Histology Image Artifact Restoration with Lightweight Transformer Based Diffusion Model

Chong Wang[1], Zhenqi He[2], Junjun He[3], Jin Ye[3], and Yiqing Shen[4(✉)]

[1] School of Biological Science and Medical Engineering, BeiHang University, Beijing, China
[2] Department of Mathematics, Faculty of Science, The University of Hong Kong, Kowloon, Hong Kong SAR
[3] Shanghai AI Laboratory, Shanghai, China
[4] Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA
yshen92@jhu.edu

Abstract. Histology whole slide images (WSIs) are extensively used in tumor diagnosis and treatment planning. However, the presence of artifacts resulting from improper operations during WSI collection can impede both manual and deep learning-based analysis. To bridge this gap, we propose an innovative denoising diffusion model tailored for inpainting artifact-laden regions in histology WSIs. Our method focuses on preserving the intricate morphological structures, which are essential for accurate diagnosis. To ensure the preservation of morphological structures during regional artifact inpainting, we have developed a novel lightweight Transformer-based denoising network, that can capture the correlations between the regional artifact with the global morphological structures. In comparison to existing generative adversarial network (GAN) based solutions, our method minimizes changes in morphology while maximizing preservation of the stain style during the restoration of the artifact. By providing a more reliable and accurate restoration of artifact-affected areas, our model facilitates better analysis and interpretation of histological images, thereby potentially improving the accuracy of tumor diagnosis and treatment decisions. The code is available at https://github.com/zhenqi-he/artifact-restoration.

Keyword: Image Inpainting · Artifacts Restoration · Denoising Diffusion Model

1 Introduction

The analysis of histology whole slide images (WSIs) is crucial for various medical tasks, such as tumor diagnosis, nuclei localization, and treatment planning. However, the integrity of WSIs in histological analyses is frequently jeopardized by a multitude of variables, among which inaccuracies arising during the scanning or collection phases represent a predominant factor [12]. Such procedural deviations can induce significant morphological alterations in tissue structures, which are manifest as artifacts, including but not limited to tissue folds and bubbles [15]. The presence of such artifacts substantially complicates the accurate evaluation and interpretation of histological samples.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Finkelstein et al. (Eds.): AIME 2024, LNAI 14845, pp. 81–89, 2024.
https://doi.org/10.1007/978-3-031-66535-6_9

This complication elevates the likelihood of misdiagnosis among medical professionals and compromises the performance efficacy of Computer-Aided Diagnosis (CAD) systems [17]. Particularly concerning is the detrimental impact on deep learning models, which are increasingly deployed in histological analyses. These computational models are intrinsically vulnerable to the distorting influence of artifacts, thereby engendering potential misclassification or mislabeling of critical tissue structures. This, in turn, exacerbates the risk factors associated with both false positive and false negative diagnoses [19].

In clinical applications, although it is technically possible to rectify the issue of artifacts through re-scanning or re-preparation of the affected tissue slides, such a course of action imposes significant logistical and financial burdens, often requiring multiple iterations to achieve a satisfactory WSI [4]. As for CAD, the strategy of identifying and excluding artifact-laden regions does offer a solution; however, this tactic inevitably results in a forfeiture of valuable contextual information, thereby undermining the overall reliability of the diagnostic model [2]. Existing techniques for mitigating various types of artifacts, including but not limited to folds [11], ink markings [1], and out-of-focus [8] zones, have been introduced; yet, these methods remain inadequate in restoring occluded regions. More recently, approaches leveraging generative networks have been proposed for the remediation of artifact-afflicted areas [10]. These methods, however, suffer from inherent challenges related to mode collapse, complicating their training phases. Additionally, the propensity of generative-based approaches to induce alterations in structure or coloration detracts from the precision of the restored regions.

