Self-Diffusion Driven Blind Imaging

Yanlong Yang
yyxr164579504@gmail.com

Guanxiong Luo *
luoguan5@gmail.com

Abstract

Optical imaging systems are inherently imperfect due to diffraction limits, lens manufacturing tolerances, assembly misalignment, and other physical constraints. In addition, unavoidable camera shake and object motion further introduce non-ideal degradations during acquisition. These aberrations and motion-induced variations are typically unknown, difficult to measure, and costly to model or calibrate in practice. Blind inverse problems offer a promising direction by jointly estimating both the latent image and the unknown degradation kernel. However, existing approaches often suffer from convergence instability, limited prior expressiveness, and sensitivity to hyperparameters. Inspired by recent advances in self-diffusion, we propose DeblurSDI, a zero-shot, self-supervised blind imaging framework that requires no pre-training. DeblurSDI formulates blind image recovery as an iterative reverse self-diffusion process that begins from pure noise and progressively refines both the sharp image and the blur kernel. Extensive experiments on combined optical aberrations and motion blur demonstrate that DeblurSDI consistently outperforms other methods by a substantial margin.

1. Introduction

Optical imaging systems are widely used in various sensors such as cameras [13, 27], microscopes [7], telescopes [31], and LiDARs that operate across wavelengths from visible to infrared. The lens system and the focal-plane image sensor jointly constitute a complete imaging module. In a paraxial, shift-invariant optical system, the captured image can be modeled as

$$I(x, y) = O(x, y) \circledast h(x, y), \tag{1}$$

where $O(x, y)$ denotes the object's intensity distribution and $h(x, y)$ is the lens's point spread function (PSF) given by

$$h(x, y) = \left| \mathcal{F}\{P(\rho, \theta)\, e^{i2\pi W(\rho,\theta)/\lambda}\} \right|^2. \tag{2}$$

Here $\mathcal{F}$ denotes the Fraunhofer diffraction operator [11], $P(\rho, \theta)$ is the complex pupil function, $W(\rho, \theta)$ is the wavefront, and λ is the wavelength [4]. Because of inherent imperfections such as diffraction limits [11], manufacturing tolerances [33], assembly misalignment [26], and unavoidable camera shake and object motion [14, 37], the captured images are often corrupted by optical aberrations [26, 28] and motion blur [34, 37].

A straightforward way to improve image quality is to explicitly calibrate the system's PSF by measuring optical aberrations and motion-blur kernels for each lens [7, 30, 31]. However, this calibration process is often time-consuming, requires dedicated hardware, and relies heavily on expert knowledge. Blind inverse problems offer a promising direction by jointly estimating both the latent image and the unknown degradation PSF [8, 23]. This approach provides a unified framework for both calibration and deblurring, and has been applied to a wide range of applications not limited to deblurring, such as super-resolution and image denoising.

However, existing approaches often exhibit unstable convergence [23], limited prior expressiveness [20], and pronounced sensitivity to hyperparameters [2, 5, 21, 29, 32], which make them difficult to use reliably in practice. Motivated by recent developments in self-diffusion [24], we propose DeblurSDI, a training-free and hyperparameter-insensitive blind imaging framework that achieves high stability while requiring no external priors. Inheriting the core idea of self-diffusion, which progressively restores images across hierarchical noise levels, DeblurSDI extends this principle to jointly recover the sharp image and PSF. This formulation significantly improves robustness in the otherwise unstable joint optimization, particularly in PSF estimation, which is the primary challenge in blind imaging problems.

Extensive experiments on simulated optical aberrations and motion blur demonstrate that DeblurSDI consistently outperforms state-of-the-art blind deconvolution methods by a substantial margin. Our contributions are summarized as follows: 1) We propose DeblurSDI, a training-free and self-supervised framework for blind image recovery that integrates self-diffusion and jointly estimates both the sharp

*Equal contribution to this work

image and the PSF; 2) We construct evaluation datasets that include optical-aberration PSFs and motion-blur kernels, enabling comprehensive assessment of blind imaging performance under challenging conditions; 3) We demonstrate that DeblurSDI achieves consistently superior performance and robustness compared to state-of-the-art blind deblurring methods, particularly in stability and PSF estimation.

2. Related Work

Correcting optical aberrations and motion blur generally follows one of two paradigms. In the model-based paradigm, the PSF is described using a parametric wavefront model such as Zernike polynomials [25, 28], and calibration procedures are employed to estimate the corresponding coefficients. This yields an explicit PSF that enables non-blind deconvolution. In the model-free paradigm, blind deconvolution is performed, which bypasses calibration entirely by estimating both the image and the PSF from the observation [16].

Optical Aberration Modeling and PSF Calibration: Optical aberrations arise from diffraction limits [11], manufacturing tolerances [33], lens-group misalignment [26], and environmental variations [4], and they fundamentally shape the PSF of an imaging system. Early descriptions of aberrations were grounded in classical geometric optics [36], characterizing low-order distortions such as spherical aberration, coma, astigmatism, field curvature, and distortion [26]. With the development of wave optics [11], the PSF became formally modeled as the squared magnitude of the Fourier transform of the pupil function [15], where phase distortions encode the underlying wavefront errors. Modern optical systems typically represent these wavefront errors using Zernike polynomials [16, 25, 28], an orthonormal basis on the unit disk that directly corresponds to physically interpretable aberration modes and has become the industry standard for lens characterization [16].

Estimating the actual PSF of a physical lens, however, requires calibration procedures [30, 31]. In controlled laboratory settings, wavefront sensors such as Shack–Hartmann devices [31] provide direct measurements of Zernike coefficients, while phase-diversity methods infer aberrations from pairs of focused and defocused images [10, 30]. Microscopy systems often rely on submicron fluorescent beads [7] to sample spatially varying PSFs and construct eigen-PSF [1] or product-convolution models [3] for wide-field correction. Although accurate, these calibration pipelines require specialized hardware, multiple acquisitions, or carefully prepared samples, making them impractical for general-purpose cameras and consumer-level imaging devices.

