Blur Synthesis for Artifact Segmentation in Digital Pathology

Surya Achanta[*1,2] orcid.png SURYA.ACHANTA@AIRAMATRIX.COM
Yashasvi Singh[*1] YASHASVI.SINGH@AIRAMATRIX.COM
Nilanjan Chattopadhyay[2] NILANJAN.CHATTOPADHYAY@AIRAMATRIX.COM
Nitin Singhal[†3] NITIN.SINGHAL@AIRAMATRIX.COM

Editors: Under Review for MIDL 2026

Abstract

Histopathology image datasets frequently suffer from artifacts such as out-of-focus blur due to inconsistent microscopy procedures, significantly compromising the reliability of downstream analysis. While identifying and segmenting these artifacts is essential for robust quality control, the development of supervised detection models is hindered by the scarcity of pixel-level annotations for blurred regions. To address this data limitation, we introduce a novel framework for synthesizing realistic blur artifacts utilizing deep generative modeling. Unlike conventional statistical approaches, which fail to capture the stochastic and spatially variant nature of optical aberrations, our approach leverages Conditional Generative Adversarial Networks (**cGANs**) (Mirza et al., 2014) and Conditional Denoising Diffusion Probabilistic Models (**cDDPMs**) (Dhariwal et al., 2021). These models are trained to translate sharp histological images into their blurred counterparts, preserving the textural semantics of the tissue. We demonstrate that segmentation networks trained on this synthetically generated data exhibit superior generalization in identifying blur artifacts compared to models trained on statistically degraded data or limited data.

Keywords: Histopathology, Diffusion Models, Artifact Detection, Synthetic Data Generation, Image Segmentation

1. Introduction

With the advancement of new and better imaging techniques, digital pathology has seen an unprecedented rise in the last few years. It makes getting a second opinion easy and quick, thereby decreasing the turn around times by a considerable amount. The most crucial requirement for an accurate diagnosis is the clarity and sharpness of the images and the visibility of the tissue on the slide. This is the reason why quality control plays such an important role in digital histopathology.

The Whole Slide Images (WSI) often suffer from different types of artifacts that degrade the quality of the tissue. These artifacts sometimes render the WSI completely useless for any meaningful diagnosis requiring a re-scan. WSIs can have artifacts due to a number of reasons, such as human error: while placing the tissue on the slide or during scanning the slide, systematic error: defect in the equipment, or random error: uneven tissue thickness.

[*] Contributed equally
[†] Contributed equally

Of all the different kinds of artifacts, blur is the most common and most detrimental because it has a very wide variety and at times can be very hard to detect.

The primary reasons for blur artifacts being introduced in the WSI are uneven tissue thickness, folding of tissue regions (while preparing the slide) or the scanner being out of focus. Although the blur artifact is most commonly occurring artifact, there is still not enough annotated data across different tissue morphologies, stains, scanners etc. to train a good deep learning model. Since deep learning models require data annotated at a pixel level for training it is imperative to use synthetically generated data to save costs and introduce variety.

2. Literature Survey

2.1. Statistical methods

Traditional methods for simulating image degradation rely on applying deterministic, mathematically defined convolution kernels. While computationally efficient, these approaches fundamentally model blur as a global, stationary process, failing to capture the complexities of real-world microscopy optics.

Gaussian blur is the most frequently employed statistical method for simulating general out-of-focus artifacts. It approximates the Point Spread Function (PSF) of a simple optical system.The sharp image $I(x, y)$ is convolved with a 2D Gaussian kernel $G(u, v)$ to yield the blurred image $I'(x, y)$.The 2D Gaussian kernel is defined by

$$G(u, v) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}}$$

where σ is the standard deviation.The degree of blur is controlled almost entirely by the standard deviation σ. A larger σ results in a wider kernel and a more pronounced, smooth, isotropic defocusing effect.

Radial blur simulates motion or focusing errors originating from or directed toward a central point, making it a simple model for certain types of field curvature or rotational artifacts. Instead of uniform convolution, the kernel is applied along lines or arcs extending from a defined center point (x_c, y_c) in the image. This results in a degradation where the blur is strongest near the periphery and often minimal near the center, or vice-versa. In histopathology, it can crudely simulate a scenario where the microscope stage is slightly tilted, causing focus to drop off radially, or a slight rotational movement during the scanning process.

