RePaint: Inpainting using Denoising Diffusion Probabilistic Models [1]

Minji Kim (minji@snu.ac.kr; 2020-28702), Dept. of Electrical and Computer Engineering, Seoul National University

1. Introduction

Inpainting is a process of reconstructing damaged, deteriorated, or missing parts of images or videos. Early approaches include searches for the most similar patches from the untacted pixels of the image itself, and then directly pastes the patches on the missing parts. On the other hand, there were GAN-based approaches such as Context Encoder, GLCIC. The problem of previous inpainting models is that 1) they are not generalizable to unseen mask types, and 2) training with pixel-wise and perceptual loss often leads to naive textural extensions instead of semantically meaningful generation. To cope with this limitation, the authors propose **RePaint**, which conditions the generation process by sampling from the given pixels during the reverse diffusion iterations, instead of learning a mask-conditional generative model. As shown in Fig. 1, Denoising Diffusion Probabilistic Models (DDPM) is used for inpainting. The process is conditioned on the masked input, and it starts from a random Gaussian noise sample, and than it is iteratively denoised until a high-quality output is generated. As this process is stochastic, multiple diverse outputs can be obtained.

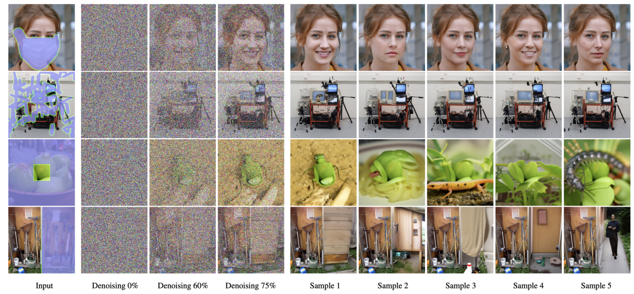

Figure 1. DDPM for inpainting.

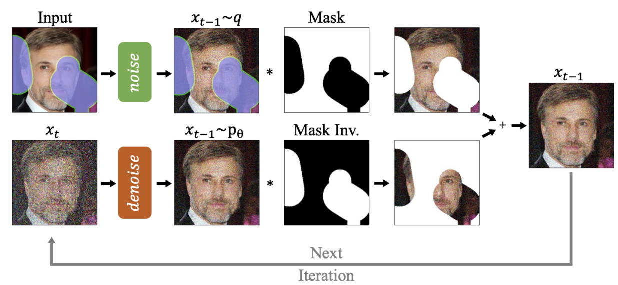

Figure 2. Overview of Repaint.

2. RePaint

2.1. Overview

The overview of Repaint is shown in Fig. 2. RePaint modifies the standard denoising process in order to condition on the given image content. In each step, it samples the known region from the input and the inpainted part from the DDPM output. Specifically, given a ground truth image x, the unknown pixels $m \odot x$ and the known pixels $(1-m) \odot x$, the known regions in diffusion step t, $(1-m) \odot x_t$, could be consistently altered. Thus, in each reverse step, x_{t-1} could be obtained by using following equations:

$$x_{t-1}^{known} \sim \mathcal{N}(\sqrt{\bar{\alpha}_t} x_0, (1 - \bar{\alpha}_t)\mathbf{I}), \tag{1}$$

$$x_{t-1}^{unknown} \sim \mathcal{N}(\mu_\theta(x_t, t), \Sigma_\theta(x_t, t)), \tag{2}$$

$$x_{t-1} = m \odot x_{t-1}^{known} + (1 - m) \odot x_{t-1}^{unknown}. \tag{3}$$

Thus, x_{t-1}^{known} is sampled using the known pixels in the given image $m \odot x_0$, and $x_{t-1}^{unknown}$ is sampled from the model, given the previous iteration x_t is given. Finally, the new sample x_{t-1} is combined using the mask.

2.2. Resampling

Using above algorithm, the inpainted region might mismatches the neighboring regions semantically, because the sampling of the known pixels is performed without considering the generated parts of the image. The model needs more time to harmonize the conditional information x_{t-1}^{known} with the generated information $x_{t-1}^{unknown}$ in one step before going to the next denoising step. Thus, they use the property of DDPM which aims at producing consistent structure, and resample by diffusing the output x_{t-1} back to x_t.

3. Conclusion

This paper has proposed mask-agnostic inpainting model. DDPM is applied during inpainting, while the training is stabilized with resampling strategy. The limitation of this paper is that RePaint is significantly slower than other approaches such as GAN-based or Autoregressive-based methods.

References

[1] A. Lugmayr, et al. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, 2022. 1