# Image Inpainting of Portraits Artwork Design and Implementation

*Hongting* Zhang

Computer Science, Zhejiang University of Technology, Hangzhou, Zhejiang, 310023, China

**Abstract.** In modern society, the restoration of artwork has become increasingly important. Generative models can provide reference images for the damaged or blurred core areas of these artworks. This paper simulates artificial damage to classic portrait paintings in the Art Portraits dataset by adding center masks during data preprocessing and then implements the image inpainting task. During the training phase, the Denoising Diffusion Probabilistic Model (DDPM) is fine-tuned by progressively adding noise to the center-masked images in the noising stage, followed by denoising in the denoising stage to generate images. The generated images are compared with the original undamaged images through loss calculations to optimize the model. Additionally, a Generative Adversarial Network (GAN), which has shown promising results on other datasets, is used as a baseline for comparison. The damaged images are used as inputs, and the generated images are compared to the ground truth to evaluate the performance of both models. In the testing phase, two widely used metrics in image evaluation, Mean Squared Error (MSE) and Fréchet Inception Distance (FID), are introduced to assess the performance. The fine-tuned DDPM achieves an MSE of 0.2622 and an FID of 16.85, while the GAN scores 0.2835 and 22.78, respectively. Since lower values indicate higher fidelity in reproducing the original image, which is crucial for art restoration, the conclusion drawn from this paper is that the fine-tuned DDPM demonstrates higher accuracy and is more suitable for restoration projects related to Art Portraits.

## 1 Introduction

The application of generative models for artistic portrait restoration represents an innovative approach that integrates artificial intelligence with traditional art restoration techniques, providing new tools for the preservation and continuation of cultural heritage.

Generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), have achieved significant advancements in the fields of image processing and generation [1]. These models are capable of learning and generating images that closely resemble the style of the original artwork, gradually finding applications in the domains of art and cultural heritage. Generative models can learn and mimic the specific styles of individual artists, ensuring that the restored works retain as much of the original artistic characteristics as possible. This is particularly crucial for historical artworks that hold significant cultural value. Furthermore, one of the advantages of using generative models for

Corresponding author: 202003340227@zjut.edu.cn

restoration is their reversibility, allowing for continual adjustments. The application of this technology fosters cross-disciplinary collaboration between the arts and technology, not only advancing artificial intelligence but also injecting new vitality and creativity into the art world [2].

One of the most significant applications of generative models is their ability to accelerate the restoration process, enabling conservators to obtain multiple restoration proposals in a short period, thereby facilitating quicker and more accurate restoration decisions. This substantially enhances the efficiency of restoration work, particularly when dealing with a large number of damaged artworks. Generative models can learn and mimic the styles of specific artists or art movements, ensuring that the original artistic characteristics are preserved during the restoration process. Additionally, conservators can leverage the extensive range of generated images, combined with their own expertise, to achieve optimal restoration outcomes [3].

## 2 Method

### 2.1 Dataset

In the context of the specific research on artistic portrait restoration, the open-source Art Portraits dataset available on Kaggle presents significant research value, comprising approximately 3.4 GB of data. The dataset includes over 6,000 updated portraits, while the regular portraits folder contains more than 4,000 artistic portraits. This paper utilizes all 4,000 portraits from the portraits folder, applying a center mask to the images before training the image painting model. The training process is conducted using the denoising_diffusion_pytorch library, employing the Mean Squared Error (MSE) loss function and the Adam optimizer. The final model is then tested and validated to assess its performance.

### 2.2 Generative models

Given the specific characteristics of image inpainting, generative models are often among the most suitable choices for accomplishing this type of task. In particular, Generative Adversarial Networks (GANs) and the Denoising Diffusion Probabilistic Model (DDPM) have emerged as two of the most prominent approaches favored in generative tasks due to their notable performance [4]. This paper will conduct an in-depth investigation into the strengths and limitations of several models within these two categories, specifically in the context of image inpainting, to assess their relative effectiveness in restoring occluded image regions.