To ameliorate these challenges, we propose a novel Denoising Diffusion Probabilistic Model (DDPM) [7] with lightweight transformer as the denoising network, specifically engineered for the inpainting of artifact-laden regions within histological WSIs. Diverging significantly from the generative-based paradigms, the proposed methodology obviates the need for artifact-laden images during training, substantially facilitating the data acquisition process and streamlining the training pipeline. The key contributions of this study can be summarized as follows: (1) The introduction of an innovative, lightweight Transformer-based denoising network, purposed for the nuanced capturing of correlations between regional artifacts and global morphological structures. (2) The implementation of a conditional framework during the restoration process, selectively targeting artifact regions while scrupulously preserving the morphological and staining features of artifact-free zones.

2 Related Work

Artifact in Histology. The challenge of mitigating artifacts in medical image analysis is both enduring and complex, particularly for the field of histopathology. The presence of artifacts can significantly compromise the efficacy of computer-aided diagnosis systems, often leading deep learning algorithms to misclassify artifact-ridden areas as tumors. Artifacts are typically spawned by a variety of factors, such as equipment specifications, data acquisition protocols, and environmental conditions [12]. These intrusions not only degrade the overall quality of WSIs but also introduce errors in subsequent computational analyses. To address this, current practices often involve manually omitting images

containing artifacts during the curation of histopathological datasets [6]. While this ensures data reliability, this procedure also poses a dilemma, namely the exclusion of these images leads to a loss of potentially valuable information, thereby generating a problem of dataset incompleteness.

Generative Model for Artifact Restoration. Generative models have shown their promise in ameliorating histological artifacts [16]. For example, CycleGAN has been adapted to tackle this specific problem [20]. However, existing generative techniques are not without their limitations. One significant drawback is the inadvertent alteration of staining styles, resulting in unnatural color variations across the histological images. Although recent advancements have mitigated some of these issues, a case in point being a refined CycleGAN model AR-CycleGAN that employs artifact and pseudo-color cycles [10], GAN-based approaches still face challenges. In particular, they can unintentionally introduce morphological changes in regions that are already artifact-free, putting at risk the integrity of critical tissue structure information.

3 Methods

The proposed method for denoising and restoring artifacts in medical images consists of two stages: training on artifact-free images and artifact restoration (Algorithm 1). During the training stage, the distribution of artifact-free regions is learned through a denoising diffusion model. In the inference stage, artifact regions are specifically targeted for restoration by substituting them with Gaussian noise and leveraging the contextual regions of artifact-free areas for recovery.

3.1 Diffusion Training Stage

In the training phase, local tissue structures are learned from artifact-free histological images for subsequent artifact restoration. Following the standard formulation of DDPM [7], it comprises a forward process and a reverse process. In the forward process, noises are gradually added until the structure of the image is fully destroyed, and the image at each timestep t can be determined by the states at the previous timestep $t-1$ through a Gaussian perturbation $q(\mathbf{x}_t|\mathbf{x}_{t-1}) = N\left(x_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}\right)$ parameterized by β_t. On the contrary, during the reverse phase, a denoising neural network is trained to approximate the conditional distribution $p(\mathbf{x}_{t-1}|\mathbf{x}_t)$, effectively reversing the effects of the noise-inducing operation $q(\mathbf{x}_t|\mathbf{x}_{t-1})$. The objective is to iteratively reconstruct the original data distribution from an initial condition of white noise ($\mathbf{x}_T$). Each reverse step, guided by the learned denoising function, refines the generated sample by removing noise and enhancing the essential features of the underlying data structure. Through this iterative process, the network essentially reverses the diffusion procedure, leading to a high-fidelity reconstruction of the original image or data point. Exemplary samples generated are demonstrated in Fig. 1.

