Journal of Image Processing Theory and Applications (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jipta.2025.080106
ISSN 2560-6239 Vol. 8 Num. 1

Image Inpainting for Defective Microscopy Images

Tao Wen[1,2,a,*], Yang Yang[1,2,b,*]

[1]*School of Information Science and Technology, Yunnan Normal University, Kunming, China*
[2]*Laboratory of Pattern Recognition and Artificial Intelligence, Yunnan Normal University, Kunming, China*
[a]*taowen_wt@163.com*, [b]*yyang_ynu@163.com*
**Corresponding author*

Keywords: Image Inpainting, Deep Learning, Diffusion Model

Abstract: Recent advancements in tissue clearing and light-sheet microscopy have transformed whole-brain imaging, enabling cellular-resolution visualization of intact murine brains. However, aggressive clearing protocols and mechanical handling during sample preparation frequently introduce structural defects—such as tissue cracks and regional loss—into microscopy images. These artifacts pose significant challenges for downstream computational analyses, particularly image registration, which depends on structural continuity for precise alignment. Despite the well-documented prevalence of such defects, their impact on registration fidelity remains underexplored, and effective computational solutions for mitigating these challenges are scarce. To bridge this gap, we introduce a mask-free generative framework for digital restoration of damaged neuroimaging data. Unlike existing methods that require labor-intensive manual annotation of defect masks, our approach eliminates mask dependency entirely during both training and inference. By leveraging a diffusion-based architecture with defect-invariant learning, the model autonomously adapts to diverse defect geometries—from fine cracks to large-scale tissue loss—without prior knowledge of corruption patterns. We validate our approach using whole-brain murine microscopy datasets containing real-world artifacts induced by tissue clearing. Quantitative evaluations show that our method not only generates photorealistic restorations of missing structures but also significantly enhances registration accuracy in defective samples.

1. Introduction

Recent advancements in tissue clearing techniques, such as iDISCO+ [1], combined with light-sheet microscopy have enabled high-resolution imaging of intact murine brains at the cellular level. While these innovations support large-scale neuroimaging, adapting computational pipelines originally designed for human brain analysis to murine models presents distinct challenges. A key issue stems from morphological artifacts introduced during tissue processing: Dehydration steps inherent to clearing protocols can cause tissue shrinkage and structural distortions, complicating volumetric analyses of whole-brain microscopy datasets.

Traditional approaches to image restoration often rely on classical inpainting methods to reconstruct missing or corrupted regions. However, the heterogeneous nature of defects in cleared tissues—ranging from micro-cracks to large-scale structural discontinuities—combined with the scarcity of annotated training data, underscores the need for more generalized solutions. Denoising diffusion probabilistic models (DDPMs) offer a scalable alternative but remain limited in biological imaging contexts due to their reliance on manually annotated defect masks, a labor-intensive requirement that impedes high-throughput neuroscience applications.

To overcome these challenges, we introduce Mask-Free Diffusion Imputation (MFDI), a novel blind inpainting framework that eliminates the need for defect masks during both training and inference. Unlike existing diffusion-based methods that require explicit defect segmentation, MFDI autonomously learns defect distributions from unannotated microscopy data while preserving structural fidelity. Comparative evaluations show that MFDI matches the performance of state-of-the-art mask-dependent approaches across various defect types, including irregular shrinkage artifacts and stochastic imaging noise.

2. Related Work

2.1. Image Inpainting

Classical image inpainting frameworks, such as mask-guided partial convolution and two-stage visual consistency networks, pioneered data-driven restoration by training models to reconstruct occluded regions based on synthetically generated masks. While effective in controlled settings, these approaches are inherently limited by their reliance on prior knowledge of defect locations, restricting their applicability to real-world biological imaging scenarios.

The emergence of diffusion models has significantly advanced image restoration. Techniques like RePaint [2] employ spatially guided sampling strategies, applying distinct diffusion processes to masked (defective) and unmasked (intact) regions. Although this approach enables high-fidelity imputation, its dependence on explicit defect segmentation limits its utility for biological imaging, where stochastic morphological defects—such as irregular tissue cracks in cleared mouse brains—lack well-defined masking criteria. Annotating defects across terabyte-scale microscopy datasets remains prohibitively labor-intensive, posing a major scalability challenge.