As a result, industrial imaging systems such as smartphone cameras [13, 27], surveillance sensors, and automotive optical modules—typically rely on one-time factory calibration combined with lookup tables [27, 38] or compact neural modules embedded in the ISP pipeline. These approaches correct only approximate or averaged aberrations and cannot account for camera shake, object motion, or per-instance PSF variations in real scenes. Therefore, while model-based PSF calibration offers high fidelity, its cost, hardware dependency, and lack of adaptability highlight the need for calibration-free alternatives such as blind deconvolution methods [8, 23].

Blind Deconvolution with Training-free or Trained Priors: Blind deconvolution aims to recover both the latent sharp image and the unknown blur kernel from a single degraded observation. Classical approaches typically formulate this as an alternating optimization problem that minimizes an energy functional over the image and the PSF [8, 23]. Early methods rely on handcrafted priors such as sparsity, heavy-tailed gradient distributions, or total variation to regularize the ill-posed joint estimation [5, 20]. Although effective in limited settings, these methods often exhibit strong sensitivity to initialization, kernel size, and hyperparameter choices, which leads to kernel drift and unstable convergence behaviors.

To alleviate the limitations of handcrafted priors, implicit neural priors have emerged as a powerful alternative. Deep Image Prior (DIP) demonstrates that the structure of an untrained convolutional network can serve as a natural image prior, enabling image restoration without external data [35]. Subsequent works extend this idea to blind deblurring by optimizing both the image and the kernel through coupled implicit networks [32]. These methods offer improved flexibility and avoid the need for large training datasets, yet the coupled optimization is still highly unstable and prone to overfitting, especially when the blur kernel is large or spatially complex.

Another line of work introduces pretrained generative models such as GANs, VAEs, and more recently diffusion models as strong image priors for inverse problems [9, 18, 39]. Diffusion-based solvers, including diffusion posterior sampling (DPS) and related score-based formulations, have demonstrated impressive performance in non-blind deblurring and other linear inverse problems [6, 12]. However, these methods rely heavily on large-scale pretraining and are restricted by domain mismatch between the training data and the target scene. More importantly, pretrained-prior approaches are generally incapable of estimating the blur kernel, and therefore cannot be directly applied to fully blind deconvolution.

Because existing methods either require dedicated calibration hardware or struggle with the inherent instability of joint image–kernel estimation. These limitations highlight the need for a framework that can both capture realistic

optical degradations and robustly estimate the PSF without external training or calibration. Motivated by this gap, we now introduce our degradation modeling pipeline and the proposed DeblurSDI framework.

3. Method

In this section, we first introduce our aberration simulation framework, which generates a family of physically realistic PSFs derived from wavefront distortions modeled by Zernike polynomials. These simulated PSFs allow us to systematically evaluate blind imaging performance under controlled optical conditions. We then present DeblurSDI, a training-free blind deblurring framework that jointly recover the latent sharp image and the unknown PSF. By combining physically grounded PSF synthesis with a robust blind inverse solver, our approach enables comprehensive analysis and effective restoration across a wide range of optical aberrations and motion blur scenarios.

3.1. Modeling Optical Aberrations

To simulate realistic optical degradations, we adopt the standard wavefront–pupil–PSF formulation widely used in computational optics. Optical aberrations are represented through a weighted sum of Zernike polynomials [25], which form an orthonormal basis on the unit disk and are known to accurately model low- and high-order aberration modes such as defocus, astigmatism, coma, spherical aberration, trefoil, and quadrafoil.

Zernike-based wavefront modeling: Let (ρ, θ) denote polar coordinates on the normalized pupil domain $\{(x, y) : x^2 + y^2 \leq 1\}$. The Zernike polynomial of radial order n and azimuthal order m is defined as

$$Z_n^m(\rho,\theta) = \begin{cases} R_n^{|m|}(\rho)\cos(m\theta), & m \geq 0, \\ R_n^{|m|}(\rho)\sin(|m|\theta), & m < 0, \end{cases} \tag{3}$$

where the radial term $R_n^m(\rho)$ is given by

$$R_n^m(\rho) = \sum_{k=0}^{\frac{n-m}{2}} (-1)^k \, \frac{(n-k)!}{k!\left(\frac{n+m}{2}-k\right)!\left(\frac{n-m}{2}-k\right)!} \, \rho^{\,n-2k}. \tag{4}$$

A wavefront corrupted by multiple aberrations can therefore be written as

$$W(\rho,\theta) = \sum_{(n,m)\in\mathcal{A}} a_{n,m} \, Z_n^m(\rho,\theta), \tag{5}$$

where each coefficient $a_{n,m}$ controls the severity of the corresponding aberration mode. In our simulations, we include all Zernike modes up to order $n = 4$ (defocus, astigmatism, coma, trefoil, spherical, quadrafoil), which are sufficient to reproduce a wide range of realistic aberration patterns observed in lens-based imaging systems.

Pupil function and PSF generation: Given the aberrated wavefront W, the complex pupil function is

$$P(\rho,\theta) = \mathbf{1}_{\{\rho\leq 1\}} \, \exp\!\left(\frac{2\pi i}{\lambda} \, W(\rho,\theta)\right), \tag{6}$$

where λ denotes the wavelength and the indicator ensures that the phase modulation is applied only within the physical aperture. The PSF is obtained as the squared magnitude of the Fourier transform of the pupil:

$$h(x,y) = \frac{\left|\mathcal{F}\{P(\rho,\theta)\}\,(x,y)\right|^2}{\max\limits_{x,y}\left|\mathcal{F}\{P(\rho,\theta)\}\,(x,y)\right|^2}, \tag{7}$$

where normalization ensures that $\max h = 1$. We use an odd-sized Fourier grid so that the PSF peak is centered exactly at the middle pixel. A central crop of h is used as the final convolution PSF in all experiments.