Denoising diffusion methods predicated on UNet architecture are typically characterized by a substantial number of model parameters, resulting in protracted inference times. Inspired by SegFormer [18], to optimize efficiency for local deployment, a

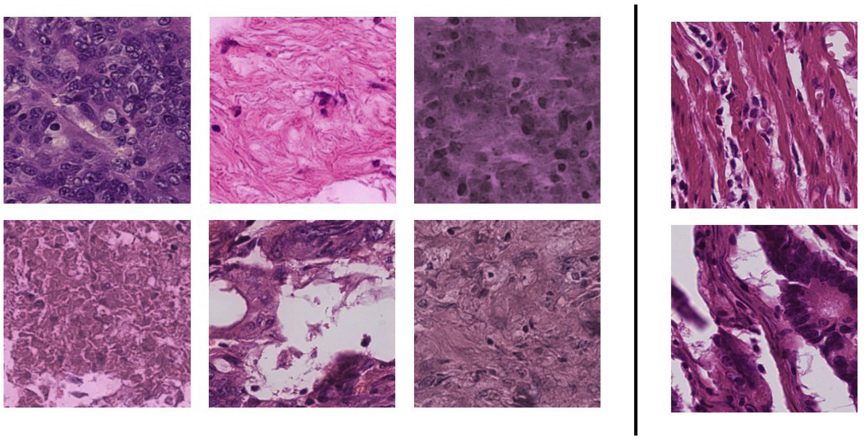

Fig. 1. The generated histological images by the diffusion model (left) and real-world histology data (right).

denoising network leveraging a lightweight Transformer architecture was engineered. Notably, relative to convolutional networks, the Transformer-based architecture demonstrates superior efficacy in capturing the intricate relationships between regional artifacts and global morphological structures (Fig. 2).

3.2 Artifact Restoration

In the artifact restoration stage, we utilize the diffusion model to restore the artifact region from an inpainting perspective (Fig. 3). Firstly, a threshold-based segmentation method is first used to detect the artifact regions in the input image $\mathbf{x}_0$. We then use this information to condition the denoising process only to the artifact regions, while maintaining the artifact-free regions to preserve the morphological structure and stain style. Formally, we write the artifact region as $\mathbf{x}_0 \odot \mathbf{m}$ and the artifact-free region as $\mathbf{x}_0 \odot (1 - \mathbf{m})$, where $\mathbf{m}$ is a boolean mask generated in the threshold method, and $\odot$ is the pixel-wise multiplication operator. It follows that the selective denoising process to inpaint the artifact region is formulated as $\mathbf{x}_t^{in} = \mathbf{x}_t^{sample} \odot (1 - \mathbf{m}) + \mathbf{x}_{t+1}^{out} \odot \mathbf{m}$, where $\mathbf{x}_t^{sample}$ follows the Gaussian perturbation in the forward process, and $\mathbf{x}_{t+1}^{out}$ is the output from the denoising network in the previous reverse step. We employ the Denoising Diffusion Implicit Model (DDIM) [13] to accelerate the restoration process.