#### *2.2.1 Preprocessing*

After introducing the Art Portraits dataset, the initial images are first resized to 128x128 resolution and normalized, followed by conversion into tensors. Subsequently, a preprocessing step is applied before training, where a center mask is used to set the central region of the image—specifically the height and width dimensions—to zero. This masking operation facilitates the preparation of the dataset by occluding the central part of the image, thereby creating the necessary conditions for the model to focus on reconstructing these missing areas during training. The Art Portraits dataset is divided into training and testing sets, with a ratio of 0.8 and 0.2, respectively (Fig. 1).

**Fig. 1** The sample of the original image and center masked (impaired) image (Photo/Picture credit: Original).

#### *2.2.2 GAN inversion*

The core idea of GAN inversion is—a process that maps a real image to a latent code within a pre-trained GAN model [5]. GANs are excellent at generating realistic images, and the latent code in the GAN's latent space represents a compressed version of the image. By adding a binary mask to the original image and using a pretrained GAN model (like Creative Adversarial Network-CAN) [6, 7]. During the inversion process, the missing region of the image is optimized to match the surrounding context.

#### *2.2.3 Denoising diffusion probabilistic model (DDPM)*

In the image inpainting tasks using the diffusion model, a linear noise schedule, where noise is gradually added over a fixed number of timesteps [8]. This is crucial for generating noisy images from the original image. Then the diffusion model will apply a traditional Unet structure to calculate the forward diffusion and the reverse diffusion. By doing this, the training process will be able to figure out the loss between the predicted noisy image and the actual noisy image (Fig.2) [9].

This paper utilizes the denoising_diffusion_pytorch library to implement the image inpainting task for artistic portrait restoration. Given that the objective of the model training is to recover occluded regions, partially masked images are generated as input, with the original, unmasked images serving as the target. Subsequently, an L2 loss function is defined to measure the differences between the restored images generated by the model and the original images. The mask is employed to ensure that the model focuses solely on restoring the occluded areas. The model architecture largely follows the conventional diffusion model framework. By calculating the loss in the masked regions and performing gradient updates, a fully trained model is obtained (Fig. 3).

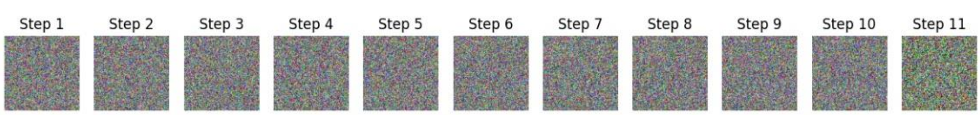


**Fig. 2** The working process of an image in the Diffusion model (Photo/Picture credit: Original).

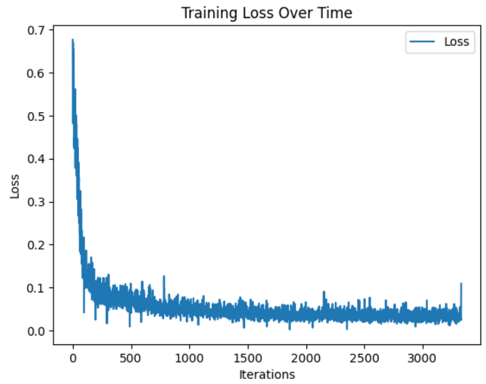


**Fig. 3** The training loss of DDPM (Photo/Picture credit: Original).

### 2.3 Experiments

Pixel-Level Metric-MSE:

The first evaluation metric this paper uses is simply the MSE loss. And it is a typical Pixel-Level metric. By comparing the average value of the generated masked portraits and the original portraits in the test dataset. This paper can determine which model is relatively better [8].

Perceptual-Level Metric-Fréchet Inception Distance (FID):

Besides the MSE metric, this paper also selects another metric called FID which is from another important perceptual level. The FID metric compares the distribution of features extracted from a neural network for the original and inpainted images. Lower FID indicates that the inpainted image is closer to the distribution of real images [10].