Lens blur attempts to model the shape of the physical lens aperture (the bokeh effect) rather than simple diffusion. This method uses a kernel shaped like a polygon or a simple disk, rather than a Gaussian curve. This kernel is applied via convolution. The kernel is often based on a Pillbox function or a shape corresponding to the iris of the camera system. The resulting blur is distinct from Gaussian blur, featuring sharper edges and highlights that take on the shape of the aperture, which is a closer approximation to the true Airy disk or aberration function of real optical systems.

2.2. GAN

As shown by Goodfellow et al. (2014), the GAN framework consists of a Generator (G) and a Discriminator (D) trained in a zero-sum game. G learns to map a latent noise vector (z) to data samples that mimic the training distribution, while D is trained to distinguish between real data and G's synthetic output. The training minimizes the Jensen-Shannon Divergence (JSD) between the real data distribution (p_{data}) and the generated distribution (p_g). This is framed as a minimax objective:

$$\min_G \max_D \mathbb{E}_{\mathbf{x}\sim p_{data}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_z}[\log(1 - D(G(\mathbf{z})))]$$

Early GANs were applied to medical tasks like data augmentation for rare diseases and unconditional image synthesis to increase dataset size and diversity. Standard GANs primarily focus on generating samples from noise ($\mathbf{z}$). For blur synthesis, we need an image-to-image translation framework where the input is the sharp image, and the output is the blurred image, a capability that requires conditioning. cGANs extend the original GAN framework by feeding conditioning information ($\mathbf{c}$) into both the Generator and the Discriminator. This allows for controlled image generation based on a specific input or label.The loss function is modified to include the conditioning vector $\mathbf{c}$:

$$\min_G \max_D \mathbb{E}_{\mathbf{x}\sim p_{data}}[\log D(\mathbf{x}|\mathbf{c})] + \mathbb{E}_{\mathbf{z}\sim p_z}[\log(1 - D(G(\mathbf{z}|\mathbf{c})))]$$

Architectures like Pix2Pix (Isola et al., 2021) formalized cGANs for pixel-to-pixel mapping, making them highly relevant for our task.cGANs have been widely used in digital pathology for stain normalization (HE to Masson's Trichrome), virtual staining (HE to IHC), and super-resolution. Despite their success, cGANs trained with the JSD-based loss often suffer from mode collapse (failing to capture the full diversity of target blur patterns) and training instability (oscillating loss and difficulty converging to Nash Equilibrium), which leads to generated images that lack fine, high-frequency details. To address the stability and quality issues of traditional GANs, Wasserstein GANs were introduced, replacing the JSD divergence with the Earth Mover's (EM) distance or Wasserstein-1 distance. The EM distance provides a cost metric that is continuous and differentiable almost everywhere, even when the distributions are disjoint. This provides a more stable and meaningful gradient signal to the Generator throughout training.The original WGAN required weight clipping on the Discriminator, which limited the model's capacity. WGAN-GP (Gulrajani et al., 2017), replaced weight clipping with a Gradient Penalty term (λ_{GP}):

$$\mathcal{L}_{GP} = \mathbb{E}_{\hat{\mathbf{x}}\sim p_{\hat{\mathbf{x}}}}\left[(\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}})\|_2 - 1)^2\right]$$

where $\hat{\mathbf{x}}$ is sampled from between the real and generated distributions. WGAN-GP enforces the Lipschitz constraint on the Critic via this penalty, leading to minimizing the chances of catastrophic mode collapse and the robust loss helps the Generator produce sharper, more realistic textures, which is essential for synthesizing the subtle, complex structure of histopathology blur. WGAN-GP has been successfully employed in diverse medical image-to-image tasks, including super-resolution reconstruction in MRI and synthetic retinal image generation, specifically because of its ability to generate medically plausible and visually sharp results. Our work leverages WGAN-GP's stability within the cGAN framework to ensure the generated blur is not just present, but realistic and diverse.