Algorithm 1 Overall procedure for artifact restoration

Require: Original image x_0

Ensure: Denoised image x_T^{out}

▷ *Non-Artifact Image Training Phase*

Train Denoising Diffusion Model to learn $p(x)$

▷ *Inference Phase*

Identify artifact regions using threshold-based method

$m \leftarrow$ Boolean mask of artifact regions

▷ *Artifact Restoration Phase*

for each timestep $t = 0, 1, \ldots, T$ **do**

  if t is forward timestep **then**

    $x_t \leftarrow \text{ApplyNoise}(x_{t-1}, \beta_t)$

  else

    $x_t \leftarrow \text{Denoise}(x_{t+1}, m)$

  end if

  $x_t^{in} \leftarrow x_t^{sample} \odot (1 - m) + x_{t+1}^{out} \odot m$

end for

return x_T^{out}

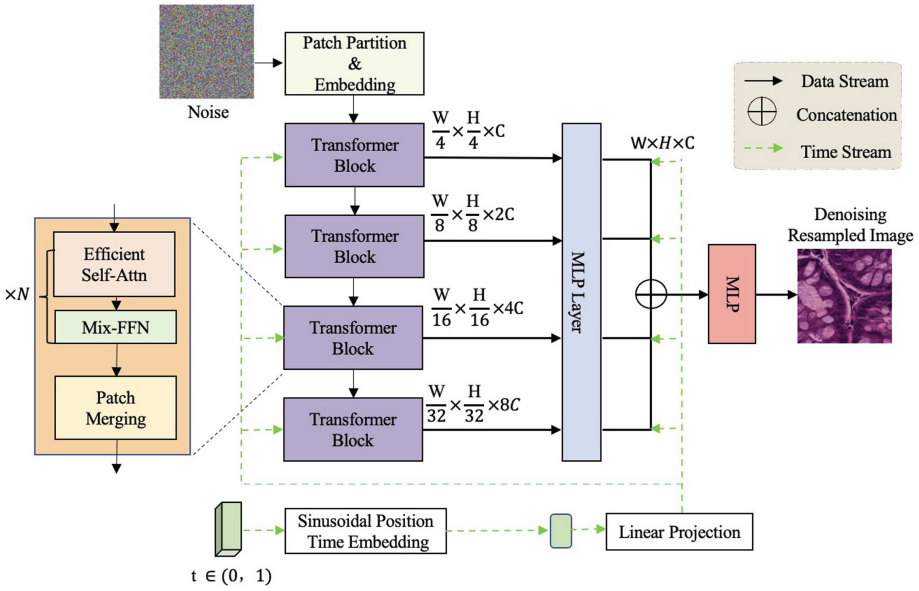

Fig. 2. The proposed lightweight Transformer-based denoising network. FFN, Feedforward Network; MLP, Multilayer Perceptron.

4 Experiments

4.1 Datasets and Settings

The efficacy of the proposed framework for artifact restoration has been rigorously evaluated in the context of both artifact removal and downstream tissue structure classification tasks. For the purpose of ensuring the reliability of the findings, the publicly accessible dataset, NCT-CRC-HE-100K [9], was employed for the training phase, while validation was conducted using the CRC-VAL-HE-7K dataset. We regard the images from the original dataset as clean and proceed to artificially synthesize Artifact images. Owing to the scarcity of comparable studies and open-source codes in the existing literature, CycleGAN [20] was chosen as the benchmark for comparison, and its performance was also evaluated on downstream classification tasks.

4.2 Implementation Details

We implemented the proposed method along with the associated downstream classification tasks using Python 3.8.10 and the PyTorch 1.10.0 library. To accelerate computations and facilitate parallelism, all experiments were run on a dual-GPU setup, specifically using two NVIDIA RTX A4000 graphics cards, each outfitted with 16 GiB of memory. Within the context of our denoising diffusion model, we adhered to a specific set of hyperparameters to optimize performance. We employed the Adam optimizer with a learning rate of the 1×10^4, and designated a total of 250 time steps for the artifact restoration for computational efficiency.

4.3 Evaluation of Artifact Restoration

In the evaluation of artifact restoration, one crucial aspect that must not be overlooked in the realm of medical imaging is the preservation of tissue structure and staining variations within artifact-free regions. As evidenced by the Fig. 4, the proposed method exhibits significant efficacy in retaining the structural integrity of the image. These results suggest that by strategically utilizing the distribution of artifact-free regions as a guide for the restoration process within artifact-afflicted areas, it becomes feasible to accurately replicate the tissue structure and staining characteristics inherent to artifact-free regions. This in turn culminates in a more authentic restoration of tissue images, thereby enhancing their clinical relevance.

4.4 Comparisons of Model Complexity

A noteworthy feature of our proposed method lies in its efficiency in terms of parameter count. Relative to other extant models, our approach operates with a substantially reduced number of parameters (Table 1). Specifically, when contrasted with UNet-based denoising models [7], our model manifests a 91.1% reduction in the number of parameters, diminishing from 1.081×10^8 to 9.62×10^6. This reduction in parameter complexity translates to accelerated inference times, which have been curtailed from 112.37 s to a mere 24.64 s. Such an improvement substantially augments the suitability of our model for real-world clinical deployment.