3.2. Blind Imaging via Self-Diffusion

This section introduces the DeblurSDI method, which addresses blind optical correction by extending the self-diffusion framework [24]. The original self-diffusion method was designed for non-blind problems where the degradation operator is known. Our key contribution is to adapt this powerful framework to the more challenging blind imaging setting, where the degradation operator is also unknown. To achieve this, we introduce a novel process that jointly recovers the clean image and the PSF by optimizing two coupled, untrained neural networks within a self-contained reverse diffusion process.

Self-Diffusion: Self-diffusion is a pretraining free method designed to solve general linear inverse problems of the form $\mathcal{A}\mathbf{x}_{\text{true}} = \mathbf{y}$, where $\mathcal{A}$ is a known forward operator, $\mathbf{x}_{\text{true}}$ is the unknown solution, and $\mathbf{y}$ is the observation. It operates via an iterative reverse diffusion process that starts from pure Gaussian noise. At each step t, a noisy version of the current estimate $\mathbf{x}_t$ is created with

$$\hat{\mathbf{x}}_t = \mathbf{x}_t + \sigma_t \cdot \epsilon_t \, . \tag{8}$$

A single, randomly initialized network at the first time step, the self-denoiser D_θ, is then optimized continuously by minimizing a data fidelity loss with respect to the original observation $\mathbf{y}$ for each time step t,

$$\mathcal{L}_t(\theta) = \|\mathcal{A}D_{\theta,t}(\hat{\mathbf{x}}_t) - \mathbf{y}\|_2^2 \, . \tag{9}$$

The effectiveness of this process stems from a principle known as noise-regulated spectral bias. The noise schedule σ_t implicitly regularizes the optimization, forcing the network to first learn low-frequency components and progressively refine high-frequency details in a multi-scale manner.

Joint Image and PSF Estimation: Based on above observations, we apply such implicit regularization to PSF estimation. The forward model for blind image deblurring is given by

$$\mathbf{y} = \mathbf{x}_{\text{true}} \circledast \mathbf{k} + n \,, \tag{10}$$

where the $\circledast$ denotes convolution, the sharp image $\mathbf{x}_{\text{true}} \in \mathbb{R}^{H\times W\times C}$ and the PSF $\mathbf{k} \in \mathbb{R}^{K\times K\times 1}$ are both unknown. To adapt the self-diffusion framework to this blind setting, both variables must be estimated simultaneously. We achieve this by employing two dedicated, randomly initialized networks: an image denoiser D_θ to restore $\mathbf{x}_{\text{true}}$, and a PSF generator G_ϕ to produce $\mathbf{k}$. Our method simulates a reverse diffusion process over T discrete time steps, starting with random noise for both the image estimate, $\mathbf{x}_T$, and the PSF estimate, $\mathbf{z}_T$. At each time step $t \in \{T, T-1, ..., 1\}$, the current estimates are perturbed with scheduled noise,

$$\hat{\mathbf{x}}_t = \mathbf{x}_t + \sigma_t \cdot \epsilon_x, \quad \text{and} \quad \hat{\mathbf{z}}_t = \mathbf{z}_t + \sigma'_t \cdot \epsilon_z, \tag{11}$$

where $\epsilon_x \sim \mathcal{N}(0, \mathbf{I})$ and $\epsilon_z \sim \mathcal{N}(0, \mathbf{I})$. The noise schedule is $\sigma_t = \sqrt{1-\bar{\alpha}_t}$, where $\bar{\alpha}_t = \prod_{i=0}^{t}(1-\beta_i)$ and $\beta_t = \beta_{\text{end}} + \frac{t}{T-1}(\beta_{\text{start}} - \beta_{\text{end}})$, and $\sigma'_t = \mu\sigma_t$, μ is a adjutable hyperparameter. It is often known that the PSF is to be sparse in practice. We thus define $R(\cdot)$ as L1-norm and use many layers of ReLU in the PSF generator to enforce sparsity on the generated PSF. The both networks are jointly optimized within an inner loop by minimizing the following objective:

$$\mathcal{L}_t(\theta, \phi) = \|(D_\theta(\hat{\mathbf{x}}_t) \circledast G_\phi(\hat{\mathbf{z}}_t)) - \mathbf{y}\|_2^2 + \lambda_k R(G_\phi(\hat{\mathbf{z}}_t)) \tag{12}$$

After the inner optimization loop, the improved networks produce cleaner estimates for the next time step, $\mathbf{x}_{t-1} = D_\theta(\hat{\mathbf{x}}_t)$ and $\mathbf{z}_{t-1} = G_\phi(\hat{\mathbf{z}}_t)$, continuing the coarse-to-fine reconstruction inherent to the self-diffusion process. The detailed algorithm is presented in Algorithm 1.

Algorithm 1 Blind Imaging via Self-Diffusion

Require: Blurry image $\mathbf{y}$, total steps T, inner iterations S, learning rate η, ℓ_1 weights λ_k

