Artifact Restoration in Histology Images with Diffusion Probabilistic Models

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

[1] The University of Hong Kong, Pokfulam, Hong Kong
[2] Shanghai AI Laboratory, Shanghai, China
[3] Johns Hopkins University, Baltimore, USA
yshen92@jhu.edu

Abstract. Histological whole slide images (WSIs) can be usually compromised by artifacts, such as tissue folding and bubbles, which will increase the examination difficulty for both pathologists and Computer-Aided Diagnosis (CAD) systems. Existing approaches to restoring artifact images are confined to Generative Adversarial Networks (GANs), where the restoration process is formulated as an image-to-image transfer. Those methods are prone to suffer from mode collapse and unexpected mistransfer in the stain style, leading to unsatisfied and unrealistic restored images. Innovatively, we make the first attempt at a denoising diffusion probabilistic model for histological artifact restoration, namely `ArtiFusion`. Specifically, `ArtiFusion` formulates the artifact region restoration as a gradual denoising process, and its training relies solely on artifact-free images to simplify the training complexity. Furthermore, to capture local-global correlations in the regional artifact restoration, a novel Swin-Transformer denoising architecture is designed, along with a time token scheme. Our extensive evaluations demonstrate the effectiveness of `ArtiFusion` as a pre-processing method for histology analysis, which can successfully preserve the tissue structures and stain style in artifact-free regions during the restoration. Code is available at https://github.com/zhenqi-he/ArtiFusion.

Keywords: Histological Artifact Restoration · Diffusion Probabilistic Model · Swin-Transformer Denoising Network

1 Introduction

Histology is critical for accurately diagnosing all cancers in modern medical imaging analysis. However, the complex scanning procedure for histological whole-slide images (WSIs) digitization may result in the alteration of tissue structures, due to improper removal, fixation, tissue processing, embedding, and storage [11]. Typically, these changes in tissue details can be caused by various extraneous factors such as bubbles, tissue folds, uneven illumination, pen marks, altered staining, and *etc* [13]. Formally, the changes in tissue structures are known as artifacts. The presence of artifacts not only makes the analysis more challenging

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 518–527, 2023.
https://doi.org/10.1007/978-3-031-43987-2_50

for pathologists but also increases the risk of misdiagnosis for Computer-Aided Diagnosis (CAD) systems [14]. Particularly, deep learning models, which have become increasingly prevalent in histology analysis, have shown vulnerability to the artifact, resulting in a two-times increase in diagnosis errors [18].

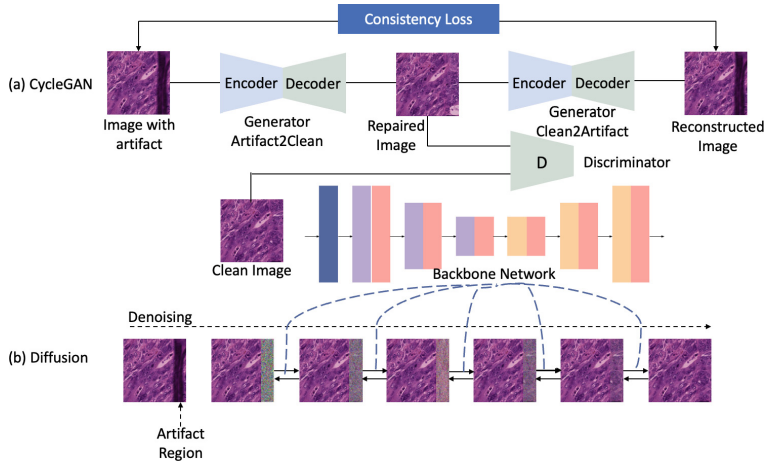

Fig. 1. Learning-based artifact restoration approaches. (a) CycleGAN [19] formulates the artifact restoration as an image-to-image transfer problem. It leverages two pairs of the generator and discriminator to learn the transfer between the artifact and artifact-free image domains. (b) Diffusion probabilistic model [5] (ours) formulates artifact restoration as a regional denoising process.

