Physics-Driven Turbulence Image Restoration with Stochastic Refinement

Ajay Jaiswal[1*], Xingguang Zhang[2*], Stanley H. Chan[2], Zhangyang Wang[1]

[1]University of Texas at Austin, [2]Purdue University

{ajayjaiswal, atlaswang}@utexas.edu {zhan3275, stanchan}@purdue.edu

Abstract

Image distortion by atmospheric turbulence is a stochastic degradation, which is a critical problem in long-range optical imaging systems. A number of research has been conducted during the past decades, including model-based and emerging deep-learning solutions with the help of synthetic data. Although fast and physics-grounded simulation tools have been introduced to help the deep-learning models adapt to real-world turbulence conditions recently, the training of such models only relies on the synthetic data and ground truth pairs. This paper proposes the Physics-integrated Restoration Network (PiRN) to bring the physics-based simulator directly into the training process to help the network to disentangle the stochasticity from the degradation and the underlying image. Furthermore, to overcome the "average effect" introduced by deterministic models and the domain gap between the synthetic and real-world degradation, we further introduce PiRN with Stochastic Refinement (PiRN-SR) to boost its perceptual quality. Overall, our PiRN and PiRN-SR improve the generalization to real-world unknown turbulence conditions and provide a state-of-the-art restoration in both pixel-wise accuracy and perceptual quality. Our codes are available at https://github.com/VITA-Group/PiRN.

1. Introduction

Atmospheric turbulence (AT) is one of the major sources of degradation in long-range passive imaging systems. It arises due to random spatiotemporal fluctuations in the index of refraction [19, 61]. When accumulating over the distance, turbulence oftentimes leads to degraded image quality with random pixel displacement and blurring [5]. Such degradation is very common in settings when the object distance is long, and the exposure is short [61], causing a substantial drop in the performance of long-range passive imaging and its downstream tasks, such as human recognition, detection, and tracking. Developing restoration meth-

ods to mitigate the turbulence effect is important. However, anisoplanatic turbulence has two major properties: ① the geometric warping and blur are entangled with each other; ② the point spread function is spatially and temporally varying, which makes the problem harder than other image restoration problems.

Turbulence mitigation algorithms have been studied for several decades by the image processing community. Since capturing real-world corrupted and clean image pairs is almost impossible, those conventional methods are model-based [16, 20, 27, 49, 2, 17, 23, 74, 39]. Recently, data-driven approaches have been introduced with synthetic data from various turbulence simulators. Existing deep learning works explored both deterministic and stochastic methods to solve the turbulence mitigation problem. In the deterministic approaches [41, 72, 1, 28, 34, 47, 45, 26, 51, 18], a turbulence mitigation network is trained on a synthetic dataset to minimize pixel-level distortion between the output and ground truth images. The generalization capability of those works is fully limited by the training images. Moreover, the deterministic models learn to "fill up" a probability space that covers all possible clean images, which leads to unnatural output with an average effect. Although adversarial training [28, 34, 51] has been used to alleviate this problem, it could make the model more vulnerable to small perturbation on the input [10, 11]. The stochastic approaches [44, 12] could produce more natural images with certain robustness. However, they are more likely to hallucinate as they do not have a physics-grounded degradation model, and the generative models they used are unconditional diffusion models trained on general datasets. The large domain gap between the testing images and the training set distribution and the lack of a good forward model as a connection oftentimes cause the output of generative models unreliable.