Initialize:
Image estimate $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$; D_θ with random weights θ
PSF estimate $\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I})$; G_ϕ with random weights ϕ
Adam optimizer for (θ, ϕ)
Noise schedule σ_t for $t \in \{1, ..., T\}$
for $t = T, T-1, ..., 1$ **do**
  Sample noise $\epsilon_{\mathbf{x}}, \epsilon_{\mathbf{z}} \sim \mathcal{N}(0, \mathbf{I})$
  Create noisy inputs: $\hat{\mathbf{x}}_t \leftarrow \mathbf{x}_t + \sigma_t \cdot \epsilon_{\mathbf{x}}$, and $\hat{\mathbf{z}}_t \leftarrow \mathbf{z}_t + \sigma'_t \cdot \epsilon_{\mathbf{z}}$
  for $s = 1, ..., S$ **do**
    Generate PSF: $\mathbf{k} \leftarrow G_\phi(\hat{\mathbf{z}}_t)$
    Compute denoised image: $\mathbf{x}_t \leftarrow D_\theta(\hat{\mathbf{x}}_t)$
    Calculate loss: $\mathcal{L}(\theta, \phi) \leftarrow \|(\mathbf{x}_t \circledast \mathbf{k}) - \mathbf{y}\|_2^2 + \lambda_k R(\mathbf{k})$
    Update parameters via gradient descent: $(\theta, \phi) \leftarrow (\theta, \phi) - \eta \nabla_{(\theta,\phi)} \mathcal{L}(\theta, \phi)$
  end for
  Update image estimate: $\mathbf{x}_{t-1} \leftarrow D_\theta(\hat{\mathbf{x}}_t)$
  Update PSF latent code: $\mathbf{z}_{t-1} \leftarrow G_\phi(\hat{\mathbf{z}}_t)$
end for
return Reconstructed image $\mathbf{x} \leftarrow \mathbf{x}_0$ and final PSF $\mathbf{k} \leftarrow \mathbf{k}_0$

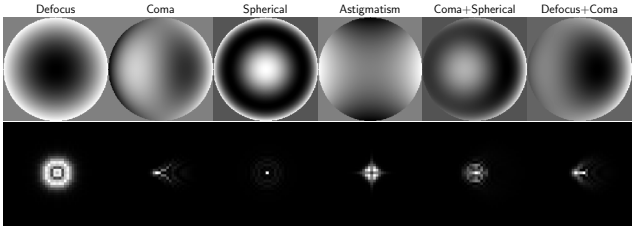

Figure 1. PSFs for individual and combined aberrations (defocus, coma, astigmatism, spherical, coma+spherical, defocus+coma). Uprow: wavefronts. Downrow: PSFs.

4. Implementation and Experiments

4.1. Dataset

Optical Aberration: The simulated optical aberration process is implemented numerically by evaluating the Zernike basis on a 255×255 Cartesian grid, constructing the pupil (6), and computing the PSF via FFT according to (7). Figure 1 shows the resulting wavefronts and PSFs for individual and combined aberrations (defocus, coma, astigmatism, spherical, coma+spherical, defocus+coma). These aberrations are all applied on sharp images collected from above datasets to obtain blurred images. Then our proposed DeblurSDI is applied and evaluated on them.

Motion Blur: To systematically evaluate the performance of our method, we collected four datasets including Levin [22], Cho [5], Kohler [19], and FFHQ [17]. The first three datasets are widely used benchmarks for blind image deblurring, containing various synthetic blur kernels applied to natural images. The FFHQ dataset consists of 20 random selected human face images and 4 blur kernels from Cho [5]. For our experiments, we use $T = 30$ outer reverse diffusion steps. At each step, we run $S = 200$ inner optimization iterations. For each dataset, we use the provided blur kernels and generate blurred images by convolving the sharp images with these kernels to simulate real-world conditions. Each blur kernel is applied to every image in the dataset. The details of our comprehensive datasets are

shown in Table 1.

Table 1. Details of datasets for evaluation

Dataset	Image Size	Kernel Sizes
Levin[1]	255×255	{19, 17, 15, 27, 13, 21, 23, 23}
Cho[2]	622×463, 780×580, 1006×665, 1002×661	{27, 23, 19, 21}
Kohler[3]	800×800	16, 14, 9, 13, 29, 17, 19, 98, 102, 62, 40, 29
FFHQ[4]	256×256	{27, 23, 19, 21}
Total Pairs		128

4.2. Network architecture and training

Due to the low-dimensional nature of the blur kernel, we employ a fully-connected network (FCN) to implement the kernel generator, G_ϕ. To ensure the output corresponds to a physical blur kernel, a softmax activation is applied to the final layer, enforcing non-negativity and a sum-to-one constraint. The 1D output of G_ϕ is subsequently reshaped into a 2D blur kernel. Besides, we introduced standard mode for ablation study where the latent vector z is sampled from a normal distribution and kept fixed during training, and diffusion mode where the $\mathbf{z}_t$ evolves through the self-diffusion process. K is the kernel size, n is the number of hidden layers and H_d is the hidden dimension size, Table 2 shows the architecture in details.

Table 2. The architecture of the kernel generator G_ϕ

Mode	Layer	Specification
Standard	Input	$z \in \mathbb{R}^{200} \sim \mathcal{N}(0, \mathbf{I})$
	Hidden layer	Linear(200, 2000); ReLU6
	Output layer	Linear(2000, $K \times K$); Softmax
Diffusion	Input	$\mathbf{z}_t \in \mathbb{R}^{K \times K}$
	Hidden layer 1	Linear($K \times K$, H_d); ReLU
	Hidden layer 2	Linear(H_d, H_d); ReLU
	...	...
	Hidden layer n	Linear(H_d, H_d); ReLU
	Output layer	Linear(H_d, $K \times K$); Softmax

For the image denoiser D_θ, we employ an encoder-decoder network with skip connections, following a U-Net-like structure. The network consists of five hierarchical levels. Each level in the encoder path consists of two convolutional blocks and a stride-2 convolution for downsampling. Correspondingly, the decoder path uses bilinear upsampling. Skip connections concatenate features from each encoder level to the corresponding decoder level. Non-Local Blocks are integrated into the deeper encoder levels (levels 3, 4, and 5) to capture long-range dependencies. The architecture is detailed as in Table 3.

Table 3. The architecture of the image denoiser D_θ, consisting of five encoder units (e_i) and five decoder units (d_i). The form is Conv(input channels, output channels, kernel size).