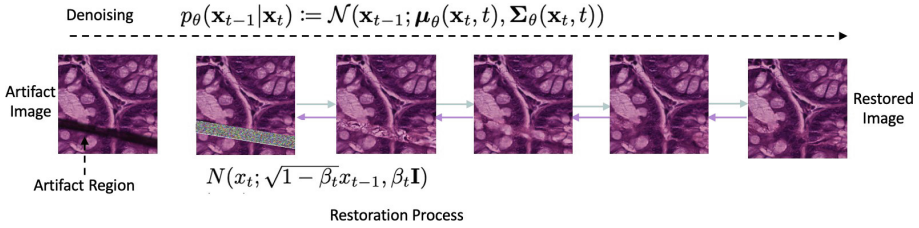

Fig. 3. The proposed artifact restoration pipeline.

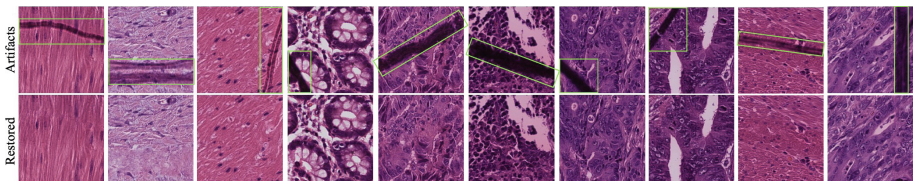

Fig. 4. Illustration of histology images with artifact, where the artifact region is highlighted by green rectangles (□); and restored images by our proposed denoising diffusion model.

4.5 Evaluation on Downstream Classification Tasks

To evaluate the effectiveness of the proposed artifact restoration framework, an assessment was conducted on downstream tissue structure classification tasks. Two baselines, denoted as 'Clean' and 'Artifacts', served as comparative standards. 'Clean' represents artifact-free images, while 'Artifacts' encapsulates images afflicted by artifacts. These baselines are critical for contextualizing the improvements furnished by our model. In Table 2, a comprehensive assessment against multiple classification benchmarks revealed that our method outperforms the CycleGAN approach across all metrics. Specifically, it was observed that the presence of artifacts led to a significant degradation in classification accuracy, with an average decline of 8.82%. However, the artifact removal augmented the classification accuracy of afflicted images. For instance, on the ResNet50 architecture, compared to the 'Artifacts' baseline, our method enhanced the accuracy by 7.13%, while CycleGAN improved it by 2.98%. Compared to the CycleGAN benchmark, our approach exhibited superior accuracy with an improvement of 3.54%, indicating enhanced capabilities for artifact restoration.

Table 1. Model complexity comparison in terms of the number of parameters and averaged inference time.

Methods	#Params ($\times 10^6$)	Time (s)
CycleGAN	28.28	1.065
DDPM (UNet)	108.41	112.37
Ours	9.62	24.64

Table 2. The effectiveness of the proposed artifact restoration framework on the downstream task-tissue classification task. We report the classification accuracy on the test set (%) with different network architectures of ResNet [5], RexNet [3], and EfficientNet [14].