In this work, we propose a two-stage method to improve both the fidelity [43] and perceptual quality [66] for single-frame turbulence mitigation. To improve fidelity, we propose to tightly couple a physics-based turbulence simulator [40, 9] in the training paradigm of our turbulence restoration backbone. Specifically, we re-degrade the reconstructed image and align it with the original input to enforce the con-
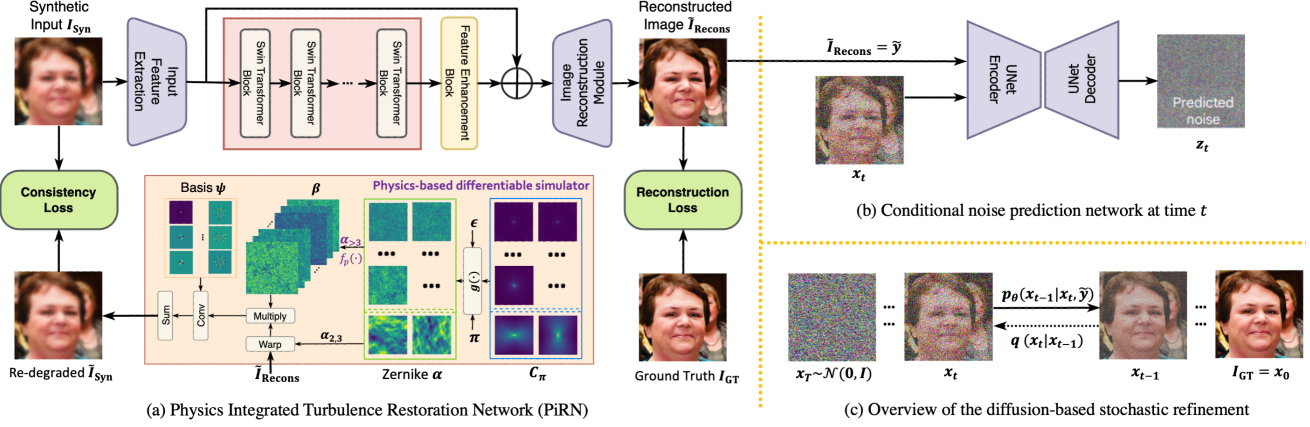
---

*Equal contribution.

Figure 1. The overview of our PiRN with stochastic refinement model. The PiRN (a) training utilizes a physics-based differentiable simulator to map the stored image back to the input to enforce the consistency between the forward and restoration processes. We use the intermediate restoration from PiRN as a reliable condition of a diffusion model at each denoising step (b). We perform 10-20 stochastic iterations (c) to sample a natural image with high perceptual quality from the well-constraint image space.

sistency between image formation and restoration. Our consistency enforcement could facilitate the separation of image semantics and turbulence profiles by injecting the turbulence conditions in the training loop. With the physics-integrated restoration network (PiRN), we experimentally found the above strategy significantly improves the generalization across multiple real-world datasets with varying turbulence strength.

To improve perceptual quality, unlike [28, 34, 51], which used the adversarial training, we build a stochastic posterior sampler by training a conditional denoising diffusion probabilistic model (DDPM) [25] that takes the output of the PiRN as a reliable condition and generates high-quality natural images within the constraint sample space. Our experiments found direct conditioning on the degraded images could destabilize diffusion training due to the failure to capture the complicated degradation. We take advantage of the deterministic reconstructor PiRN and use its restored image as a more constrained condition. Our divide-and-conquer strategy suggests PiRN focus on handling turbulence strength variations and facilitating *stochastic refinement (SR)* to mitigate the gap between the rough restoration and real-world image distribution.

Our overall framework PiRN-SR enjoys the benefit of high fidelity and perceptual quality while adapting to a wide range of physical attributes of turbulence degradation (i.e., generalization). PiRN-SR combines PiRN with stochastic refinement in a plug-and-play fashion. It iteratively performs 10-20 denoising steps to significantly boost *perceptual quality and robustness to additional perturbation* without compromising the fidelity. Our primary contributions can be briefly summarized as follows:

☐ We show how a fully differentiable physics-based simulator can be tightly coupled with the DL restoration paradigm. We improve the generalization to multiple

physical attributes of turbulence (distance to object, camera settings, etc.) with the help of our diligently curated synthetic data generation strategy.

☐ Our proposed framework PiRN-SR demonstrates how a carefully trained conditional diffusion model can be used as a plug-and-play stochastic refiner to generate high perceptual quality results from turbulence-degraded input images at marginal inference overhead.

☐ Extensive experiments and ablation across synthetic and multiple real-world popular turbulence benchmarks demonstrate our method achieves state-of-the-art reconstruction quality in both pixel-wise accuracy and perceptual quality.