Layer	Specification
Input	Noisy image $\hat{x}_t \in \mathbb{R}^{C \times H \times W}$
Output	Denoised image $x_{t-1} \in \mathbb{R}^{C \times H \times W}$
Encoder unit 1	$e_1(\cdot, 128, 3)$
Encoder unit 2	$e_2(128, 128, 3)$
	...
Encoder unit 5	$e_5(128, 128, 3)$
Decoder unit 5	$d_5(128, 128, 3)$
	...
Decoder unit 2	$d_2(128, 128, 3)$
Decoder unit 1	$d_1(128, 128, 3)$
Output layer	Conv(128, C, 1); Sigmoid

We use a single Adam optimizer to jointly update the parameters of both the image denoiser, D_θ, and the kernel generator, G_ϕ. The initial learning rate for the image denoiser D_θ is set to 1×10^{-3}. The kernel generator G_ϕ uses a lower learning rate, typically 25% of the denoiser's rate (i.e., 2.5×10^{-4}), as the small change in the kernel can make a bigger impact on the image. The smaller learning rate helps stable convergence of the kernel estimate. Furthermore, we employ an optional adaptive learning rate schedule for the kernel generator. The learning rate is decayed by a factor of 0.95 at the end of each outer time step t, down to a minimum threshold of 1×10^{-5}. The L1 regularization weight for the kernel prior is set to $\lambda_k = 2 \times 10^{-3}$. The noise level σ_t for the image perturbation at each step t is determined by a pre-defined variance schedule. Following common practice in diffusion models, we use a linear schedule where the variance β_t interpolates from $\beta_{\text{start}} = 1 \times 10^{-4}$ to $\beta_{\text{end}} = 2 \times 10^{-2}$ over T steps. The noise level σ_t is then derived from the cumulative product of these variances.

4.3. Experiments

In the following, we present the deblurring process described in Algorithm 1 and evaluated DeblurSDI from different perspectives, including robustness to the choice of PSF generator, sensitivity to hyperparameters, and noise degradation.

Hyperparameters sensitivity: The number of outer diffusion steps T and inner optimization iterations S in Algorithm 1 are two critical hyperparameters that directly impact both the reconstruction quality and computational cost. Intuitively, more diffusion steps allow for a finer coarse-to-fine reconstruction, while more inner iterations enable better convergence of the networks at each step. However, increasing either parameter also leads to longer runtimes. To evaluate the sensitivity of our method to these parameters,

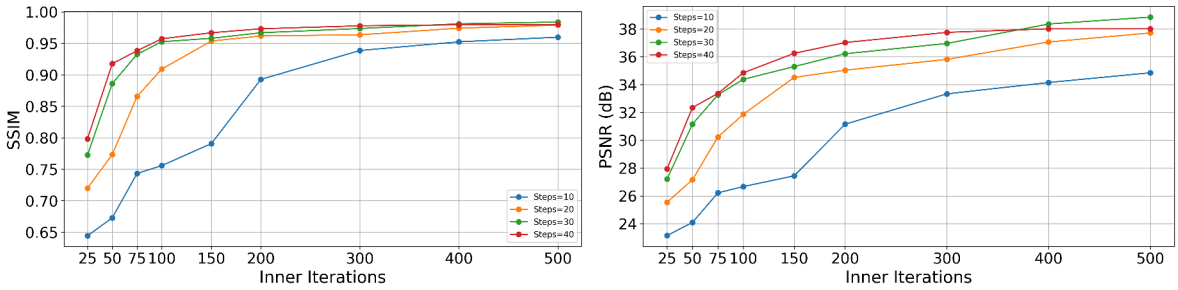

Figure 2. Sensitivity of noise scheduled steps, T, and inner iterations, S. The graphs show the SSIM (left) and PSNR (right) scores for different numbers of outer diffusion steps ($T \in \{10, 20, 30, 40\}$) and inner optimization iterations ($S \in \{25, ..., 500\}$).

we conduct experiments varying T from 10 to 40 and S from 25 to 500. As shown in Figure 2, performance improves with higher values of T and S. The most significant gains occur when increasing T from 10 to 30. However, the improvements tend to saturate beyond certain thresholds (e.g., $T = 30$, $S = 400$), with the performance curve for $T = 40$ closely tracking that of $T = 30$. This indicates that our approach can achieve strong deblurring performance without requiring excessively high iteration counts.

Recovering process: Figure 3 shows the evolution of estimates of image and kernel through the deblurring process. The left and right subfigures shows the SSIM and PSNR between the original image and the reconstruction over noise steps. The estimates of images and kernels at noise steps $5, 10, 15, 20, 30$ are shown on the top. Unlike traditional optimization processes where evaluation metrics typically increase monotonically, our curves exhibit an up-down-up behavior (especially for PSNR curve), which is attributed to the noise scheduling strategy. By injecting noise into intermediate reconstruction results, we effectively enlarge the search space of the inverse solution. The initial reconstructions are smooth and lack fine detail, while later steps recover sharper features. This aligns with the coarse-to-fine nature of self-diffusion.

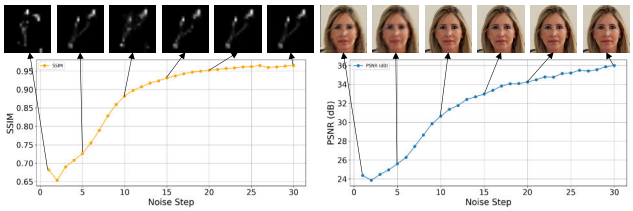

Figure 3. Evolution of image and kernel estimates during DeblurSDI's reverse diffusion process. Left/right subfigure: SSIM/PSNR between the original image and the reconstruction over noise steps. The joint estimation of images and kernels at noise steps $5, 10, 15, 20, 30$ are shown on the top.