Recent efforts toward mask-free imputation, exemplified by SDEdit [3], propose an alternative by leveraging intensity discontinuities between defective and intact regions. By introducing noise perturbations into anomalous areas and iteratively denoising through a reverse diffusion process, SDEdit circumvents the need for explicit defect masks. However, its reliance on localized intensity anomalies proves insufficient for neuroimaging applications. The method often generates hallucinated structures in intact regions while failing to fully correct diffuse, low-contrast defects common in cleared tissues. This limitation arises from an inability to effectively distinguish genuine anatomical structures from artifact-induced intensity variations, particularly in high dynamic-range microscopy data.

2.2. Denoising Diffusion Probabilistic Model

Diffusion model falls into the regime of generative models that progressively corrupt data by adding noise through a forward diffusion process, and then learn to reverse this process for generating new data.

Following the notation of denoising diffusion probabilistic model [3] (DDPM) the forward process adds Gaussian noise in a sequence of steps, where the "corrupted" data $x_t(t = 1, \dots, T)$ at t^{th} step is

obtained by applying a small noise to the previous state $x_{t-1}$by a Markov chain with learned Gaussian transitions by:

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \quad (1)$$

where β_t denotes variance at step t.

In the reverse diffusion process, the goal is to recover true data from the noisy data step by step, moving from pure noise back to the original data distribution via a trained neural network (with parameter θ):

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)\right) \quad (2)$$

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha_t}}}\epsilon_\theta(x_t, t)\right) \quad (3)$$

Specifically, ϵ_θ represents the noise predicted by the model, α_t is a noise scheduling parameter, and $\overline{\alpha_t}$ is the cumulative product of α_t over time. By iteratively applying the reverse process, the denoising diffusion model gradually refines a noisy input x_T sampled from a Gaussian distribution to approach the original data distribution x_0.

3. Method

3.1. Network structure

In this study, we chose the classical U-Net with cosine positional encoding as the backbone architecture for our diffusion model, as shown in Figure 1. This selection was driven by several key considerations. First, U-Net has demonstrated exceptional performance across various image generation and restoration tasks, particularly in scenarios that require capturing multi-scale information. Its encoder-decoder structure allows for effective feature extraction at different scales while preserving local details through skip connections, which is essential for reconstructing complex structures. Second, the integration of cosine positional encoding enhances the model's sensitivity to timestep information, thereby improving the progressive generation process of the diffusion model. Additionally, using a fundamental diffusion model enables us to validate the general applicability of our approach. Notably, our method can be seamlessly applied to any denoising diffusion model.

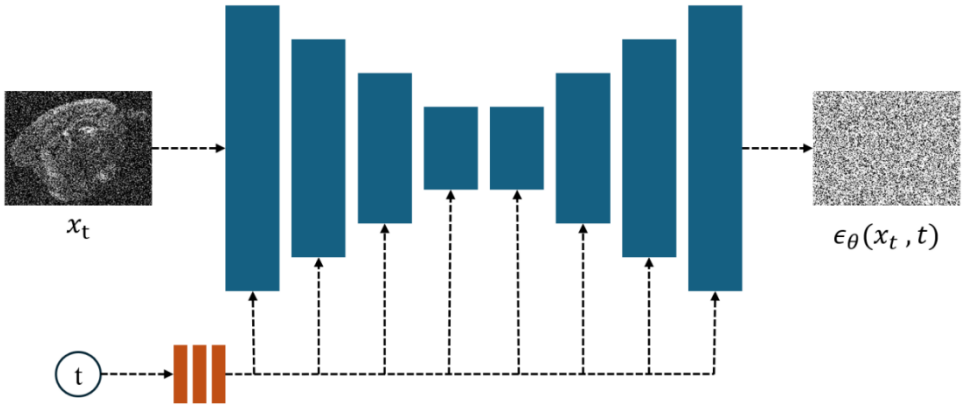

Figure 1: The overall architecture of our proposed method.