2. Method

2.1. Zernike-based turbulence forward model

The forward model of the degradation caused by atmospheric turbulence to an image $\boldsymbol{I}$ is [24]:

$$\boldsymbol{O} = \boldsymbol{h} \circledast \boldsymbol{I} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{I} \in \mathbb{R}^{H \times W}$ is the clean input image, $\boldsymbol{O} \in \mathbb{R}^{H \times W}$ is the captured image degraded by turbulence, $\circledast$ is the convolution, is the 2-D spatial position of pixels and $\boldsymbol{n} \in \mathbb{R}^{H \times W}$ is the Gaussian random noise. Since the point spread function (PSF) $\boldsymbol{h}$ varies with , the degradation is *spatially varying*. According to [5], may be written as a function of pixel-shifting $\mathcal{T}$ (closely associated with "tilt" in the optical literature) and blur $\mathcal{B}$ in the strict order:

$$\boldsymbol{O} \approx \mathcal{B} \circledast \mathcal{T}(\boldsymbol{I}) + \boldsymbol{n}. \tag{2}$$

The recent Zernike-based turbulence simulators [8, 9] model the degradation as a wide sense stationary (WSS)

random field of the phase distortion, which is represented by Zernike polynomials $\{_i\}$ [48] as a basis, with coefficients $_{,i}$, where $i \in \{1, 2, 3, \cdots, 36\}$. Given a set of camera and atmospheric *protocol* $\boldsymbol{\pi}$, the autocorrelation $\boldsymbol{C_\pi}$ of the Zernike random field can be drawn by this simulator. From $\boldsymbol{C_\pi}$ and Gaussian white noise $\boldsymbol{\epsilon} \in \mathbb{R}^{H \times W \times 36}$, the WSS Zernike random field can be generated via Fourier Transform. We denote this function by $= g(\boldsymbol{C_\pi}, \boldsymbol{\pi}, \boldsymbol{\epsilon})$

Among all 36 coefficients, $i = 1$ denotes the current component, $i = 2, 3$ controls the $\mathcal{T}$ by a constant scale, and the rest high order Zernike coefficients contribute to the blur effect. The phase distortion corresponding to high-order Zernike coefficients can be efficiently translated to 100 PSFs basis $\boldsymbol{\psi}$ and coefficients $\boldsymbol{\beta}$ via the phase-to-space (P2S) [40] transform $\boldsymbol{\beta} = f_p(_{,\{i \geq 4\}})$. The spatially varying blur kernel can be finally written as:

$$\boldsymbol{\mathcal{B}} \approx \sum_{k=1}^{100} \boldsymbol{\beta}_{,k} \boldsymbol{\psi}_k = \sum_{k=1}^{100} f_p(_{,\{i \geq 4\}})_k \boldsymbol{\psi}_k. \quad (3)$$

For the detailed expression of $g(\cdot)$ and $f_p(\cdot)$, we refer readers to read section I of the supplementary material.

2.2. Physics-integrated restoration network (PiRN)

Since g can be implemented by spectrum decomposition, f_p is a small neural network [40], and ψ are fixed basis kernels, the forward model conditioned on the turbulence protocol $\boldsymbol{\pi}$ and random seed $\boldsymbol{\epsilon}$ is differentiable. This property suggests the forward process can be embedded into the training loop and effectively provide the turbulence prior through gradients, facilitating the invariance of the reconstruction towards the stochasticity of the degradation.

The degradation profile controlled by $\boldsymbol{\epsilon}$ is different for each frame, which makes anisoplanatic turbulence get spatially and temporally varying distortions. Conventional CNN-based general image restoration methods applying the fixed kernels on all spatial locations may not be adequate to solve the location adaptive problem [41]. Motivated by the *necessity of input-adaptive and location-adaptive filtering effect*, our physics-integrated restoration network (PiRN) uses a transformer-based network to capture and recover the spatially- and instance-varying turbulence effects.

Overall, PiRN architecture is composed of a Swin-based deep backbone module, a convolution-based image reconstruction module, and a physics-based differentiable forward model, as described in detail in the following.

2.2.1 Phase-to-Space differentiable forward process

The primary contribution to PiRN design is the integration of the Phase-to-Space differentiable turbulence simulator in the training paradigm. Conventional turbulence mitigation networks [41, 72, 28, 1, 45] trained their models only with

the low-quality and reference high-quality pairs, the generalization capability of their networks only relies on their synthetic method and training data, hence inept at adapting to varying turbulence protocols and profiles. [41, 35, 18] proposed to re-map the restored image to the degraded image. However, the re-mapping of [41] is based on an empirical design without a clear physical meaning. [35, 18] are multi-frame methods, their reconstruction requires minutes or hours of refinement, and the adaptation to real-world cases is yet highly limited.