Architecture of kernel generator: Figure 4 shows how the architecture of PSF kernel generator G_ϕ affects the performance of DeblurSDI. The "Standard" mode is G_ϕ without noise schedule, while "k-diff" stands for the kernel generator G_ϕ with noise schedule as Table 2 describes. The results show that the "k-diff" mode outperforms the "Standard" mode in terms of PSNR and SSIM. The results indicate that increasing the depth of the kernel generator network leads to a more accurate kernel estimation.

Figure 4. Ablaion study on the network for PSF estimation. This figure compares the performance of the "Standard" mode against the "Diffusion" mode architecture with a varying number of hidden layers ($n = 1$ to $n = 5$).

Noise degradation: To evaluate the robustness of our method to additive noise, we test it across noise levels $\sigma \in 0.01, 0.02, 0.03, 0.05$ using the degradation model $\mathbf{y} = \mathbf{x} \circledast \mathbf{k} + \sigma\mathbf{n}$, $\mathbf{n} \sim \mathcal{N}(0, \mathbf{I})$. The image is normalized to $[0, 1]$. As shown in Figure 5, our method remains stable for noise levels up to $\sigma = 0.02$, achieving PSNR above 28.0 and SSIM above 0.73. The recovered PSFs also remain visually accurate up to $\sigma = 0.03$.

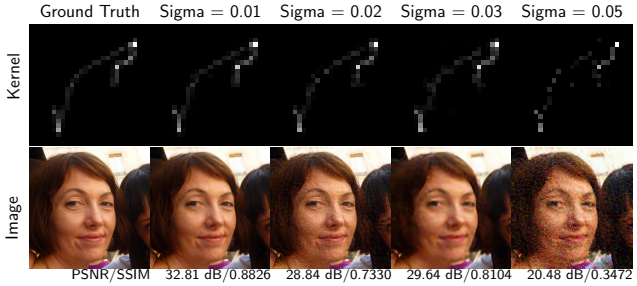

Figure 5. Performance comparison under different addictive noise on the blurred image. The first and second rows of the figure show the Kernel and Image reconstructions according to the Noise factor $\sigma \in 0.01, 0.02, 0.03, 0.05$.

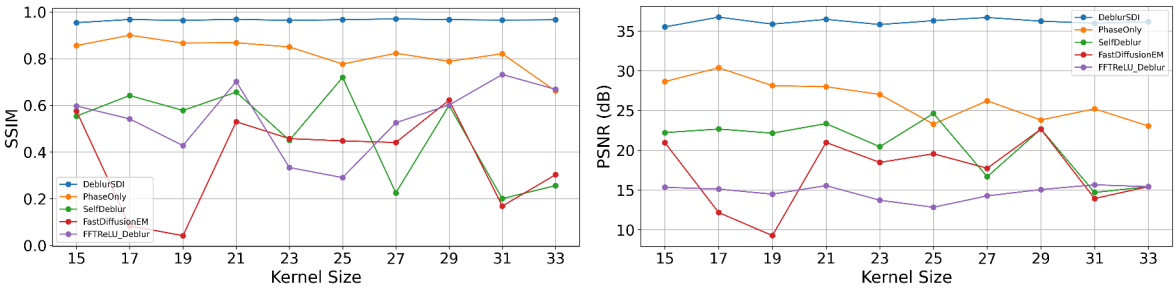

Figure 6. Performance and stability comparison across different kernel sizes. The graphs show the SSIM (left) and PSNR (right) scores for five deblurring methods evaluated on kernel sizes ranging from 15 to 33. Our method, DeblurSDI (blue), consistently achieves the highest scores and demonstrates remarkable stability, with its performance remaining largely unaffected by changes in kernel size. In contrast, other methods exhibit significant volatility, underscoring the superior robustness of our approach.

5. Evaluation

In this section, we evaluated DeblurSDI on datasets described in Section 4.1 comparing to other methods. The joint estimation of the image denoiser and the blur kernel often collapses to trivial solutions, such as Dirac kernels or reproducing the blurred image itself, especially when the chosen kernel size is incompatible with the image content. Larger kernels are harder to be recovered accurately, while smaller kernels may fail to capture long-range motion. For this reason, SelfDeblur [32] carefully selects the kernel size for each image. In contrast, our method exhibits much greater robustness. As shown in Figure 6, we evaluate ten different kernel sizes from 15 to 33, and compare the performance of several approaches. Our method not only achieves consistently superior performance across all kernel sizes but also demonstrates remarkable stability.

Figure 7. The reconstruction results of DeblurSDI on the optical aberration situations. The reconstruction results of aberration PSFs and blurred images are shown in the left and right subfigures.

Optical Aberration Correction: Based on the simulated six optical aberrations from 3.1, all the clean images are blurred by each one of them to obtain the corresponding blurred images. Then the proposed DeblurSDI and all the compared methods are applied and evaluated on the blurred images. The results are shown in Figure 7 and the quantitative results are shown in Table 4. It can be seen that the proposed DeblurSDI outperforms all the compared methods in all the optical aberration situations.

Table 4. Quantitative results of optical aberration correction performance (PSNR/SSIM) on four datasets.

	Phase-Only[29]	FFT-ReLU Deblur[2]	SelfDeblur[32]	FastDiffusionEM[21]	DeblurSDI (Ours)
Levin[22]	15.52/0.3722	19.57/0.5655	18.13/0.4706	18.68/0.5085	28.36/0.8598
Cho[5]	22.04,0.8272	23.07/0.8573	20.69/0.7793	15.66/0.4768	25.60/0.9234
Kohler[19]	27.37/0.7889	29.89/0.8358	20.76/0.5409	19.83/0.5242	32.07/0.9061
FFHQ[17]	26.31/0.7703	23.21/0.6942	19.65/0.5591	17.90/0.4508	33.00/0.9343

Motion Deblur: Figure 8 shows deblurring results of different methods on the FFHQ dataset [17]. For each estimated image, the recovered kernel is displayed at the top-left corner, and three zoomed-in regions highlight fine details. As observed, FastDiffusionEM [21] performs the worst. Despite being pre-trained on the FFHQ dataset, its performance remains poor: the estimated kernels degenerate into trivial point- or line-like structures. Without reliable kernel estimation, even a strong image prior cannot yield satisfactory deblurring. In contrast, other four optimization-based methods produce visibly superior results.