In real clinical practice, rescanning the WSIs that contain artifacts can partially address this issue. However, it may require multiple attempts before obtaining a satisfactory WSI, which can lead to a waste of time, medical resources, and deplete tissue samples. Discarding the local region with artifacts for deep learning models is another solution, but it may result in the loss of critical contextual information. Therefore, learning-based artifact restoration approaches have gained increasing attention. For example, CycleGAN [19] formulates the artifact restoration as an image-to-image transfer problem by learning the transfer between the artifact and artifact-free image domains from unpaired images, as depicted in Fig. 1(a). However, existing artifact restoration solutions are confined to Generative Adversarial Networks (GANs) [2], which are difficult to train due to the mode collapse and are prone to suffer from unexpected stain style mistransfer. To address these issues, we make the first attempt at a diffusion probabilistic model for artifact restoration approach [5], as shown in Fig. 1(b). Innovatively, our framework formulates the artifact restoration as a regional denoising process, which thus can to the most extent preserve the stain style and avoid the loss of contextual information in the non-artifact region. Furthermore, our approach is trained solely with artifact-free images, which reduces the difficulty in data collection.

The major contributions are two-fold. (1) We make the first attempt at a denoising diffusion probabilistic model for artifact removal, called `ArtiFusion`. This approach differs from GAN-based methods that require either paired or unpaired artifacts and artifact-free images, as our `ArtiFusion` relies solely on artifact-free images, resulting in a simplified training process. (2) To capture the local-global correlations in the gradual regional artifact restoration process, we innovatively propose a Swin-Transformer denoising architecture to replace the commonly-used U-Net and a time token scheme for optimal Swin-Transformer denoising. Extensive evaluations on real-world histology datasets and downstream tasks demonstrate the superiority of our framework in artifact removal performance, which can generate reliable restored images while preserving the stain style.

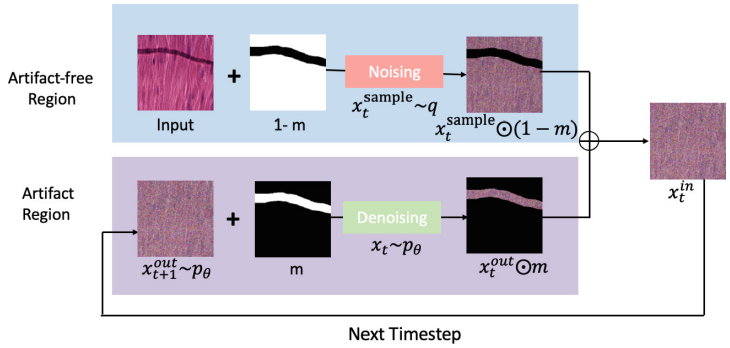

Fig. 2. The semantic illustration of inference stage in `ArtiFusion` for local regional artifact restoration.

2 Methodology

Overall Pipeline. The proposed histology artifact restoration diffusion model `ArtiFusion`, comprises two stages, namely the training, and inference. During the training stage, `ArtiFusion` learns to generate regional histology tissue structures based on the contextual information from artifact-free images. In the inference stage, `ArtiFusion` formulates the artifact restoration as a gradual denoising process. Specifically, it first replaces the artifact regions with Gaussian noise, and then gradually restores them to artifact-free images using the contextual information from nearby regions.

Diffusion Training Stage. The proposed `ArtiFusion` learns the capability of generating local tissue representation from contextual information during the training stage. To achieve this, we follow the formulations of DDPM [5], which involve a forward process that gradually injects Gaussian noise into an

artifact-free image and a reverse process that aims to reconstruct images from noise. During the forward process, we can obtain a noisy version of $\mathbf{x}_t$ for arbitrary timestep $t \in \mathbb{N}[0, T]$ using a Gaussian transition kernel $q(\mathbf{x}t|\mathbf{x}t-1) = \mathcal{N}(x_t; \sqrt{1-\beta_t}\mathbf{x}t-1, \beta_t\mathbf{I})$, where $\beta_t \in (0, 1)$ are predefined hyper-parameters [5]. Simultaneously, the reverse process trains a denoising network $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t^{in})$, which is parameterized by θ, to reverse the forward process $q(\mathbf{x}_t|\mathbf{x}_{t-1})$. The overall training objective L is defined as the variational lower bound of the negative log-likelihood, given by:

$$\mathbb{E}[-\log p_\theta(\mathbf{x}_0)] \leq \mathbb{E}_q[-\log p(\mathbf{x}_T) - \sum_{1\leq t\leq T} \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_t|\mathbf{x}_{t-1})}] = L. \tag{1}$$

This formulation is extended in DDPM [5] to be further written as:

$$L = \mathbb{E}_q[\underbrace{D_{KL}(q(\mathbf{x}_T|x_0))||p(\mathbf{x}_T)}_{L_T} + \sum_{t>1} \underbrace{D_{KL}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0))||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}_{L_{t-1}} - \underbrace{\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}],$$

where $D_{KL}(\cdot||\cdot)$ is the KL divergence.