In PiRN, we propose to integrate the well-established physics of turbulence described in section 2.1 and explore its experimental benefits on unseen turbulence protocols in both synthetic and real-world datasets. More specifically, we store the degradation protocol $\boldsymbol{\pi}$ and random seed $\boldsymbol{\epsilon}$ along with the degraded image $\mathbf{I}_{\text{Syn}}$ in the synthetic data generation stage. During training, the degraded image is first restored by the PiRN backbone and reconstruction module, the restored image $\widetilde{\mathbf{I}}_{\text{Recons}}$ is then passed through the simulator with $\boldsymbol{\pi}$ and $\boldsymbol{\epsilon}$ to re-degrade the original input to $\widetilde{\mathbf{I}}_{\text{Syn}}$. Precisely, the function of this module $\mathbf{SIM}(\cdot)$ can be summarized as:

$$\underbrace{\{\widetilde{\mathbf{I}}_{\text{Recons}}, \boldsymbol{\pi}, \boldsymbol{\epsilon}\}}_{\substack{\text{Reconstructed Clean PiRN Output} \\ \text{and Data Synthesis Protocol}}} \rightarrow \mathbf{SIM}(\cdot) \rightarrow \underbrace{\widetilde{\mathbf{I}}_{\text{Syn}}}_{\substack{\text{Re-degraded} \\ \text{Output Image}}} \quad (4)$$

We force $\mathbf{I}_{\text{Syn}}$ and $\widetilde{\mathbf{I}}_{\text{Syn}}$ to be aligned using the ℓ_1 loss. Since this remapping enforces the restoration to be consistent with the degradation process, we call it *consistency loss*. During training, by gradient descent, the turbulence prior from the simulator can be injected into the network and help enhance the invariance of the turbulence effect. Despite its simplicity, we found it significantly improves the adaptability across multiple real-world datasets with varying turbulence strength.

2.2.2 Swin-based deep feature backbone

Swin Transformer [38] has recently shown great success in modeling long-range dependency with shifted window schemes. Although it has been explored for many image restoration works [36, 69, 64, 7], its potential for turbulence mitigation which requires spatially varying operations is still unexplored. As shown in Figure 1, the deep feature backbone of PiRN architecture is a sequence of Swin Transformer blocks (RSTB). The RSTB utilizes several Swin Transformer layers for local attention and cross-window interaction. Finally, for feature enhancement, we add a convolution layer at the end of RSTB blocks for feature enhancement and use a residual connection to provide a shortcut for feature aggregation [36]. Before RSTB, the input feature extraction uses convolution layers to extract shallow

features. Those features preserve low-frequency information of the image, induce convolutional inductive bias in the early stage, and improve the representation learning capability of transformer blocks [67]. The details of the architecture are provided in the supplementary material.

2.2.3 Image Reconstruction Network

Our upsample image reconstruction module restores the high-quality image by decoding the deep features generated by the Swin-based backbone with reference to shallow features from input feature extraction. With the long skip connection, our network can transmit the low-frequency information directly to the reconstruction network, which can help the deep feature extraction module focus on high-frequency information and stabilize training. Our Image reconstruction module is a sequence of convolutional layers with LeakyReLU that projects enriched features back to low dimension feature map corresponding to the reconstructed clean image $\tilde{\mathbf{I}}_{\text{Recons}}$. Precisely, the role of the reconstruction module $\mathbf{REC}(\cdot)$ can be summarized as:

$$\underbrace{\{\mathbf{D}_{\mathbf{I}_{\text{Syn}}}, \mathbf{S}_{\mathbf{I}_{\text{Syn}}}\}}_{\substack{\text{Deep and Shallow Features of} \\ \text{Input Image } \mathbf{I}_{\text{Syn}}}} \rightarrow \mathbf{REC}(\cdot) \rightarrow \underbrace{\tilde{\mathbf{I}}_{\text{Recons}}}_{\substack{\text{Reconstructed Clean} \\ \text{Output Image}}} \quad (5)$$

