RePaint: Inpainting using Denoising Diffusion Probabilistic Models

Kate Liang, Huining Liu, Hongrui Tang, Jennie Wu, Mengjia Zhao

Date: May 1st, 2025

BACKGROUND & MOTIVATION

Image Inpainting
- Filling missing regions within an image

Context & Motivation
- Existing AR and GAN methods struggle with large or sparse masks.
- Diffusion models lacked harmonization of the known and generated part.

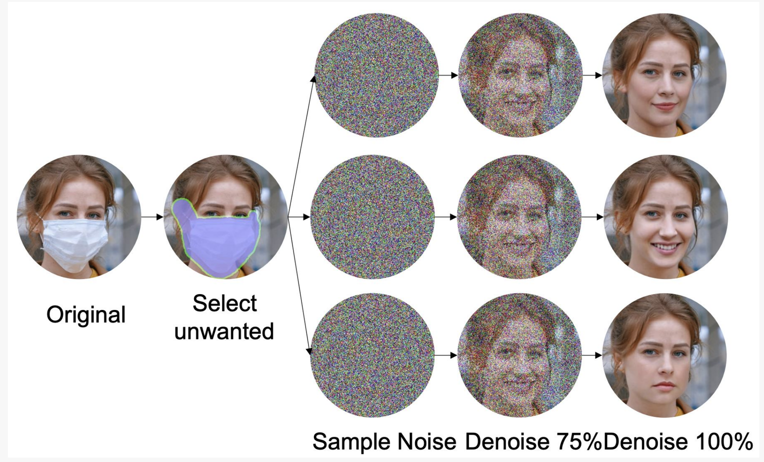


Figure 1: Diffusion-based inpainting

Goal
- Implement the reverse diffusion + resampling approach introduced by the paper
- Reproduce LPIPS results on the Celeb-A-HQ dataset

DATA & MODEL

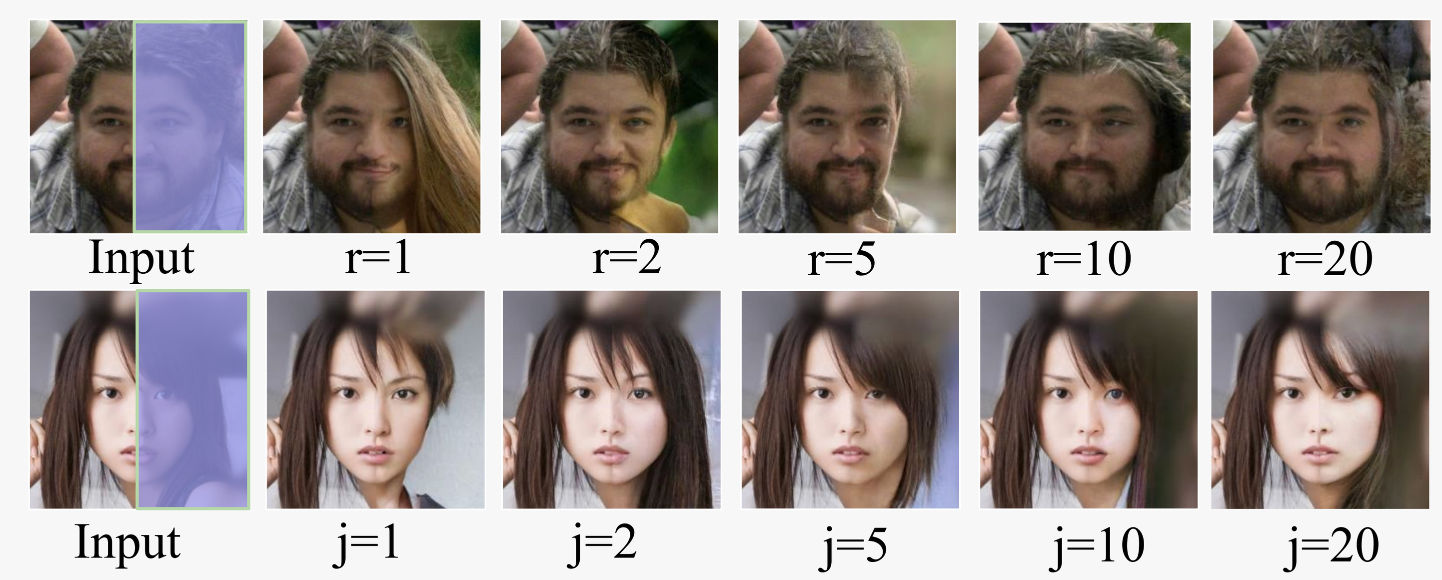
- **Dataset:** CelebA-HQ-256
- **Masks:** Wide, Narrow, Half, Expand, Alternate Lines, Super-Resolution
- **Model:** Pretrained google/ddpm-celebahq-256 DDPM from Hugging Face