Figure 8. Deblurring results on the FFHQ dataset [17].

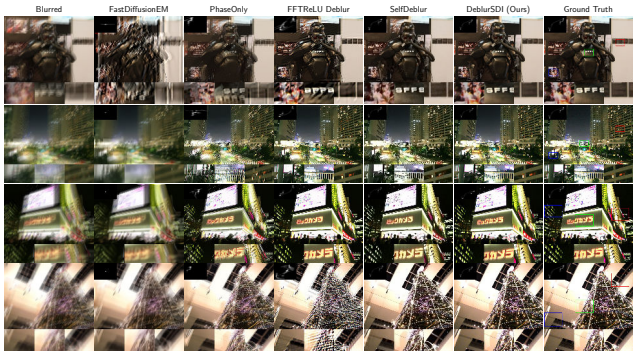

Figure 9. Deblurring results on the Cho dataset [5].

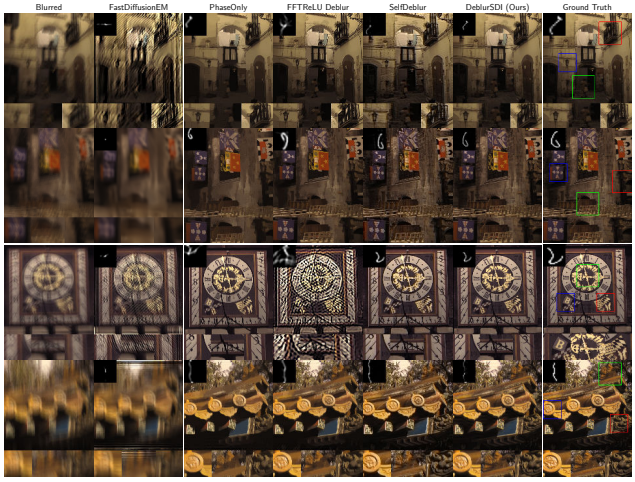

Figure 10. Deblurring results on the Kohler dataset [19].

Figure 11. Deblurring results on the Levin dataset [22].

PhaseOnly [29] and FFTReLU Deblur [2] provide good results in some cases, but lack of robustness when facing various blur kernels. SelfDeblur [32] shows the most promising performance out of previous methods, yet still suffers from poor generalizability, reconstruction shiftting and color distortion. The proposed DeblurSDI method consistently outperforms all compared methods, achieving significant improvements in the long-standing shiftting and robustness issues of blind deblurring techniques. Moreover, DeblurSDI is also able to recover accurate blur kernels, which gurantees the performance of the method. More results of other datasets are shown in Figure 9, 10, and 11.

Table 5. Quantitative results of blind motion deblurring performance (PSNR/SSIM).

	Phase-Only[29]	FFT-ReLU Deblur[2]	SelfDeblur[32]	FastDiffusionEM[21]	DeblurSDI (Ours)
Levin[22]	20.68/0.6061	15.56/0.3845	25.06/0.7301	16.55/0.4005	31.85/0.7911
Cho[5]	19.89/0.6746	18.73/0.6546	20.37/0.6844	15.39/0.4687	28.73/0.8859
Kohler[19]	28.23/0.8092	25.33/0.7140	21.97/0.5995	18.85/0.4813	29.17/0.7653
FFHQ[17]	25.80/0.7904	21.71/0.6579	19.82/0.5563	15.59/0.3592	33.90/0.9064

We compared our DeblurSDI method with several other blind deblurring approaches, including Phase-Only [29], FFT-ReLU Deblur [2], SelfDeblur [32], and FastDiffusionEM [21] and calculated Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) on each dataset. The quantitative results are summarized in Table 5. This demonstrates the effectiveness and generalizability of our self-diffusion approach in recovering sharp images and accurate blur kernels.

6. Discussion and Conclusion

Through solving blind inverse problem to recover the image under optical aberrations and motion blur, we investigate a novel and robust solution for image deblurring. The plausible performance and robustness of our proposed method are brought in by the noise schedule, which is a key component of the self-diffusion framework. We astutely observed this characteristic and boldly extended it to solving extremely unstable joint optimization inverse problems. This greatly solves the problem of unstable solutions or collased in trivial solutions. Of course, because we simultaneously constrained both the image and the kernel with neural networks and added a hierarchical noise schedule, our method requires a longer runtime compared to others. However, this time investment is worthwhile because we significantly improved the realism of the recovered image across the continuous spectrum, especially high-frequency details.

This paper presents a novel self-diffusion-based approach for optical aberration correction and motion deblurring, which we call DeblurSDI. Our method leverages the self-diffusion principle to recover accurate blur kernels and sharp images in a single framework. Experimental results on four benchmark datasets as well as six simulated optical aberrations show that DeblurSDI consistently outperforms other blind deblurring methods on various datasets, demonstrating its effectiveness and generalizability in recovering sharp images and accurate blur kernels.