2.3. DDPM

DDPMs (Ho et al., 2020) define a Markov chain process that first corrupts a normal image by adding noise to it via a fixed forward diffusion process and then denoises it by a learned reverse denoising process. The forward process gradually adds Gaussian noise (ϵ) to the clean image ($\mathbf{x}_0$) over T time steps until the image becomes pure isotropic Gaussian noise ($\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$). A deep neural network (often a U-Net) learns to reverse this noise process. The network is trained to predict the noise ϵ added at any given step t, allowing it to iteratively recover the image from noise. The training loss is an L2-loss (or equivalent) that minimizes the difference between the actual noise added (ϵ) and the noise predicted by the network (ϵ_θ): $\mathcal{L}_{\text{DDPM}} = \mathbb{E}_{\mathbf{x}_0,t,\epsilon}\left[\|\epsilon - \epsilon\theta(\mathbf{x}_t, t)\|^2\right]$ This simple, non-adversarial objective is the source of DDPMs' superior training stability compared to GANs.

DDPMs are uniquely suited for complex medical image synthesis as they excel at modeling complex, high-dimensional probability distributions, resulting in synthetic images with high perceptual quality and fine textural details mimicking subtle blur patterns. Unlike GANs, which can suffer from mode collapse, DDPMs utilize the Gaussian noise space to ensure greater coverage of the target data distribution, leading to more diverse synthetic artifacts. Similar to how cGANs enabled image-to-image mapping, conditioning is applied to DDPMs to enable the Sharp Image $\rightarrow$ Blurred Image translation required for our approach.

In Conditional DDPMs (cDDPMs), the conditioning input (the sharp image $\mathbf{x}_{\text{sharp}}$) is concatenated or encoded and injected into the noise prediction network ($\epsilon\theta$) at every reverse diffusion step t. The noise prediction is then conditioned on the input: $\epsilon_\theta(\mathbf{x}_t, t, \mathbf{x}_{\text{sharp}})$. This guides the network to denoise the input noise into a blurred image that corresponds to the given sharp image.

The UNIT-DDPM (Sasaki et al., 2021) architecture provides image translation, particularly when annotation pairs may not be perfect. It employs two coupled diffusion processes, one for each domain, with a loss function that encourages the latent representations to be shared. The generative process is then conditioned on the input image using a denoising Markov Chain Monte Carlo approach. UNIT-DDPM formalizes the use of a DDPM backbone to learn the mapping between image characteristics (e.g., texture, contrast, blur level) between two related domains using the shared-latent space assumption of the earlier UNIT framework (Liu et al., 2017). It assumes that images from both the source domain (Sharp) and the target domain (Blurred) can be mapped to a common, domain-invariant latent representation. This assumption is crucial to translate a sharp histopathology image into its intricate, realistic and stochastic blurred counterparts that defy simple statistical modeling.

3. Methodology

3.1. Data

The training of the generative synthesis framework and the subsequent evaluation of the downstream segmentation network relies on three distinct image datasets: a clean source

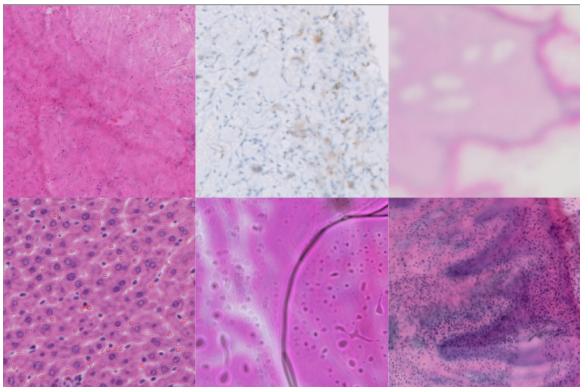

Figure 1: Examples from real world dataset

domain, target domain and a manually annotated set for testing. The source domain $\mathcal{D}_{\text{Sharp}}$ comprises the high-quality, artifact-free histological images used as the input for our statistical and generative models providing the tissue context that must be preserved during the blur synthesis process. The target domain $\mathcal{D}_{\text{Blurred}}$ is used to teach the generative models the characteristics and distribution of real-world blur. This set consists histopathology which are $> 90\%$ blurred. The test set $\mathcal{D}_{\text{Test}}$ is reserved exclusively for the evaluation of the downstream segmentation network. Crucially, these images feature mixed regions of sharp and blurred tissue, reflecting real-world clinical data. Pixel-level ground truth masks are manually annotated and used to compute quantitative metrics for the segmentation networks trained on the synthetically augmented data.