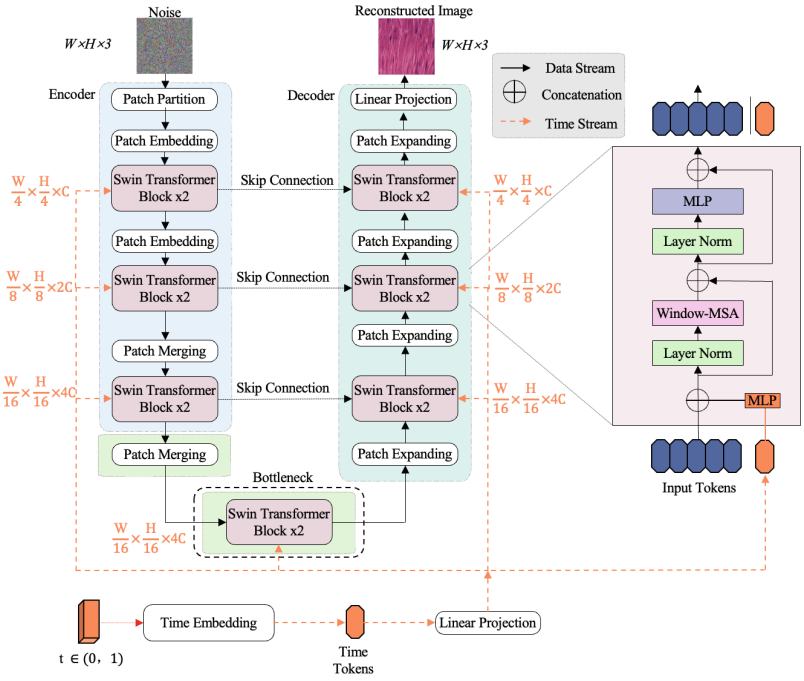

Fig. 3. The proposed Swin-Transformer denoising network.

Artifact Restoration in Inference Stage. During the inference stage, we first use a threshold method to detect the artifact region in the input image $\mathbf{x}_0$. Then, unlike the conventional diffusion models [5] that aim to generate the entire image, `ArtiFusion` selectively performs denoising resampling only in the artifact region to maximally preserve the original morphology and stain style in the artifact-free region, as shown in Fig. 2. Specifically, we represent the artifact-free region and the artifact region in the input image as $\mathbf{x}_0 \odot (1-\mathbf{m})$ and $\mathbf{x}_0 \odot \mathbf{m}$, respectively [10], where $\mathbf{m}$ is a Boolean mask indicating the artifact region and $\odot$ is the pixel-wise multiplication operator. To perform the denoising resampling, we write the input image $\mathbf{x}_t^{in}$ at each reverse step from t to $t-1$ as the sum of the diffused artifact-free region and the denoised artifact region, *i.e.,*

$$\mathbf{x}_t^{in} = \mathbf{x}_t^{sample} \odot (1-\mathbf{m}) + \mathbf{x}_{t+1}^{out} \odot \mathbf{m}, \tag{2}$$

where $\mathbf{x}_t^{sample} o \odot (1-\mathbf{m})$ is artifact-free region diffused for t times using the Gaussian transition kernel *i.e.* $\mathbf{x}_t^{sample} \sim \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t\mathbf{I}))$ with $\bar{\alpha}_t = \prod_{i=1}^{t}(1-\beta_i)$; and $\mathbf{x}_{t+1}^{out}$ is the output from the denoising network in the previous reverse step *i.e.,* $p_\theta(\mathbf{x}_{t+1}^{out}|\mathbf{x}_{t+1}^{in})$. Consequently, the final restored image is obtained as $\mathbf{x}_0 \odot (1-\mathbf{m}) + \mathbf{x}_0^{out} \odot \mathbf{m}$.

