PatchDDM-Based Inpainting Pipeline for CelebA

Diffusion-based Image Completion on Masked Facial Images

1. Dataset Preparation

- CelebA faces, resized to 64×64.
- Custom PatchMaskDataset applies square patch masking (e.g., 25%).
- Returns (masked_image, original_image) pairs for training.

2. Denoising Model: UNet with Time Embedding

- Encoder–decoder structure using residual blocks.
- Downsampling with MaxPool2d; upsampling with interpolation + skip connections.
- Time embedding encodes timestep t for conditional denoising.

3. Diffusion Framework (DDPM)

- Forward process: adds Gaussian noise to image ($x_0 \rightarrow x_t$).
- Reverse process: model predicts noise to denoise x_t.
- Loss: MSE between predicted noise and true noise.

4. Training

- Input: noisy image x_t from q_sample(original_image, t).
- Model output: predicted noise $\varepsilon_t(x_t, t)$.
- Loss: $MSE(\varepsilon, \varepsilon_t(x_t, t))$.
- Trained on masked images to enable inpainting.

5. Sampling / Inpainting

- Start from masked image + added noise.
- Iteratively denoise using the trained model.
- RePaint (optional): reinject known pixels during sampling to improve realism.

6. Visualization

- Side-by-side display: Masked Input → Reconstruction → Ground Truth.
- Early epochs may produce random outputs.
- Overfitting small batches confirms learning.