3.2. DDPM-based image inpainting

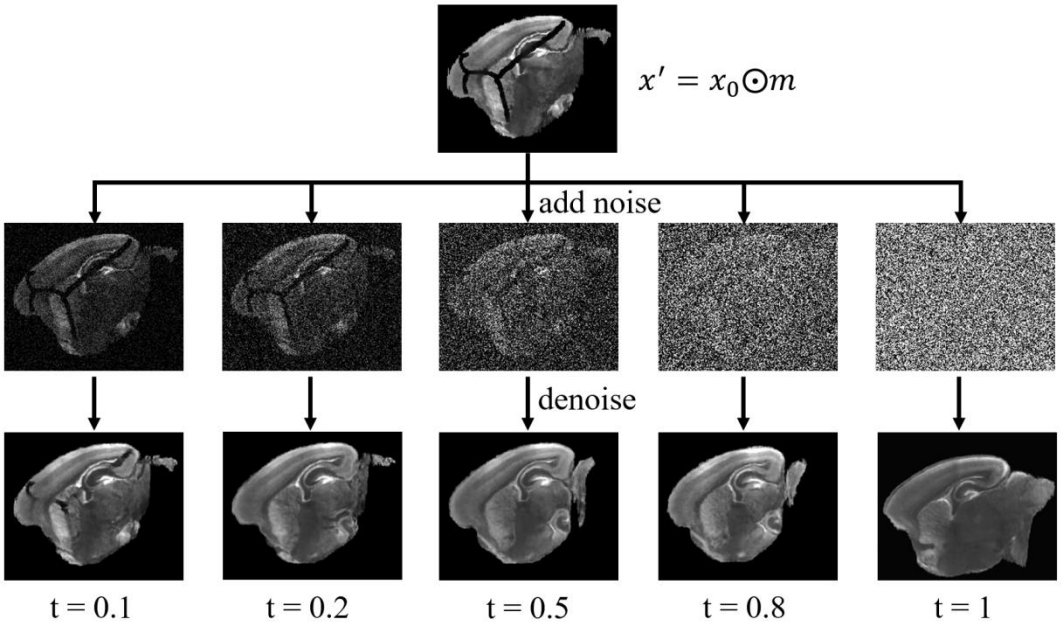

Figure 2: Performance of DDPM-based image inpainting at different time steps.

We denote the original complete image as x_0 and the binary mask as m, where 0 and 1 indicate defective and intact regions, respectively. The masked input image, or the defective image, is obtained via element-wise multiplication $x' = x_0 \odot m$.

Unlike standard diffusion models, which generate images from pure noise, our inpainting process starts from an intermediate step to incorporate known intact regions into the diffusion model. To ensure consistency in noise levels, we introduce noise into the defective image during the forward process, aligning it with the noise distribution of step t in the diffusion model. The formulation is as follows:

$$x'_t = \sqrt{\overline{\alpha_t}}x' + (1 - \overline{\alpha_t})\epsilon \tag{4}$$

where ϵ is randomly sampled from a standard Gaussian distribution, i.e. $\epsilon \sim N(0, I)$. The noisy defective image is then progressively denoised for t steps. Since the diffusion model is trained on complete images, it naturally learns to inpaint missing regions during the denoising process, reconstructing them to resemble the original image.

As illustrated in Figure 2, the choice of timestep t as the starting point for denoising significantly affects the final inpainting quality due to variations in the level of added noise. Experimental observations reveal that as the number of denoising steps increases, the discrepancy between the reconstructed and original images also grows. When t is close to 0, fewer denoising steps are performed, resulting in minimal modification to the intact regions but suboptimal inpainting for the missing areas. Conversely, when t approaches 1, more denoising steps lead to better inpainting results, but at the cost of excessive modification to intact regions. In particular, when $t = 1$ the model effectively reduces to a standard DDPM, generating a complete image from pure noise. However, in the absence of guidance from the binary mask, excessive diffusion steps can cause the model to not only reconstruct missing regions but also alter intact areas—a drawback that is unacceptable in medical image inpainting.

Experimental results indicate that setting $t = 0.2$ achieves an optimal balance, allowing the model to effectively inpaint missing regions while preserving as much of the intact areas as possible. Therefore, we adopt this value in our experiments.