PiRN optimization requires the joint optimization of reconstruction loss with ground truth and the consistency loss as shown in Figure 1. We formulate the loss as follows:

$$\mathcal{L}_{total} = \alpha \cdot ||\mathbf{I}_{\text{GT}}, \tilde{\mathbf{I}}_{\text{Recons}}||_1 + (1 - \alpha) \cdot ||\mathbf{I}_{\text{Syn}}, \tilde{\mathbf{I}}_{\text{Syn}}||_1 \quad (6)$$

2.3. Diffusion-based stochastic refinement

Although the turbulence simulation tool is physics-grounded, the domain gap still exists between synthetic and real-world turbulence. Besides, the restored images from the PiRN also suffer from the problem of deterministic methods. In order to overcome this and make the output more natural and closer to the target dataset distribution, we use Denoising Diffusion Probabilistic Models (DDPM) [57, 25, 60] as a stochastic sampler. DDPMs are generative models with a Markov chain that constructs samples iteratively from a joint distribution:

$$p_\theta(_{0:T}) = p_\theta^{(T)}(_T) \prod_{t=0}^{T-1} p_\theta^{(t)}(_t|_{t+1}) \quad (7)$$

Where $_T$ is random Gaussian noise and $_0$ is the ground truth. In our framework, the high-quality image space follows a posterior distribution conditioned on the pre-restored image $\tilde{} = \tilde{\mathbf{I}}_{\text{Recons}}$ from our deterministic reconstructor:

$$p_\theta(_{0:T}|\tilde{}) = p_\theta^{(T)}(_T) \prod_{t=0}^{T-1} p_\theta^{(t)}(_t|_{t+1}, \tilde{}) \quad (8)$$

This conditional diffusion sampler could have an unconditional variational inference distribution:

$$q(_{1:T}|_0) = q^{(T)}(_T|_0) \prod_{t=1}^{T-1} q^{(t)}(_t|_{t+1}, _0) \quad (9)$$

Whose conditional transition distribution of the forward and backward diffusion process can be modeled by Gaussian parameterization. For $p_\theta^{(t)}$ and $q^{(t)}$, we reduce the evidence lower bound (ELBO) objective by solving the following noise prediction loss [25]:

$$\mathcal{L} = \sum_{t=1}^{T-1} \gamma_t \mathbb{E}_{(_0, _t, \tilde{})} \left[||_t -_\theta (_t, \tilde{}, t)||^2 \right] \quad (10)$$

Where $_t$ is the noise added to the forward diffusion process at time t, $_t = \sqrt{\overline{\alpha}_t}_0 + \sqrt{1 - \overline{\alpha}_t}_t$ is the noised clean image at denoising step t, and $_\theta(_t, \tilde{}, t)$ is a network with learnable parameters to predict the noise $_t$. $\overline{\alpha}_t$ is a fixed scalar to control the diffusion schedule [25].

We perform gradient descent to train $_\theta(_t, \tilde{}, t)$ like [54]. When $_\theta(_t, \tilde{}, t)$ is trained, the restoration can be converted to a diffusion posterior sampler $\hat{}_0 \sim p_\theta(_0|\tilde{})$ defined in Eq 8. As shown in Figure 1, along with the ground truth $\mathbf{I}_{\text{GT}}$, we pair one restored image from PiRN. Our experiments found direct conditioning on the degraded images ($\mathbf{I}_{\text{Syn}}$) could destabilize diffusion training because the turbulence profiles are random, and the diffusion model can't capture the complicated degradation. During inference with PiRN-SR, we iteratively perform 10-20 denoising steps with our diffusion sampler upon the restoration output of PiRN to boost its perceptual quality.
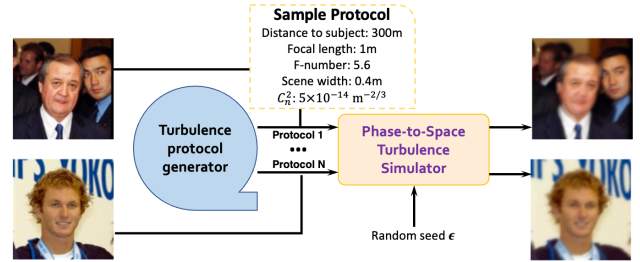

Figure 2. Overview of our synthetic data (PiRN-Syn) simulation.

