DIFFGANPAINT: FAST INPAINTING USING DENOISING DIFFUSION GANS

Anonymous authors
Paper under double-blind review

ABSTRACT

Free-form image inpainting is the task of reconstructing parts of an image specified by an arbitrary binary mask. In this task, it is typically desired to generalize model capabilities to unseen mask types, rather than learning certain mask distributions. Capitalizing on the advances in diffusion models, in this paper, we propose a Denoising Diffusion Probabilistic Model (DDPM) based model capable of filling missing pixels fast as it models the backward diffusion process using the generator of a generative adversarial network (GAN) network to reduce sampling cost in diffusion models. Experiments on general-purpose image inpainting datasets verify that our approach performs superior or on par with most contemporary works.

1 INTRODUCTION

Image Inpainting is the task of reconstructing missing regions in an image. As an inpainting approach requires strong generative capabilities, most of the contemporary works rely on GANs (Zheng et al., 2022) or Autoregressive Modeling (Yu et al., 2021). Capitalizing on the advances in diffusion models, a different line of research is the DDPM-based image synthesis (Meng et al., 2021; Lugmayr et al., 2022). Despite their impressive results, DDPM-based models suffer from computationally expensive sampling procedures. To circumvent this, we propose DiffGANPaint, an inpainting method that leverages trained DDPM, and uses a trained GAN model during the reverse process to generate the inpainted image. Thus, our model is a mask-agnostic approach that allows the network to generalize to any arbitrary mask during inference using the generation capabilities of DDPMs. Our experiments on the diverse datasets demonstrate generalization in inpainting semantically meaningful regions.

2 RELATED WORK

Most Existing literature on inpainting methods follow a standard configuration and use diverse GAN-based structures (Cha & Kim, 2022). Despite remarkable image synthesis performance, in these methods, still, pixel artefacts or colour inconsistency occur in synthesized images during the generation process. In a different direction, (Lugmayr et al., 2022) use image prior and a pre-trained Denoising Diffusion Probabilistic Model for generic inpainting. Similar to this, we propose a novel method which uses a trained GAN in the reverse diffusion process to ameliorate the rapidity and sample quality performance (Elaborated in section 3).

3 METHODOLOGY

Our approach (see Figure 1) is comprised of a diffusion model that denoises an image using the diffusion process, which is then used to prepare the image for inpainting using the generator of a trained GAN model that generates the image. Specifically, the inpainting is performed by first denoising the input image using the diffusion process, then extracting the masked region from the original image, and finally inpainting the masked region using the GAN generator. Therefore, we can concurrently leverage the structure consistency attained by DDPM, and high-quality rapid samples achieved by GAN generator. Our approach is illustrated more in Figure 1.

Figure 2: Visual examples of DiffGANPaint results on CelebA-HQ faces. From left to right, shows the original image, input masked image and result image.

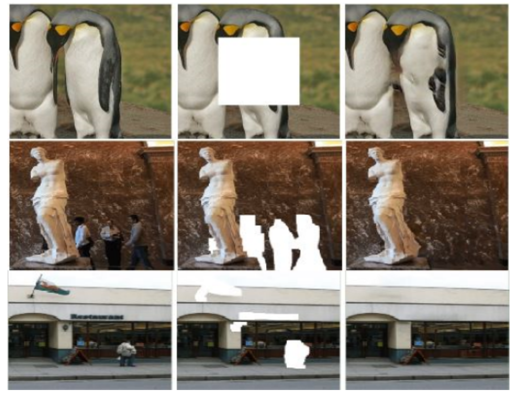

Figure 3: Visual examples of DiffGANPaint results on generic images. From left to right, shows the original image, input masked image and result image.

4 Experiments and Results

We perform experiments for inpainting tasks on generic data and the CelebA-HQ faces datasets [1]. We use the trained guided diffusion and GAN model on Imagenet (Dhariwal & Nichol, 2021). The visual results of the generation process is provided in Figure 2 and 3. As shown, DiffGANPaint produces higher visual quality with a low computational budget. Concretely, our approach can produce samples in fewer steps while trading off the sample quality.

```
timesteps=100
def denoise_diffusion(x, model):
    noise = torch.randn_like(x)
    for i in range(timesteps):
        eps = torch.randn_like(x)
        x = x + torch.sqrt(torch.tensor(2.0)) *
        eps
        model_input = torch.cat([x, noise])
        out = model(model_input)
        x = x + out * np.sqrt(1/timesteps)
    return x

def test_inpainting(img, gan,mask):
    denoised_img = denoise_diffusion(img,gan)
    model_input = torch.cat([denoised_img, mask
                                            ])
    output = gan(model_input)
    output = output * mask + masked_image * (1
                                            - mask)
```

Figure 1: **DiffGANPaint Inpainting Procedure in Pytorch Style:** Pseudo code to generate the inpainted image using the corresponding mask, trained diffusion model and generator.

5 Conclusion

In this paper, we proposed to leverage the benefits of a novel denoising diffusion probabilistic model and GAN model solution for the image inpainting task. Specifically, we exploit a trained diffusion model and modify the reverse diffusion by using a GAN generator to paint images with better mode coverage and sample diversity. Showcased on a variety of datasets, our model demonstrates strong visual capabilities at a low computational cost.

[1]We will make the code publicly available on GitHub after the double-blind review process is complete, to ensure that the anonymity of the authors is not compromised during the review.

References

Dongmin Cha and Daijin Kim. Dam-gan: Image inpainting using dynamic attention map based on fake texture detection. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4883–4887. IEEE, 2022.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11461–11471, 2022.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Yingchen Yu, Fangneng Zhan, Rongliang Wu, Jianxiong Pan, Kaiwen Cui, Shijian Lu, Feiying Ma, Xuansong Xie, and Chunyan Miao. Diverse image inpainting with bidirectional and autoregressive transformers. In *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 69–78, 2021.

Haitian Zheng, Zhe Lin, Jingwan Lu, Scott Cohen, Eli Shechtman, Connelly Barnes, Jianming Zhang, Ning Xu, Sohrab Amirghodsi, and Jiebo Luo. Cm-gan: Image inpainting with cascaded modulation gan and object-aware training. *arXiv preprint arXiv:2203.11947*, 2022.