Progress Report

Training a Stripe-Mask ControlNet

Saruultugs Batbayar

12/04/2025

UNIVERSITY of ROCHESTER

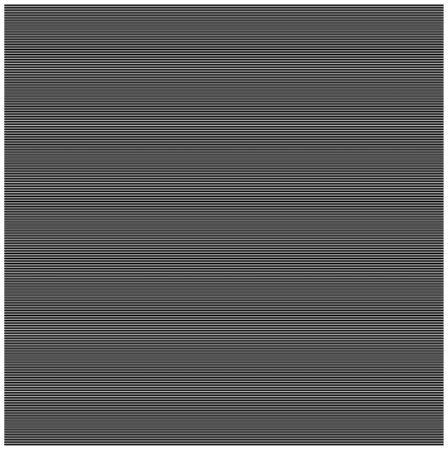
Problem Statement & Motivation

- Light-sheet microscopy and other imaging pipelines often produce **row-missing / stripe-like artifacts**.
- Existing inpainting models aren't optimized for **structured missing patterns**, especially high-frequency stripe damage.
- We want a model that can **reconstruct full images** from only ~20% - 25% visible rows.
- ControlNet gives us a way to **condition the UNet** on structured masks → stable training + controllable restoration.
- Training our own ControlNet allows us to support **custom masking patterns**, **custom datasets**, and **domain-specific reconstructions**.

GT (resized + random crop)

Striped RGB

Stripe Control Mask



Training Architecture

Input:

- **Striped RGB (3 channels):** masked image with missing rows
- **Binary Mask (1 channel):** indicates which rows are visible
- **4-channel ControlNet condition**

Text prompt → "high quality photograph" → encoder → embeddings

VAE encodes GT → latent representation

Noise sampled → added using DDPMScheduler

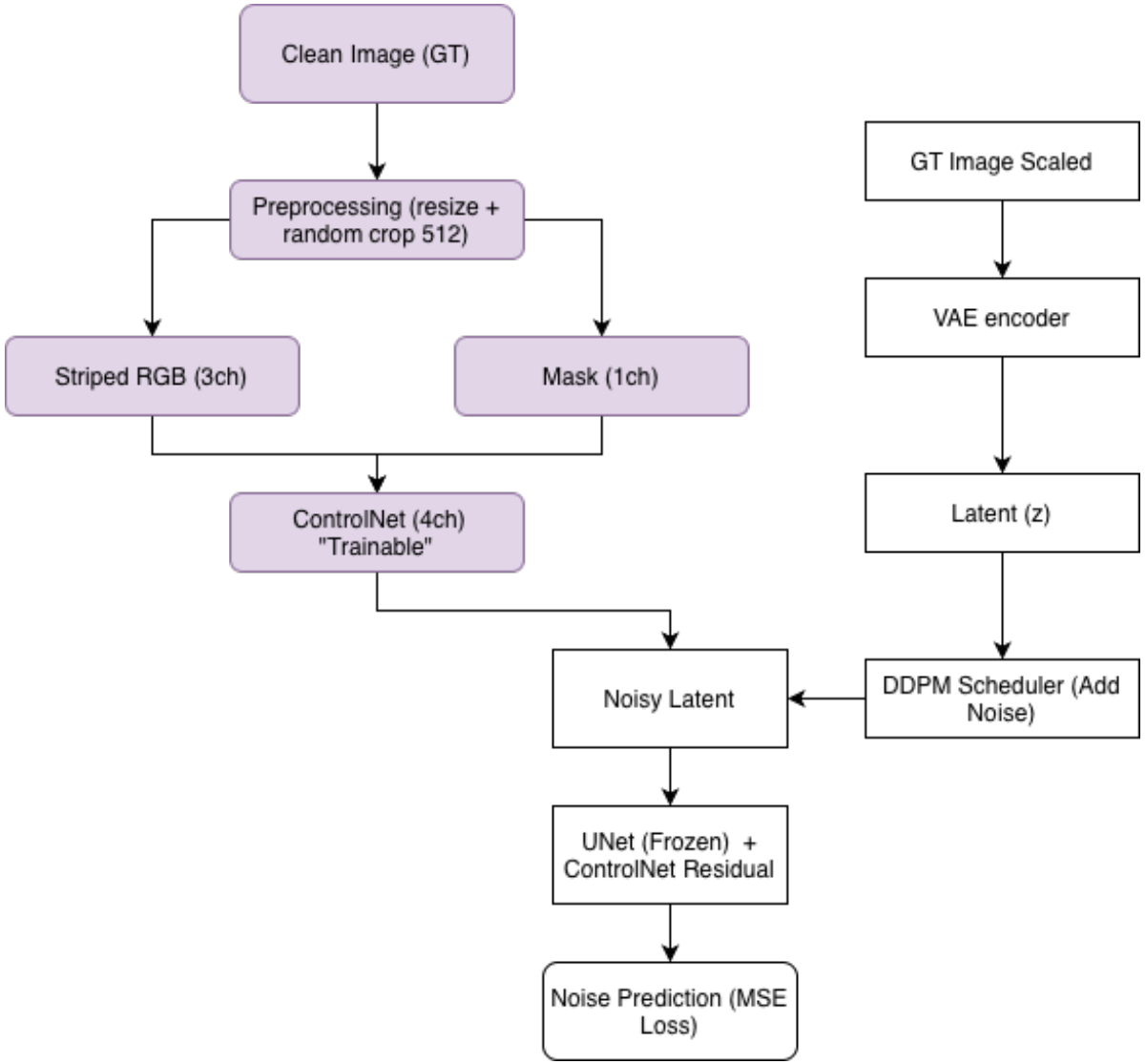
ControlNet processes control image → produces **residual feature maps**

UNet predicts the added noise using:

- noisy latents (H)
- timestep (t)
- text embeddings
- ControlNet residuals

Loss = **MSE(noise_pred, actual_noise)**

Training Architecture



Pipeline Steps

Gathered **43,525 COCO images** as clean training data.

1. Created a **dynamic online dataset**:

- Resize if needed (min side < 512)
- Preserve aspect ratio
- Random crop to 512×512
- Generate stripe masks using `skip_options=[3,4,5,6]`, `skip_probs=[0.15,0.5,0.25,0.1]`

2. Prepared **ControlNet inputs**:

- Striped RGB (scaled to [-1,1])
- Stripe mask (stays [0,1])

3. Loaded **Stable Diffusion 1.5**:

- VAE, Text Encoder, UNet (all frozen)

4 . Initialized **ControlNet with 4-channel conditioning**

5. Trained using:

- batch_size=42, grad_accum=3
- bf16
- Cosine LR schedule
- 100 epochs/ Trained 20 epochs so far (Approximately, 20 hours)

Key Improvements & Changes Made

1. Dynamic Online Dataset Generation

- Replaced pre-saved `gt_512 / striped_512 / stripe_control_512` folders
- All preprocessing (resize → crop → stripe mask) happens on the fly
- Increases data diversity and avoids huge storage

2. Improved Resizing Strategy

- Old: center-crop + resize (could distort images)
- New: aspect-ratio-preserving resize when min side < 512
- Then random 512×512 crop for better augmentation

3. Robust Error Handling for COCO Dataset

- Skips unreadable or corrupted images (`UnidentifiedImageError`, `OSError`)
- Re-samples a new image instead of stopping training
- Ensures stable, uninterrupted training across images

Visual Comparison On The Results

Avg Best LPIPS: 0.5885 | Avg Best SSIM: 0.1834

Avg Best LPIPS: 0.7487 | Avg Best SSIM: 0.0985

UNIVERSITY of ROCHESTER