Method	ResNet18	ResNet50	RexNet100	EfficientNetB0
Clean	95.529	94.833	95.487	95.808
Artifacts	80.302	85.012	90.446	90.626
CycleGAN	86.326	87.994	90.776	91.811
Ours	91.676	92.144	93.786	93.449

5 Conclusions

One salient finding was the marked degradation in classification accuracy in the presence of image artifacts—a challenge our method adeptly mitigates. Beyond merely enhancing accuracy rates, our framework sets itself apart through its exceptional efficiency in utilizing parameters. This efficiency not only showcases the robustness of our approach but also its practical viability in clinical environments. In such settings, the dual requirements of high accuracy and computational economy are critical, and our framework meets these demands admirably. Moreover, it is important to acknowledge that while our model does incur longer inference times, especially when compared to alternatives like CycleGAN, this outcome is a direct result of its intricate sequential diffusion steps. These steps, though time-consuming, are integral to the model's ability to accurately process and improve upon the input images. Future work will explore the application of distillation techniques to reduce the number of inference steps, thereby aiming to optimize the duration of the inference phase.

References

1. Ali, S., Alham, N.K., Verrill, C., Rittscher, J.: Ink removal from histopathology whole slide images by combining classification, detection and image generation models. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 928–932. IEEE (2019)
2. Bautista, P.A., Yagi, Y.: Detection of tissue folds in whole slide images. In: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3669–3672. IEEE (2009)
3. Han, D., Yun, S., Heo, B., Yoo, Y.: RexNet: diminishing representational bottleneck on convolutional neural network (2020)
4. Han, L., Su, H., Yin, Z.: Phase contrast image restoration by formulating its imaging principle and reversing the formulation with deep neural networks. IEEE Trans. Med. Imaging **42**(4), 1068–1082 (2022)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR abs/1512.03385 (2015). http://arxiv.org/abs/1512.03385
6. He, Z., He, J., Ye, J., Shen, Y.: Artifact restoration in histology images with diffusion probabilistic models. arXiv preprint arXiv:2307.14262 (2023)

7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. CoRR abs/2006.11239 (2020). https://arxiv.org/abs/2006.11239
8. Hosseini, M.S., Brawley-Hayes, J.A., Zhang, Y., Chan, L., Plataniotis, K.N., Damaskinos, S.: Focus quality assessment of high-throughput whole slide imaging in digital pathology. IEEE Trans. Med. Imaging **39**(1), 62–74 (2019)
9. Kather, J.N., et al.: Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLOS MED **16**(1), e1002730 (2019)
10. Ke, J., et al.: Artifact detection and restoration in histology images with stain-style and structural preservation. IEEE Trans. Med. Imaging **42**, 3487–3500 (2023)
11. Kothari, S., Phan, J.H., Wang, M.D.: Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade. J. Pathol. Inform. **4**(1), 22 (2013)
12. Seoane, J., Varela-Centelles, P., Ramírez, J., Cameselle-Teijeiro, J., Romero, M.: Artefacts in oral incisional biopsies in general dental practice: a pathology audit. Oral Dis. **10**(2), 113–117 (2004)
13. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv:2010.02502, October 2020. https://arxiv.org/abs/2010.02502
14. Tan, M., Le, Q.: EfficientNet: rethinking model scaling for convolutional neural networks (2020)
15. Taqi, S.A., Sami, S.A., Sami, L.B., Zaki, S.A.: A review of artifacts in histopathology. J. Oral Maxillofacial Pathol. JOMFP **22**(2), 279 (2018)
16. Venkatesh, B., Shaht, T., Chen, A., Ghafurian, S.: Restoration of marker occluded hematoxylin and eosin stained whole slide histology images using generative adversarial networks. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 591–595. IEEE (2020)
17. Wang, N.C., Kaplan, J., Lee, J., Hodgin, J., Udager, A., Rao, A.: Stress testing pathology models with generated artifacts. J. Pathol. Inform. **12**(1), 54 (2021)
18. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: simple and efficient design for semantic segmentation with transformers. In: Neural Information Processing Systems (NeurIPS) (2021)
19. Zhang, Y., Sun, Y., Li, H., Zheng, S., Zhu, C., Yang, L.: Benchmarking the robustness of deep neural networks to common corruptions in digital pathology. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 13432, pp. 242–252. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16434-7_24
20. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV (2017). https://doi.org/10.1109/iccv.2017.244