Swin-Transformer Denoising Network. To capture the local-global correlation and enable the denoising network to effectively restore the artifact regions, we propose a novel Swin-Transformer-basedr [9] denoising network for `ArtiFusion`. As shown in Fig. 3, our network follows a U-shape architecture, where the encoder, bottleneck, and decoder modules all employ Swin-Transformer as the basic building block. Additionally, we introduce an innovative auxiliary time token to inject the time information. In an arbitrary time step t during the training process, to obtain a time token, we first embed the scalar t by learnable linear layers, with weights that are specific to each Swin-Transformer block. In contrast to existing U-Net based denoising networks [5], we propose a better interaction between hidden features and time information by concatenating the time token to feature tokens before passing them to the attention layers. The resulting tokens are then processed by the attention layers, and the auxiliary time token is discarded to retain the original feature dimension to fit the Swin-Transformer block design after the attention layers.

3 Experiments

Dataset. To evaluate the performance of artifact restoration, a training set is curated from a subset of Camelyon17 [8][1]. It comprises a total number of 2445 artifact-free images and another 2547 images with artifacts, where all histological images are scaled to the resolution of 256×256 pixels at the magnitude of $20\times$. The test set uses another public histology image dataset [6] with 462 artifact-free

[1] Available at https://camelyon17.grand-challenge.org.

images[2], where we obtain the paired artifact images by the manually-synthesized artifacts [18].

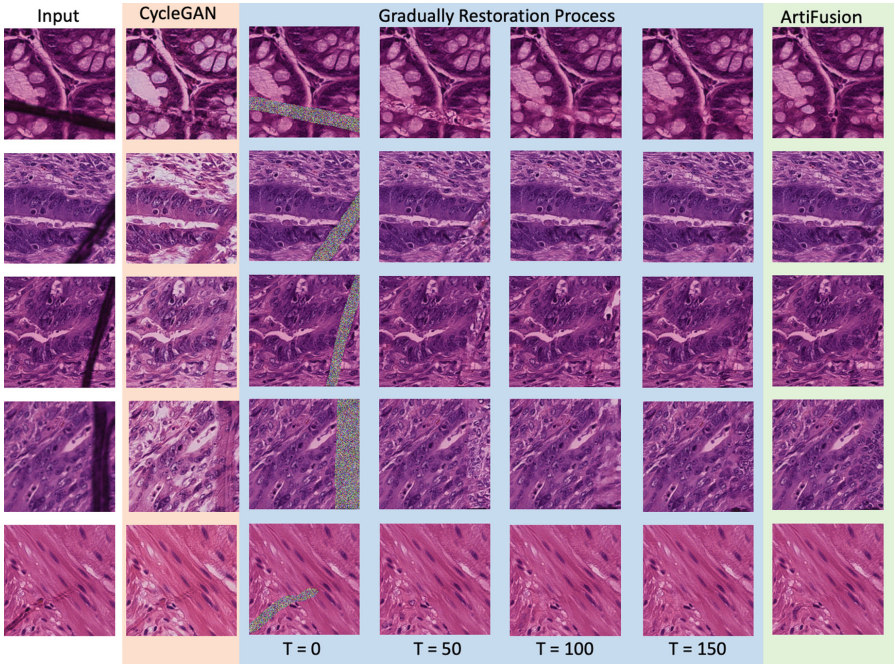

Fig. 4. Artifact restoration on five real-world artifact images. We observe that `ArtiFusion` can successfully overcome the drawback of stain style mistransfer in CycleGAN. We also illustrate the gradual denoising process in the artifact region by `ArtiFusion`, at time step $t = 0, 50, 100, 150$. It highlights the ability of `ArtiFusion` to progressively remove artifacts from the histology image, resulting in a final restored image that is both visually pleasing and scientifically accurate.

Implementations. We implement the proposed `ArtiFusion` and its counterpart in Python 3.8.10 and PyTorch 1.10.0. All experiments are carried out in parallel on two NVIDIA RTX A4000 GPU cards with 16 GiB memory. Hyper-parameters are as follows: a learning rate of 10^{-4} with Adam optimizer, the total timesteps is set to 250.