4. Experiment

4.1. Datasets

In this study, we collected a total of 28 microscopy images of mouse brains that had undergone tissue clearing. After normalization, the images were rescaled to a resolution of 128 ×128 ×128. The dataset was split into training and testing sets with an 8:2 ratio, where 22 images were used for training and 6 for testing.

4.2. Implement details

During training, we employed the Adam optimizer along with an Exponential Moving Average (EMA) framework, using a decay factor of 0.995 to ensure training stability and efficiency. The training parameters were set as follows: the batch size was 4, the total number of training iterations (epochs) was 100,000, the learning rate was set to 1e-5, and the number of diffusion time steps was 1000.

4.3. Evaluation Metrics

We evaluate the restoration performance by existing deep models and our method using several popular quantitative measurements in Table 1, which include MAE (mean absolute error), PSNR (peak signal-to-noise ratio), SSIM (structural similarity index measure), FID (Fréchet inception distance), and LPIPS (learned perceptual image patch similarity). Our model achieves the highest score in mask-free methods.

4.4. Quantitative Results

Table 1: Quantitative contrast experiment

Methods	Mask	MAE↓	PSNR↑	SSIM↑	FID↓	LPIPS↓
AOT[4]	Y	0.072	33.419	0.951	14.263	0.055
ICT[5]	Y	0.037	34.474	0.934	24.053	0.063
Repaint[6]	Y	0.059	33.356	0.950	14.235	0.047
VCNet[7]	N	0.075	31.024	0.946	21.532	0.059
SDEdit[2]	N	0.126	29.319	0.899	34.322	0.099
Ours	N	0.063	32.575	0.948	17.954	0.049

As shown in Table 1, our method achieves performance comparable or even higher than deep models, such as AOT, ICT, and Repaint, which leverage mask information during both training and testing. This demonstrates the effectiveness of our approach in utilizing diffusion models for inpainting while maintaining structural consistency and detail preservation. Unlike traditional inpainting models that rely heavily on explicit mask guidance at both training and inference stages, our method leverages the progressive denoising process to naturally reconstruct missing regions without overfitting to predefined patterns. Additionally, our model exhibits strong generalization capabilities across different defect patterns, highlighting its robustness in real-world microscopy image restoration. These results suggest that diffusion-based inpainting offers a promising alternative

to conventional deep-learning-based methods, particularly in scenarios where high-fidelity reconstruction is essential.

5. Conclusion

In this study, we proposed a diffusion-based inpainting framework for real-world defective microscopy images of mouse brains. By incorporating a classical U-Net architecture with cosine positional encoding, our model effectively captures multi-scale features while leveraging known intact regions to guide the inpainting process. Unlike standard diffusion models that generate images from pure noise, our approach initializes from an intermediate timestep, balancing noise consistency and structural preservation. Experimental results demonstrate that selecting an optimal timestep ($t = 0.2$) allows for effective inpainting while minimizing unnecessary modifications to intact regions. Moreover, our method can be seamlessly integrated with any denoising diffusion model, highlighting its generalizability. Future work will explore adaptive timestep selection strategies and the extension of our framework to other biomedical imaging tasks.

References

[1] Renier N, Wu Z, Simon D J, et al. iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging[J]. Cell, 2014, 159(4): 896-910.
[2] Meng C, He Y, Song Y, et al. Sdedit: Guided image synthesis and editing with stochastic differential equations[J]. arXiv preprint arXiv:2108.01073, 2021.
[3] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[J]. Advances in neural information processing systems, 2020, 33: 6840-6851.
[4] Zeng Y, Fu J, Chao H, et al. Aggregated contextual transformations for high-resolution image inpainting[J]. IEEE transactions on visualization and computer graphics, 2022, 29(7): 3266-3280.
[5] Wan Z, Zhang J, Chen D, et al. High-fidelity pluralistic image completion with transformers[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 4692-4701.
[6] Lugmayr A, Danelljan M, Romero A, et al. Repaint: Inpainting using denoising diffusion probabilistic models[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 11461-11471.
[7] Wang Y, Chen Y C, Tao X, et al. Vcnet: A robust approach to blind image inpainting[C]. Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16. Springer International Publishing, 2020: 752-768.