EXPERIMENT RESULTS

Table 1: **CelebA-HQ Quantitative Results.** Comparison against state-of-the-art methods. We compute the LPIPS (lower is better) for 6 different mask settings.

CelebA-HQ	Wide LPIPS↓	Narrow LPIPS↓	Super-Res 2x LPIPS↓	Altern. Lines LPIPS↓	Half LPIPS↓	Expand LPIPS↓
AOT [6]	0.104	0.047	0.714	0.667	0.287	0.604
DSI [2]	0.067	0.038	0.128	0.049	0.211	0.487
ICT [4]	0.063	0.036	0.483	0.353	0.166	0.432
DeepFillv2 [5]	0.066	0.049	0.119	0.049	0.209	0.467
LaMa [3]	0.045	0.028	0.177	0.083	0.138	0.342
RePaint [1]	0.059	0.028	0.029	0.009	0.165	0.435
Ours	0.075	0.043	0.057	0.048	0.193	0.534

ABLATION EXPERIMENT



VISUAL RESULTS

Figure 4: Comparison against the state-of-the-art methods

METHODS

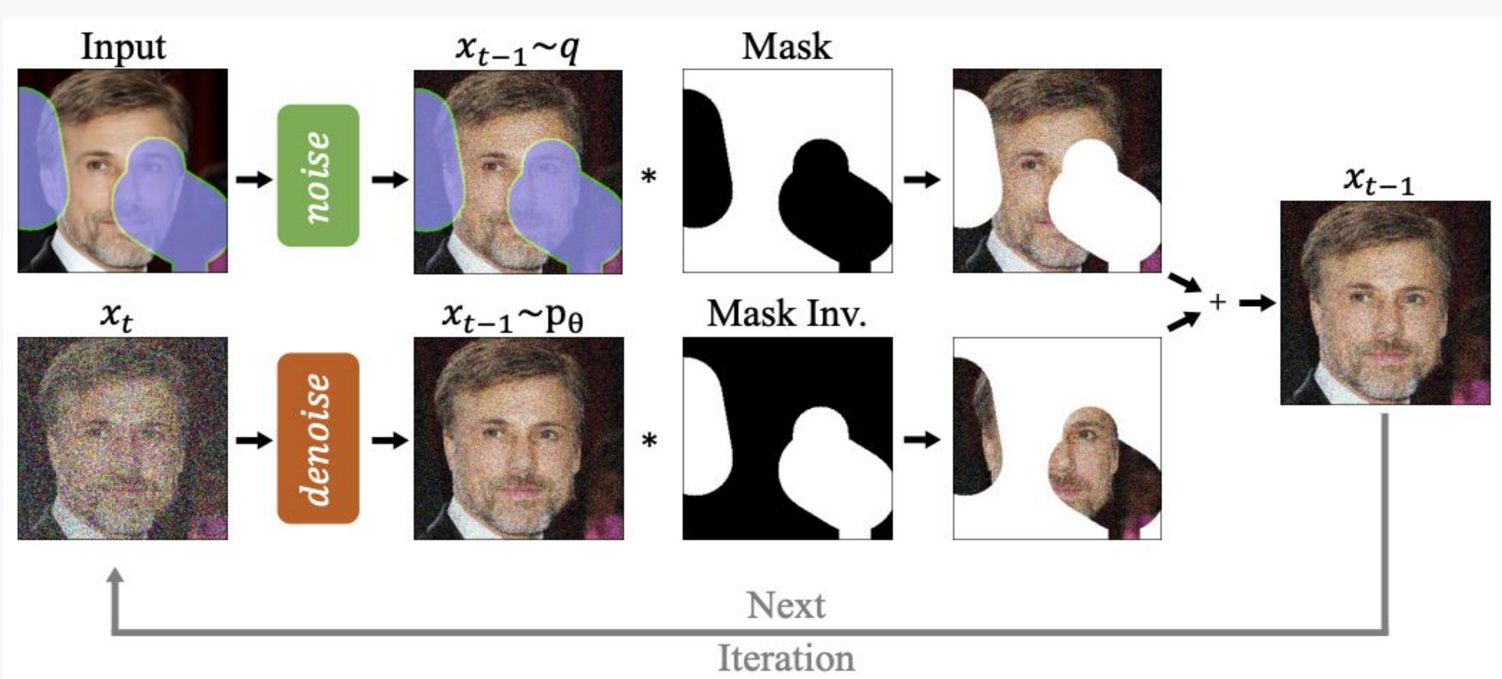


Figure 2: Overview of the approach

- **Condition Inference:** Generate image by guiding a pre-trained diffusion process conditioned on masked input
- **No Model Retraining**: Leverage pre-trained diffusion model
- **Content Harmonization:** Harmonize known and generated regions through iterative resampling
- **Bidirectional diffusion steps**: Alternate forward and backward in time
- **Larger jumps in diffusion time:** Improve perceptual quality and diversity

Design Choice:
- Utilized pre-trained model from HuggingFace
- Set total number of reverse steps to 1000
- Computed average LPIPS score for each mask based on 20 val images

Algorithm 1 Inpainting using our RePaint approach.

$x_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = T, \ldots, 1$ **do**
  for $u = 1, \ldots, U$ **do**
    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\epsilon = \mathbf{0}$
    $x_{t-1}^{\text{known}} = \sqrt{\bar{\alpha}_t} x_0 + (1 - \bar{\alpha}_t)\epsilon$
    $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
    $x_{t-1}^{\text{unknown}} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t z$