2.4. PiRN-Syn: Synthetic Data Generation Strategy

Because of the scarcity of real clean-distortion image pairs, data-driven approaches have to rely on turbulence simulators for synthetic data. Despite there have been some efforts in turbulence simulation in the image processing community [33, 46, 31, 41], they highly overlook the fact that real-world turbulence profile is affected by the aperture and focal length of the imaging system, distance to the object, field of view, wavelength, and other environmental conditions (temperature, humidity, wind speed, and so on).

Distance (m)	Focal length (m)	F-number	Scene width (m)	$C_n^2 (10^{-14} \times m^{-2/3})$
[200, 400]	[1, 2]	{8, 11}	[0.2, 0.5]	[3, 7]
		{5.6, 8, 11}	[0.5, 1]	[6, 30]
[400, 600]	[1, 2.5]	{8, 11, 16}	[0.4, 0.8]	[2, 6]
		{5.6, 8, 11}	[0.8, 1.5]	[6, 30]
[600, 800]	[1, 3]	{11, 16}	[0.5, 1.2]	[2, 5]
		{8, 11}	[1.2, 2]	[5, 30]

Table 1. Parameter range, where $[a, b]$ means uniform sampling from continuous range (a, b), and {} indicates uniform sampling from a discrete set, all rows are chosen with identical probability.

This ignorance restricted the generalization capabilities of those methods for unknown real-world degradation.

Following section 2.1, we provide an *easy-to-follow synthetic data generation strategy* capturing a wide variety of camera parameters and atmospheric conditions (represented by the measurable indicators C_n^2) using the improved Zernike-based simulator [8, 9]. Our synthetic data generation strategy (Table 1) has been curated based on the analysis to cover the potential short-exposure turbulence profiles observed in around 1000+ hours of video footage of approximately 1,000 subjects within different environmental conditions and camera settings [15]. When setting the parameters, we first select the distance and field of view, then the focal length and f-number ranges could be determined based on real-world camera models. We choose the C_n^2 range to set the turbulence effect to be neither too strong nor weak. PiRN-Syn consists of 100,000 degraded images for training and 50,000 degraded images for testing, which are generated using different synthesis protocols from Table 1 with 2000 and 1000 unique instances from [73].

In addition, to study how our new turbulence mitigation algorithm performs under different conditions, we classify the turbulence strength into multiple levels: *weak, medium, strong*. The details of how we set the threshold is provided in the supplementary material.

3. Experimental Setup

3.1. Datasets and Training Setup

To train the PiRN network, we utilized the PiRN-Syn synthetic dataset, which is generated by our curated data generation strategy as discussed in Section 2.4. The dataset comprises 100,000 degraded images for training and 50,000 degraded images for testing, generated using various synthesis protocols from Table 1, with 2000 and 1000 unique instances from [73], respectively. The diffusion network is trained using high-quality images from [29], paired with the restored output of its degraded version from PiRN. For training PiRN, we used a learning rate of $1e - 4$ and employed the cosine annealing scheduler to gradually decrease the learning rate over 100,000 iterations. The training of the diffusion network closely followed the optimal settings in [54]. In the beginning, during the first 5000 epochs of PiRN training, we set the scaling factor α in Equation 6 to 1 to provide an easy optimization landscape for PiRN, and

then set it to 0.9 for the remaining training iterations. More implementation details can be found in the supplementary.

3.2. Evaluation Protocol

Our proposed framework incorporates a strong backbone, differentiable turbulence forward model, and a diffusion posterior sampler to achieve high-quality restoration. In order to validate the effectiveness of our design, we have designed experiments to answer several key questions:

RQ1: How does the simulator-integrated PiRN design perform and generalize compared to classical state-of-the-art (SOTA) restoration architectures in both synthetic and real-world settings?

RQ2: Does simulator-in-loop training enhance the restoration adaptability of the proposed framework to varying levels of turbulence strength?

RQ3: How effective is posterior diffusion sampling in improving the perceptual quality of the restored images without impacting the standard PiRN metrics (PSNR, SSIM)?

RQ4: How effective is PiRN/PiRN-SR for ad-hoc downstream applications (e.g., detection and recognition)?