Compared Methods and Evaluation Metrics. As a proof-of-concept attempt at a generative-models-based artifact restoration framework in the histology domain, currently, there are limited available literature works and open-sourced codes for comparison. Consequently, we leverage the prevalent CycleGAN [19] as the baseline for comparison, because of its excellent performance

[2] Available at https://github.com/lu-yizhou/ClusterSeg.

in the image transfer, and also its nature that requires no paired data can fit our circumstance. Unlike CycleGAN which requires both artifact-free images and artifact images, `ArtiFusion` only relies on artifact-free images, leading to a size of the training set that is half that of CycleGAN. For a fair compaison, we train the CycleGAN with two configurations, namely (#1) using the entire dataset, and (#2) using only half the dataset, where the latter uses the same number of the training samples as `ArtiFusion`. Regarding the ablation, we compare the proposed Swin-Transformer denoising network with the conventional U-Net [5] (denoted as 'U-Net'), and the time token scheme with the direct summation scheme (denoted as 'Add'). We use the following metrics: L_2 distance (L2) with respect to the artifact region, the mean-squared error (MSE) over the whole image, structural similarity index (SSIM) [15], Peak signal-to-noise ratio (PSNR) [1], Feature-based similarity index (FSIM) [17] and Signal to reconstruction error ratio (SRE) [7].

Table 1. Quantitative comparison of `ArtiFusion` with CycleGAN on artifact restoration performance. The ↓ indicates the smaller value, the better performance; and vice versa.

Methods	L2 ($\times 10^4$) ↓	MSE ↓	SSIM ↑	PSNR ↑	FSIM ↑	SRE ↑
CycleGAN (#1) [19]	1.119	0.5583	0.9656	42.37	0.7188	51.42
CycleGAN (#2) [19]	1.893	0.5936	0.9622	42.12	0.7162	50.21
`ArtiFusion` (U-Net)	0.5027	0.2508	0.9850	47.61	0.8173	54.59
`ArtiFusion` (Add)	0.5007	0.2499	0.9850	47.79	0.8184	54.76
`ArtiFusion` (Full Settings)	**0.4940**	**0.2465**	**0.9860**	**48.08**	**0.8216**	**55.43**

Table 2. Comparison of the model complexity and efficiency in terms of the number of parameters, FLOPs, and averaged inference time.

Methods	#Params ($\times 10^6$)	FLOPs ($\times 10^9$)	Inference(s)
CycleGAN [19]	28.28	60.04	1.065
`ArtiFusion` (UNet)	108.41	247.01	112.37
`ArtiFusion` (Add)	27.74	7.69	30.14
`ArtiFusion` (Full Settings)	29.67	7.73	30.71

Evaluations on Artifact Restoration. The quantitative comparison with CycleGAN and `ArtiFusion` are shown in Table 1, where some exemplary images are illustrated in Fig. 4. Our results demonstrate the superiority of `ArtiFusion` over GAN in the context of artifact restoration, with a large margin observed in all evaluation metrics. For instance, `ArtiFusion` can reduce the L2 and MSE by

more than 50%, namely from 1×10^4 to 0.5×10^4 and from 0.55 to 0.25 respectively. It implying that our method can to the large extent restore the artifact regions using the global information. In addition, `ArtiFusion` can improve other metrics, including SSIM, PSNR, FSIM and SRE by 0.0204, 5.72, 0.1028 and 4.02 respectively, indicating that it can preserve the stain style during the restoration process. Moreover, our ablation study shows that the Swin-Transformer denoising network can outperform the conventional U-Net, highlighting the significance of capturing global correlation for local artifact restoration. Finally, the concatenating time token with feature tokens can bring an improvement in terms of all evaluation matrices, making it a better fit for the transformer architecture than the direct summation scheme in U-Net [5]. In summary, our ablations confirm the effectiveness of all the components in our method.

Table 3. The effectiveness of the proposed artifact restoration framework in the downstream task-tissue classification task. We report the classification accuracy on the test set (%) with different network architectures including ResNet [4], RexNet [3] and EfficientNet [12].