3.2. Statistical Methods

To establish a quantitative baseline for comparison against our deep generative framework, we implemented three standard statistical methods for image degradation. For all statistical methods, the blur was applied globally to the clean images from the source domain ($\mathcal{D}_{\text{Sharp}}$) to create blurred images. Gausian simulates isotropic defocusing by convolving the image with a 2D Gaussian kernel, approximating a simple point spread function (PSF). The sharp tile I_{sharp} is convolved with the kernel $G(\sigma)$. Radial blur models focus degradation or motion that is non-uniform and often centered around a single point.The blur is applied along concentric circles or radial lines originating from a pre-defined center point (x_c, y_c). This simulates a gradient of focus across the image field. Lens blur is used to model the physical shape of the lens aperture, providing a PSF with sharper edges than the Gaussian distribution. The sharp tile is convolved with a disk-shaped kernel (a Pillbox function), which simulates the out-of-focus effect of an idealized circular aperture. These statistical methods, whether Gaussian, Radial, or Lens-based, are fundamentally constrained by their assumption of stationarity and uniformity. Real blur includes noise from ambient light, sensor imperfections, and vibrations, which are inherently stochastic (random) and cannot be accurately replicated by simple deterministic convolution.

3.3. GAN

To overcome the limitations of statistical methods, we utilized a robust Conditional Generative Adversarial Network (cGAN) architecture for image-to-image translation, mapping sharp histological tiles ($\mathbf{x}_{\text{sharp}} \in \mathcal{D}_{\text{Sharp}}$) to their synthetically blurred counterparts ($G(\mathbf{x}_{\text{sharp}})$).The cGAN framework is based on the Pix2Pix architecture, adapted for histopathology.We employed a U-Net based architecture for G. This ensures that the generated blurred image retains the semantic context of the input sharp image while the decoder layers model the complex blur transformation. Discriminator is used to classify patches as real or fake. To mitigate the well-known issues of GAN instability and mode collapse, which are particularly severe when generating realistic medical images, we incorporated several stabilization techniques.

We used WGAN with gradient penalty, this replaces the unstable standard GAN loss based on Jensen-Shannon Divergence, we adopted the Wasserstein-1 distance objective.This involved re-formulating the discriminator as a Critic and integrating a Gradient Penalty (λ_{GP}) term into the Critic's loss function. This penalty enforces the 1-Lipschitz continuity constraint without requiring restrictive weight clipping. WGAN-GP provides a more stable loss metric, leading to more consistent convergence, reduced mode collapse, and higher fidelity in the generated blur textures. We set the penalty coefficient $\lambda_{GP} = 10$.

To explicitly constrain the Generator to only perform the required blur transformation and preserve the structural content, we introduced an identity mapping loss. When an image $\mathbf{x}_{\text{blurred}}$ from the target domain ($\mathcal{D}_{\text{Blurred}}$) is passed through the Generator, the output is encouraged to remain close to the input. The ℓ_1 loss is computed:

$$\mathcal{L}_{\text{Identity}}(G) = \mathbb{E}_{\mathbf{x}_{\text{blurred}} \sim \mathcal{D}_{\text{Blurred}}}[\|\mathbf{x}_{\text{blurred}} - G(\mathbf{x}_{\text{blurred}})\|_1]$$

. This penalty discourages the Generator from making unnecessary color or structural changes, focusing its learning capacity entirely on the difference between the sharp and blurred domains. The training was initiated with a fixed learning rate α=0.0002 for the first 100 epochs. This was followed by a linear decay of α to zero over the remaining 100 epochs.The initial high rate facilitates rapid convergence, while the decay helps the model settle into a fine-tuned, stable solution space, optimizing for the subtle textural details of the blur.

We used pretraining to give the Generator a robust initial mapping before subjecting it to the adversarial pressure of the Discriminator.The Generator was initially trained independently using only the ℓ_1 reconstruction loss ($\mathcal{L}_{\text{L1}}$). This required the creation of a temporary, simple statistical paired dataset in which random pixels are dropped, to define the initial mapping. This pretraining ensures the Generator can accurately map the input to the output space structurally, making the subsequent adversarial training phase more stable and accelerating convergence towards the realistic blur distribution.

