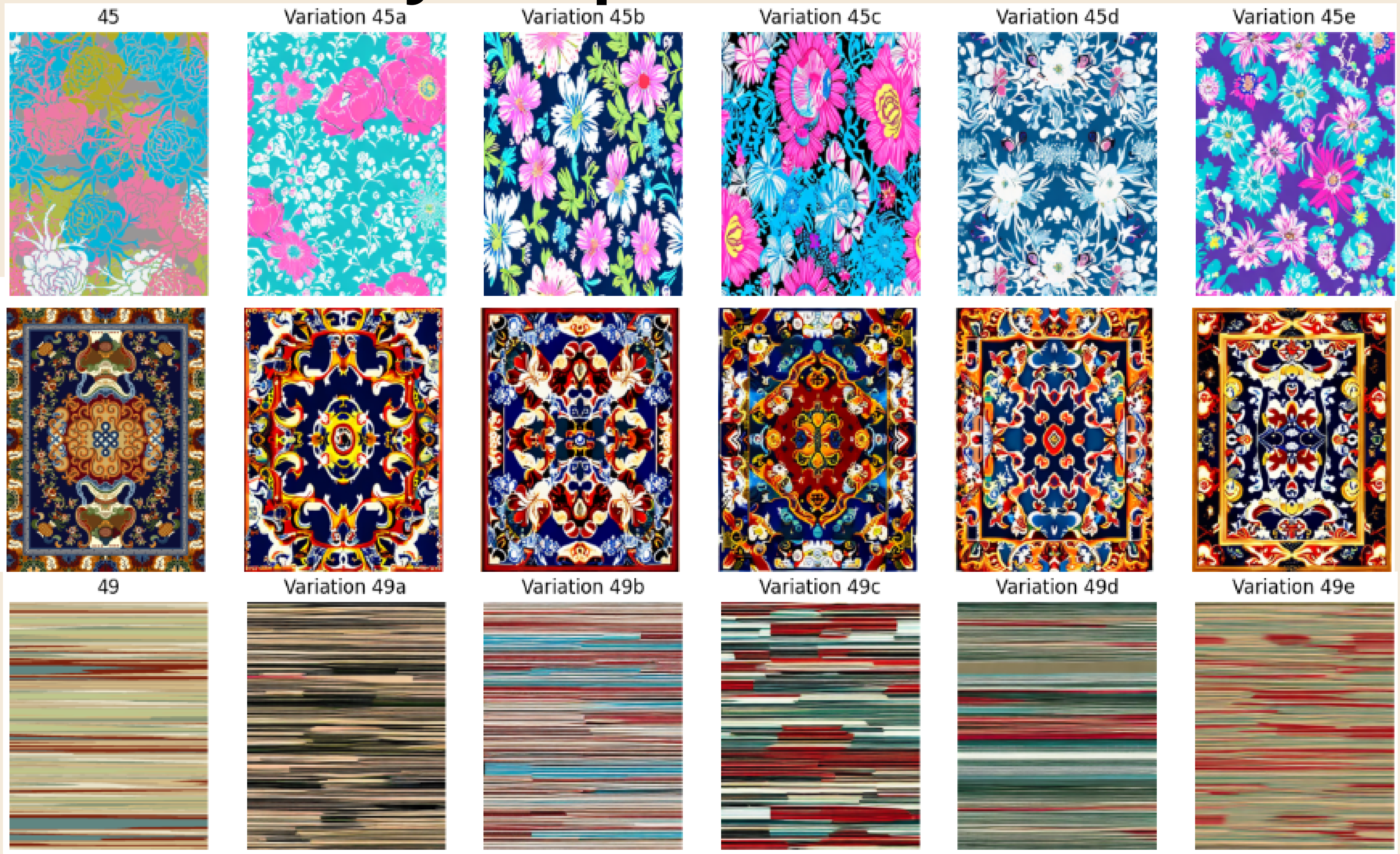
4.2 Optimization Techniques

Category	Details	Value	Rationale
Checkpoint	Save model state periodically	4000 steps	Allows resuming training and evaluating intermediate models.
8-bit Adam Optimizer	Reduce memory usage	Enabled (True)	Allows fine-tuning on GPUs with limited memory.
Mixed Precision Training	Use FP16 for reduced memory and faster training	Enabled (True)	Speeds up training and reduces memory usage.