References

[1] Sebastian Acuña, Mayank Roy, Luis E. Villegas-Hernández, Vishesh K. Dubey, Balpreet Singh Ahluwalia, and Krishna Agarwal. Deriving high contrast fluorescence microscopy images through low contrast noisy image stacks. *Biomed. Opt. Express*, 12(9):5529–5543, 2021. 2

[2] Abdul Mohaimen Al Radi, Prothito Shovon Majumder, and Md Mosaddek Khan. Blind image deblurring with fft-relu sparsity prior. In *WACV*, pages 3447–3456, 2025. 1, 7, 8

[3] Muthuvel Arigovindan, Joshua Shaevitz, John McGowan, John W. Sedat, and David A. Agard. A parallel product-convolution approach for representing depth varying point spread functions in 3d widefield microscopy based on principal component analysis. *Opt. Express*, 18(7):6461–6476, 2010. 2

[4] Max Born and Emil Wolf. *Principles of Optics*. Cambridge University Press, 7 edition, 1999. 1, 2

[5] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. In *ACM SIGGRAPH Asia 2009 papers*, pages 1–8. 2009. 1, 2, 4, 7, 8

[6] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *ICLR*, 2023. 2

[7] Michael W Davidson and Mortimer Abramowitz. Optical microscopy. *Encyclopedia of imaging science and technology*, 2(1106-1141):120, 2002. 1, 2

[8] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T Roweis, and William T Freeman. Removing camera shake from a single photograph. In *Acm Siggraph 2006 Papers*, pages 787–794. 2006. 1, 2

[9] Yosef Gandelsman, Assaf Shocher, and Michal Irani. "double-dip": unsupervised image decomposition via coupled deep-image-priors. In *CVPR*, pages 11026–11035, 2019. 2

[10] R Gonsalves. Phase retrieval and diversity in adaptive optics. *Optical Engineering*, 21(5):829–832, 1982. 2

[11] Joseph W Goodman and Mary E Cox. Introduction to fourier optics, 1969. 1, 2

[12] Shayan Mohajer Hamidi, En-Hui Yang, and Ben Liang. Coupled data and measurement space dynamics for enhanced diffusion posterior sampling. *arXiv:2510.09676*, 2025. 2

[13] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (ToG)*, 35(6):1–12, 2016. 1, 2

[14] Michael Hirsch, Christian J Schuler, Stefan Harmeling, and Bernhard Schölkopf. Fast removal of non-uniform camera shake. In *ICCV*, pages 463–470, 2011. 1

[15] Harold Horace Hopkins. On the diffraction theory of optical images. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 217(1130):408–432, 1953. 2

[16] ISO/TC 172/SC 7. Ophthalmic optics and instruments — reporting aberrations of the human eye. Standard ISO 24157:2008, International Organization for Standardization, Geneva, Switzerland, 2008. 2

[17] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, pages 4401–4410, 2019. 4, 7, 8

[18] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *NeurIPS*, 35: 23593–23606, 2022. 2

[19] Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *ECCV*, pages 27–40, 2012. 4, 7, 8

[20] Dilip Krishnan and Rob Fergus. Fast image deconvolution using hyper-laplacian priors. In *NeurIPS*, pages 1033–1041, 2009. 1, 2

[21] Charles Laroche, Andrés Almansa, and Eva Coupete. Fast diffusion em: a diffusion model for blind inverse problems with application to deconvolution. In *CVPR*, pages 5271–5281, 2024. 1, 7, 8

[22] Anat Levin, Rob Fergus, Frédo Durand, and William T Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics (TOG)*, 26 (3):70–es, 2007. 4, 7, 8

[23] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, pages 1964–1971, 2009. 1, 2

[24] Guanxiong Luo, Shoujin Huang, and Yanlong Yang. Self-diffusion for solving inverse problems. *arXiv:2510.21417*, 2025. 1, 3

[25] Virendra N Mahajan. Zernike circle polynomials and optical aberrations of systems with circular pupils. *Applied Optics*, 33(34):8121–8124, 1994. 2, 3

[26] Virendra N Mahajan. *Optical Imaging and Aberrations, Part III: Wavefront Analysis*. SPIE Press, Bellingham, Washington, 2013. PM221. 1, 2

[27] Antoine Monod, Julie Delon, and Thomas Veit. An analysis and implementation of the hdr+ burst denoising method. *Image Processing On Line*, 11:142–169, 2021. 1, 2

[28] Robert J Noll. Zernike polynomials and atmospheric turbulence. *Journal of the Optical Society of America*, 66(3): 207–211, 1976. 1, 2

[29] Liyuan Pan, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Phase-only image based kernel estimation for single image blind deblurring. In *CVPR*, pages 6034–6043, 2019. 1, 7, 8

[30] Richard G. Paxman, Timothy J. Schulz, and James R. Fienup. Joint estimation of object and aberrations by using phase diversity. *J. Opt. Soc. Am. A*, 9(7):1072–1085, 1992. 1, 2

[31] Ben C Platt and Roland Shack. History and principles of shack-hartmann wavefront sensing, 2001. 1, 2

[32] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo. Neural blind deconvolution using deep priors. In *CVPR*, pages 3341–3350, 2020. 1, 2, 7, 8

[33] Warren J Smith. *Modern Optical Engineering*. McGraw-Hill, 4 edition, 2007. 1, 2

[34] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *CVPR*, pages 769–777, 2015. 1

[35] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *CVPR*, pages 9446–9454, 2018. 2

[36] W. T. Welford. *Aberrations of optical systems*. Hilger, Bristol, repr. with minor corr. edition, 1989. 2

[37] Oliver Whyte, Josef Sivic, Andrew Zisserman, and Jean Ponce. Non-uniform deblurring for shaken images. In *CVPR*, pages 491–498, 2010. 1

[38] Bartlomiej Wronski, Ignacio Garcia-Dorado, Manfred Ernst, Damien Kelly, Michael Krainin, Chia-Kai Liang, Marc Levoy, and Peyman Milanfar. Handheld multi-frame super-resolution. *ACM Transactions on Graphics (ToG)*, 38(4):1–18, 2019. 2

[39] Zhonghe You and Xindong Zhang. Image restoration based on generative adversarial networks with hybrid attention mechanisms. In *2024 5th International Conference on Computer Engineering and Application (ICCEA)*, pages 1007–1012, 2024. 2