## 3 Result

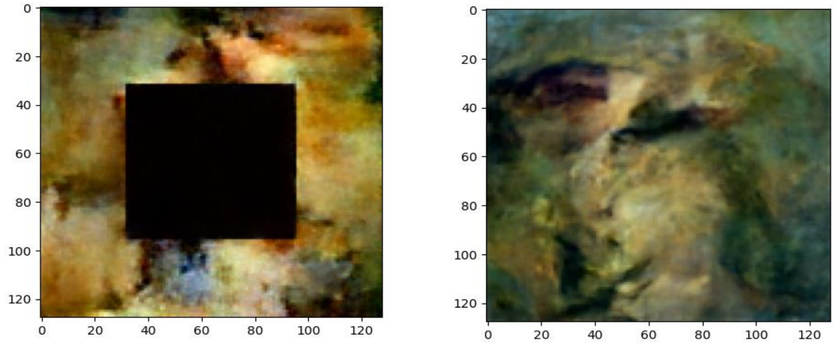

**Fig. 4** 128*128 Image Sample in the early stage of DDPM (Photo/Picture credit: Original).

**Table 1.** The evaluation metrics result on testing set about the SRNet and Fine-tuned DDPM (lower means the performance is relatively better)

| Model | Performance | |
|---|---|---|
| | MSE | FID |
| GAN | 0.2835 | 22.78 |
| DDPM | **0.2622** | **16.85** |

According to this paper's method, the fine-tuned DDPM has a better performance than the general GAN (Table 1).

As shown in Fig 4, the sample images generated in this paper share a similar artistic style with the portraits from the Art Portraits dataset. Additionally, the images were evaluated using specific metrics, confirming their high reliability. These generated images could serve as valuable references for art restorers.

## 4 Conclusion

The fine-tuned DDPM in this paper has several advantages. It demonstrates a high generation accuracy for the specified Art Portraits dataset or similarly styled portrait paintings. The model is lightweight with relatively small parameters, ensuring fast runtime and short execution time which is suitable for personal deployment and implementation. Additionally, it retains the benefits of Diffusion models, allowing for flexibility in adjusting the similarity between the generated portrait and the original artwork by modifying input parameters. If a generated image with a style significantly different from the original is desired, the input parameters can be adjusted accordingly to achieve this.

However, the fine-tuned DDPM model also has some limitations. The number of timesteps is set to 250, rather than 1000 or a larger value, leaving room for improvement in accuracy. Furthermore, the model primarily handles samples with central defects in the portraits, and there is a risk of overfitting as training progresses, especially in the later stages.

Improvements to this model can be made by adjusting the timesteps parameter, allowing the DDPM to extend the noising and denoising processes, making them more detailed and thus achieving higher accuracy.

Additionally, the data preprocessing can be optimized. By randomly masking parts of the image or manually adding masks to key areas, the images can be deliberately altered, making it harder for the model to overfit and enhancing its generalization capabilities.

The limitations of this study lie in the fact that the model achieves high accuracy only with portraits of similar styles. However, in more general image generation tasks, the relatively small model size may lead to suboptimal performance.

By implementing the methods suggested in Section 4.2 to adjust model training, and experimenting with larger, more diverse datasets, a larger model can be trained and further fine-tuned to achieve higher accuracy across a wider range of images. Additionally, developing related software to facilitate user interaction and simplify the image-generation process would be beneficial.

## References

1. R. L., Ruthotto, E. H., An introduction to deep generative modeling, arXiv:2103.05180, (2021).
2. Q. W., Chen, J., Liu, Y., Yan, D. M., Wonka, P., Deep learning-based image and video inpainting: a survey, arXiv:2401.03395 (2024).
3. L. C., Ledig, Photo-realistic single image super-resolution using a generative adversarial network, arXiv:1609.04802. (2017).
4. G. I. J., Goodfellow, Generative adversarial networks, arXiv:1406.2661 (2014).
5. Y. Y., Zhang, L., Fan, H., Luo, T., High-fidelity image inpainting with GAN inversion, arXiv:2208.11850 (2022).

6. E. A., Elgammal, Liu, B., Elhoseiny, M., Mazzone, M., CAN: Creative adversarial networks, generating ‘art’ by learning about styles and deviating from style norms, arXiv:1706.07068. (2017).
7. J. Y., Jiang, X. J., Yang, B., Xu, J., Zhu, J., Image inpainting based on generative adversarial networks, IEEE Access, 1-1, (2020).
8. H. J., Ho, Jain, A., Abbeel, P., Denoising diffusion probabilistic models, arXiv:2006.11239. (2020).
9. L. A., Lugmayr, Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L., RePaint: Inpainting using denoising diffusion probabilistic models, arXiv:2201.09865. (2022).
10. W. L., Wu, Editing text in the wild, arXiv:1908.03047. (2019).