3.4. DDPM

We implemented the UNIT-DDPM architecture to perform unpaired image-to-image translation from the sharp domain A($\mathcal{D}_{\text{Sharp}}$) to the blurred domain B ($\mathcal{D}_{\text{Blurred}}$). UNIT-DDPM operates on the fundamental assumption that images from both the source (sharp) and target (blurred) domains share a common, domain-invariant latent representation. This allows

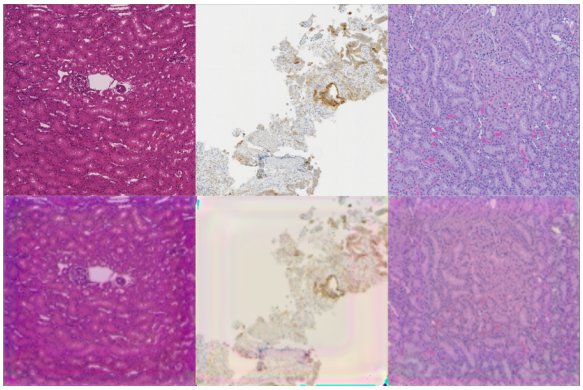

Figure 2: Original Images and their synthetically generated blurry counterparts

the model to learn the intrinsic characteristics of the tissue independent of the artifact type. This framework has two translation models and two cDDPM models. Translation models translate images from A to B and B to A. cDDPM generators (θ_A, θ_B) are trained to predict the noise added in the forward diffusion step for their respective domains, conditioned on the latent code z derived from the opposite domain. Unlike GANs which generate an image in one forward pass, generating a synthetically blurred image requires the full iterative reverse diffusion process. During sampling, the generative process is conditioned on the input source domain images that are perturbed by the forward process from t = T until an arbitrary timtestep t_r [1, T]. This is then re-generated by the reverse process from this timestep, which we denote as the release time. The case of transferring from domain A x_A^0 to domain B $\hat{x}_B^0$ is described as:

$$\hat{\mathbf{x}}_{t-1}^B = \mu_{\theta^B}(\hat{\mathbf{x}}_t^B, \hat{\mathbf{x}}_t^A, t) + \Sigma_{\theta^B}(\mathbf{x}_t, t)\epsilon^B$$

$$\hat{\mathbf{x}}_{t-1}^A = \begin{cases} \sqrt{\tilde{\alpha}_{t^A}\hat{\mathbf{x}}_0^A} + \sqrt{1-\tilde{\alpha}_{t^A}}\epsilon^A & (t > t_r) \\ \mu_{\theta^A}(\hat{\mathbf{x}}_t^A, \hat{\mathbf{x}}_t^B, t) + \Sigma_{\theta^A}(\mathbf{x}_t^A, t)\epsilon^B & (t \leq t_r) \end{cases}$$

$$\hat{\mathbf{x}}_T^B, \epsilon^A, \epsilon^B \sim \mathcal{N}(0, 1)$$

4. Results

We used the GAN to generate blurry images from existing sharp images (without blur). Next we created random polygon masks using opencv functions. Using these polygons we replace some parts of the sharp images with the blurry image. Now using the mask as the ground truth for blurry regions in the image we train a segmentation model and then test it on annotated real world data.

Qualitative analysis as well as the model performance metrics showed that the synthetic data generated by out approach successfully covered multiple modes of the real data distribution, leading to superior performance as compared to the synthetic data generated via statistical methods.

Generation type	Blur Dice
Statistical methods	23.45
GAN	**38.33**

Table 1: Performance Comparison of Synthetic data generation techniques

5. Conclusion

This study addressed the critical data scarcity challenge in digital histopathology quality control by introducing a novel framework for synthesizing realistic, spatially-variant blur artifacts. Our approach successfully demonstrated the necessity of transitioning from traditional, deterministic degradation methods to deep generative modeling to achieve robust downstream task performance.

Acknowledgments

We thank a bunch of people.

References

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems, volume 27, pages 2672–2680, 2014.

Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1125–1134, 2017.

Martin Arjovsky, Soumith Chintala, and L'eon Bottou. Wasserstein generative adversarial networks. In International Conference on Machine Learning (ICML), pages 214–223, 2017.

Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of Wasserstein GANs. In Advances in Neural Information Processing Systems, volume 30, pages 5767–5777, 2017.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 2223–2232, 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In Advances in Neural Information Processing Systems, volume 33, pages 6840–6851, 2020.

Hiroshi Sasaki, Chris G. Willcocks, and Toby P. Breckon. UNIT-DDPM: UNpaired Image Translation with Denoising Diffusion Probabilistic Models. arXiv preprint arXiv:2104.05358, 2021.

Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In Advances in Neural Information Processing Systems, volume 30, pages 700–709, 2017.