5. Model Inference & Variant Generation

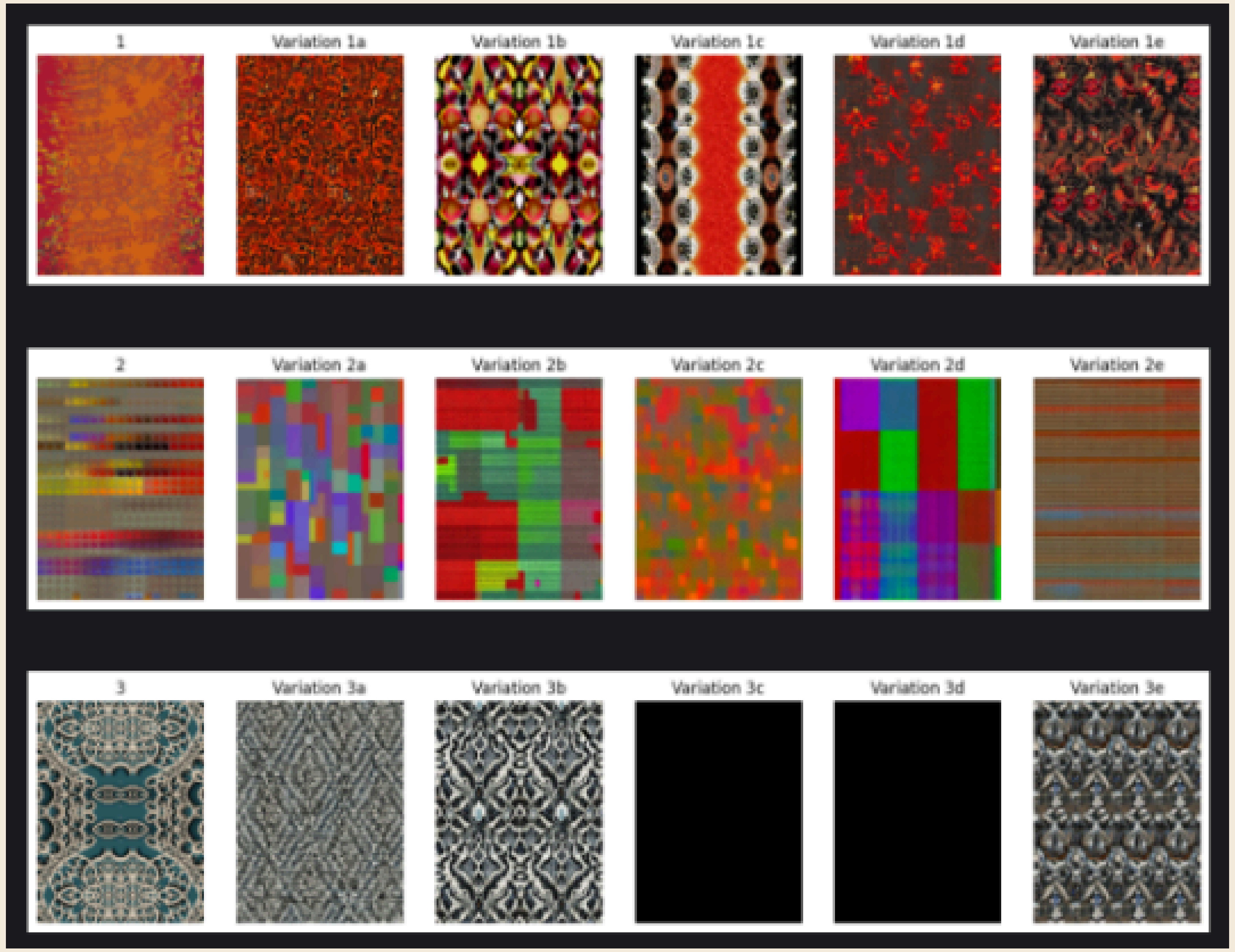
To generate the same set of images, we have used manual seed to 42

Inference by 80 epochs finetuned model:

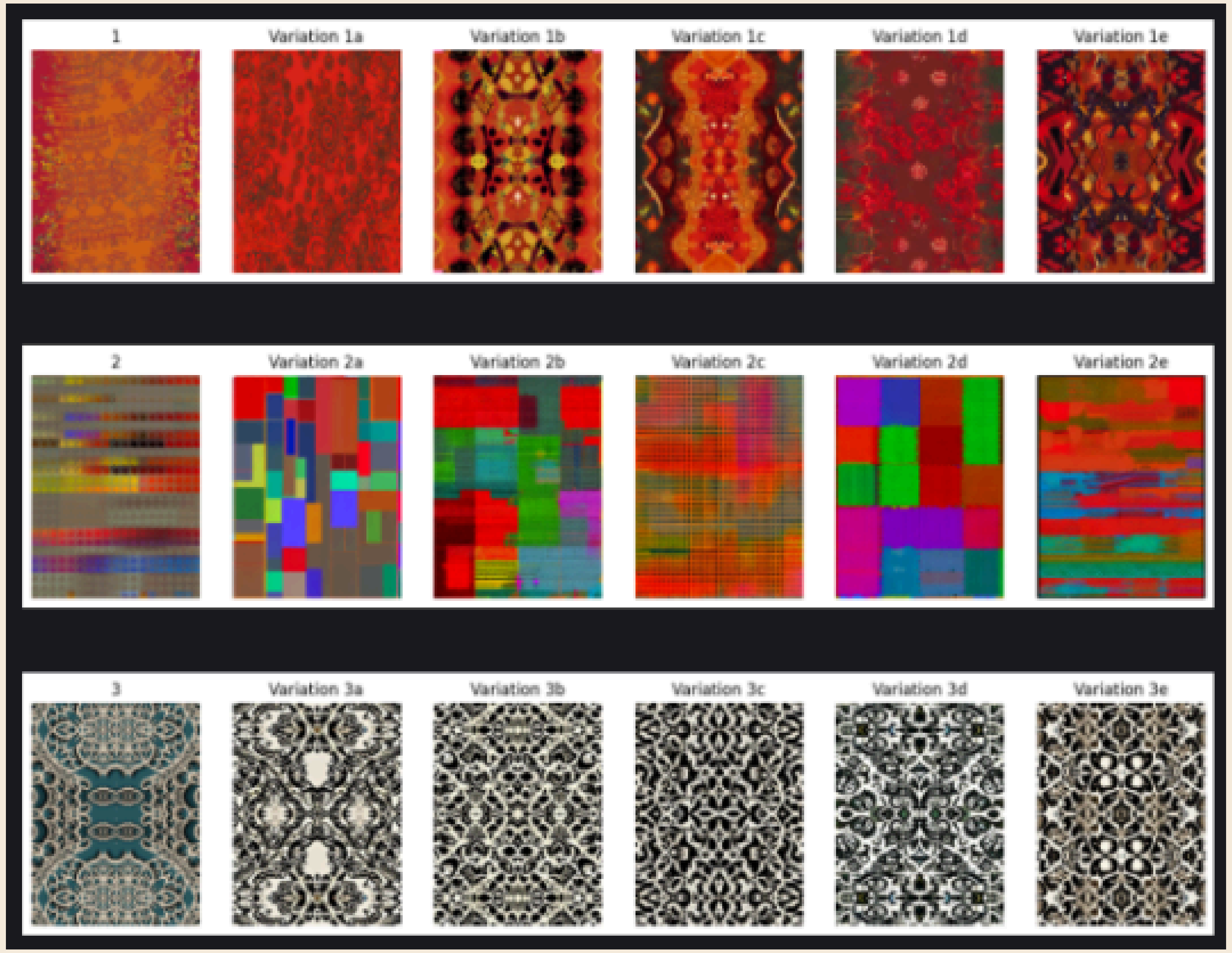


5. Comparison On Models

To generate the same set of images, we have used manual seed to 42



Pretrained ImageVariationPipeline



30 Epochs Finetuned

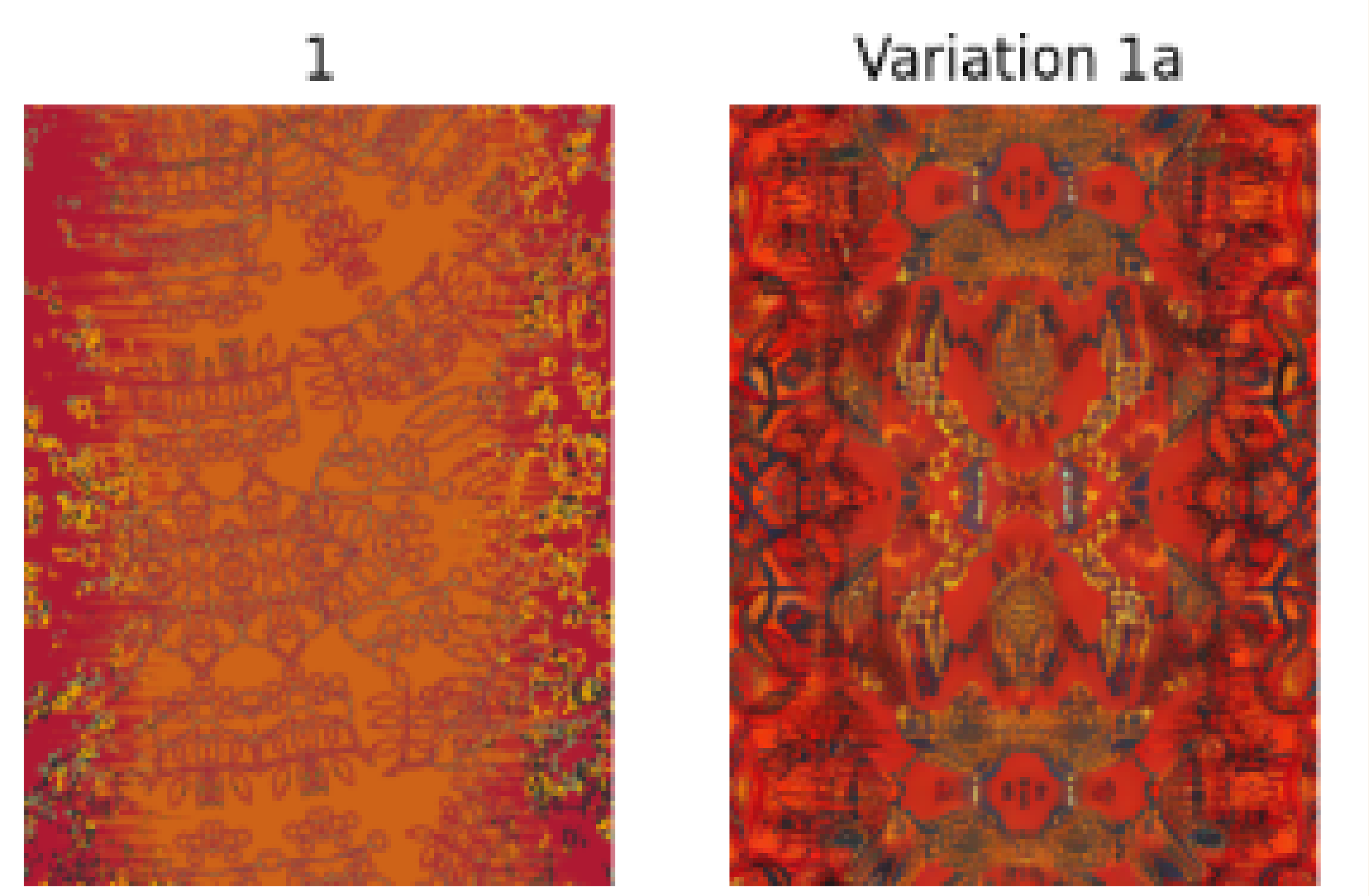
5. Model Inference & Variant Generation

50 Epochs Finetuned

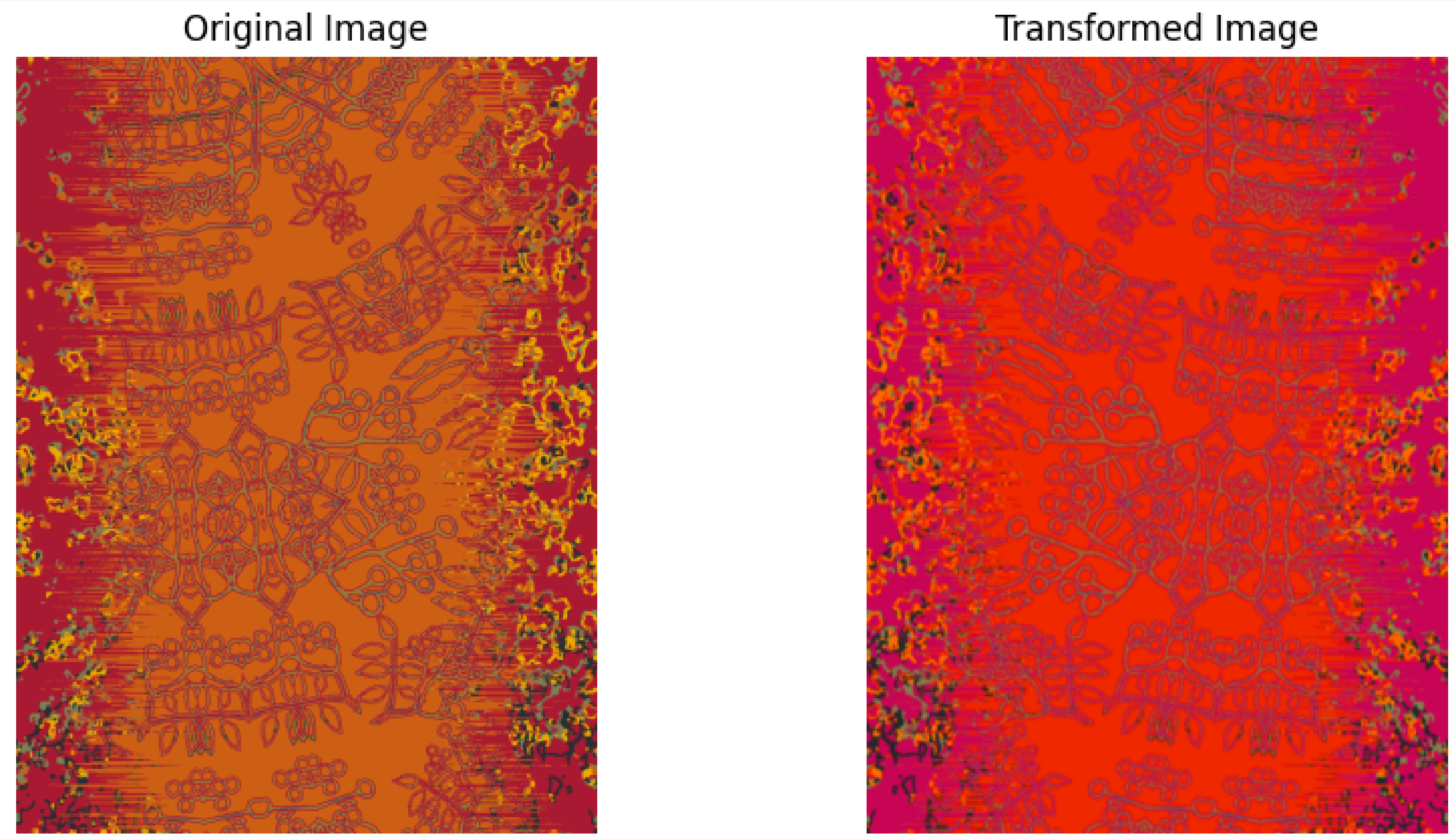
80 Epochs Finetuned

What to expect?

- Our finetuned model not only preserves the structure, it generates design.
- If you only want to preserve the structure, use basic augmentation techniques. No computation cost!



Our Model



Data Augmentation

Limitations

- Prompt based alterations in latent features for generation of variants.
- Limited to artistic and illustrative images.

Future Work

Train the model by providing an image and its variants within the same batch to explicitly learn correlated features.

Conclusion

Fine tuning the StableDiffusionImageVariationPipeline delivered the desired results.

MEET OUR TEAM

Rubika Bashyal

Sandesh Bashyal

Nijiya Maharjan

Thank you!