    $x_{t-1} = m \odot x_{t-1}^{\text{known}} + (1 - m) \odot x_{t-1}^{\text{unknown}}$
    if $u < U$ and $t > 1$ **then**
      $x_t \sim \mathcal{N}(\sqrt{1-\beta_{t-1}}x_{t-1}, \beta_{t-1}\mathbf{I})$
    end if
  end for
end for
return x_0

Figure 3: RePaint Algorithm

CONCLUSION & FUTURE WORK

RePaint is a novel conditioning method, with a unified inference schedule that works for *a variety of mask type* and *diffusion models*. It produces semantically coherent completions by harmonizing known and generated regions.

We were able to re-implement and mostly reproduce the CelebA-HQ results of the original paper.

Future Work:
- Conditioning on language instruction and known area
- Image Inpainting → Video Inpainting

References:
[1] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L., 2022. Repaint: Inpainting using denoising diffusion probabilistic models, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11461–11471.
[2] Peng, J., Liu, D., Xu, S., Li, H., 2021. Generating diverse structure for image inpainting with hierarchical vq-vae, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.10775–10784.
[3] Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., Lempitsky, V., 2022. Resolution-robust large mask inpainting with fourier convolutions, in: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 2149–2159.
[4] Wan, Z., Zhang, J., Chen, D., Liao, J., 2021. High-fidelity pluralistic image completion with transformers, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 4692–4701.
[5] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S., 2018. Generative image inpainting with contextual attention, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5505–5514.
[6] Zeng, Y., Fu, J., Chao, H., Guo, B., 2022. Aggregated contextual transformations for high-resolution image inpainting. IEEE transactions on visualization and computer graphics 29, 3266–3280.