Settings	ResNet18	ResNet34	ResNet50	RexNet100	EfficientNetB0
Clean	95.529	93.538	94.833	95.487	95.808
Artifacts	80.302	86.031	85.012	90.446	90.626
Restored w CycleGAN	86.326	88.273	87.994	90.776	91.811
Restored w `ArtiFusion`	92.376	91.252	90.408	92.310	94.232

Comparisons of Model Complexity. In Table 2, we compare the model complexity in terms of the number of parameters, Floating Point Operations Per second (FLOPs), and averaged inference time on one image. Our proposed model achieves a significant reduction in the number of parameters by 72.6%, namely from 108.41×10^6 to 29.67×10^6, compared with CycleGAN. This reduction in model size comes at the cost of longer inference time. However, a smaller model size can facilitate easier deployment in real clinical practice.

Evaluations by Downstream Classification Task. We further evaluate the proposed artifact restoration framework on a downstream tissue classification task. To this end, we use the public dataset NCT-CRC-HE-100K for training and CRC-VAL-HE-7K for testing, which together contains $100,000$ training samples and $7,180$ test samples. We consider the performance on the original unprocessed data, denoted as 'Clean', as the upper bound. Then, we manually synthesize the artifact (denoted as 'Artifact') and evaluate the classification performance with restoration approaches CycleGAN and `ArtiFusion`. In Table 3, comparisons show that the presence of artifacts can result in a significant performance decline of over 5% across all five network architectures. Importantly, the classification accuracy on images restored with `ArtiFusion` is consistently

higher than those restored with CycleGAN, demonstrating the superiority of our model. These results highlight the effectiveness of `ArtiFusion` as a practical pre-processing method for histology analysis.

4 Conclusion

In this paper, we propose `ArtiFusion`, the first attempt at a diffusion-based artifact restoration framework for histology images. With a novel Swin-Transformer denoising backbone, `ArtiFusion` is able to restore regional artifacts using the context information, while preserving the tissue structures in artifact-free regions as well as the stain style. Experimental results on a public histological dataset demonstrate the superiority of our proposed method over the state-of-the-art GAN counterpart. Consequently, we believe that our proposed method has the potential to benefit the medical community by enabling more accurate diagnosis or treatment planning as a pre-processing method for histology analysis. Future work includes investigating the extension of `ArtiFusion` to more advanced diffusion models such as score-based or score-SDE models [16].

References

1. Osama, S., et al.: A comprehensive survey analysis for present solutions of medical image fusion and future directions. IEEE Access **9**, 11358–11371 (2021)
2. Ian, J., et al.: Generative adversarial networks (2014)
3. Han, D., Yun, S., Heo, B., Yoo, Y.: Rexnet: Diminishing representational bottleneck on convolutional neural network (2020)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR, abs/1512.03385 (2015)
5. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. CoRR, abs/2006.11239 (2020)
6. Ke, J., et al.: ClusterSeg: a crowd cluster pinpointed nucleus segmentation framework with cross-modality datasets. Med. Image Anal. **85**, 102758 (2023)
7. Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., Schindler, K.: Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network. ISPRS J. Photogrammetry Remote Sens. **146**, 305–319 (2018)
8. Litjens, G., et al.: 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. GigaScience **7**(6), giy065 (2018)
9. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. CoRR, abs/2103.14030 (2021)
10. Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L.: Repaint: inpainting using denoising diffusion probabilistic models. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
11. Seoane, J., Varela-Centelles, P.I., Ramírez, J.R., Cameselle-Teijeiro, J., Romero, M.A.: Artefacts in oral incisional biopsies in general dental practice: a pathology audit. Oral Dis. **10**(2), 113–117 (2004)
12. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks (2020)

13. 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)
14. Wang, N.C., Kaplan, J., Lee, J., Hodgin, J., Udager, A., Rao, A.: Stress testing pathology models with generated artifacts. J. Pathol. Inf. **12**(1), 4 (2021)
15. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
16. Yang, L., et al.: Diffusion models: A comprehensive survey of methods and applications. arXiv preprint arXiv:2209.00796 (2022)
17. Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: a feature similarity index for image quality assessment. IEEE Trans. Image Process. **20**(8), 2378–2386 (2011)
18. 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.) MICCAI 2022. LNCS, vol. 13432, pp. 242–252. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16434-7_24
19. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.:. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV) (2017)