For our experiments, we have used our PiRN synthetic test set, along with a variety of real-world turbulence benchmark datasets, including OTIS [21], CLEAR[1], Heat Chamber and Turbulence Text Data [41]. We trained all baseline methods using identical settings and datasets, relying on their official GitHub implementation to ensure a fair comparison. Further details regarding our evaluation datasets can be found in the supplementary material.

3.3. Simulator Integrated PiRN design and state-of-the-art restoration methods

In this section, our main focus is to address research questions RQ1 and RQ2 by evaluating the advantages of integrating the turbulence forward model into the restoration training process to align with the synthetic data generation. To answer RQ1, which aims to determine the necessity of PiRN design in terms of performance and generalization, we conducted a comparative study by assessing the performance of PiRN with several state-of-the-art classical restoration methods. The results, presented in Table 2, demonstrate significant benefits of PiRN design on our synthetic test set and a real-world turbulence benchmark dataset Heat Chamber, in comparison to other baselines. It is noteworthy that our design outperforms general image restoration models significantly and also achieves higher performance than the recent turbulence-based design TurbNet [41], with a margin of $+1.68$ (PSNR) and $+0.78$ (PSNR) on our synthetic test set and Heat Chamber dataset, respectively. Moreover, Figure 3 depicts the qualitative performance of PiRN with and without our simulator integration on real-world turbulence benchmark CLEAR
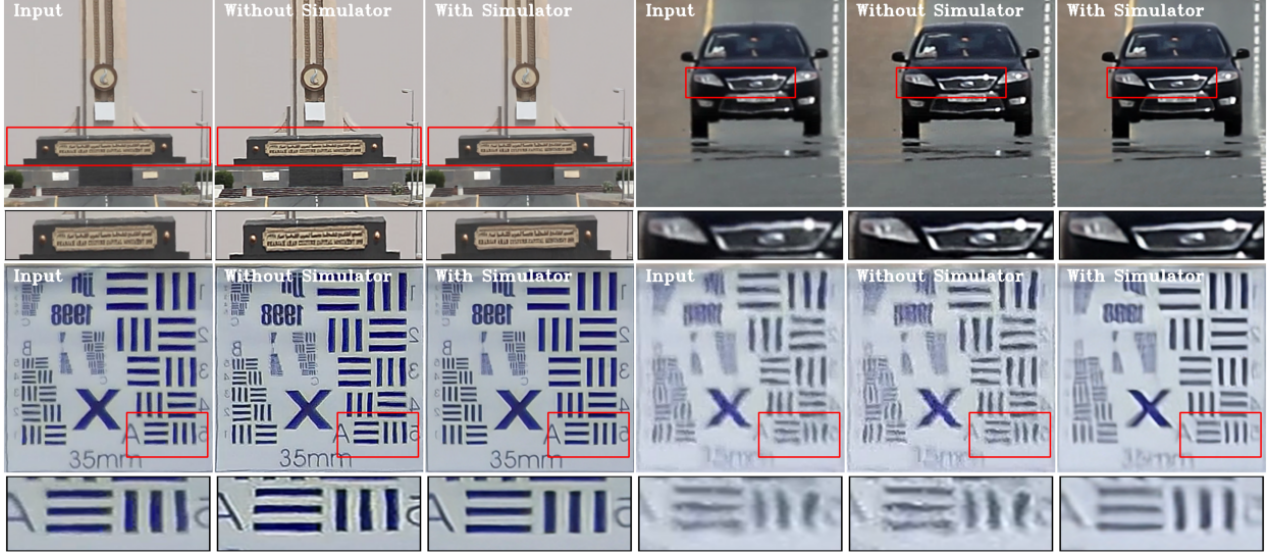
Figure 3. Qualitative performance of PiRN with and without our simulator integration on real-world turbulence benchmark CLEAR (Row1) and OTIS (Row2). Note that both models are trained using exactly the same PiRN-Sync train dataset for a fair comparison. Results in row 2 represent the ability of PiRN to generalize to different turbulence strengths.

Figure 4. Qualitative performance comparison of our proposed method on synthetic PiRN-Syn (test) dataset wrt. SOTA baselines.

and OTIS, trained on the same PiRN-Sync dataset. It is evident that our design plays a significant role in modeling the turbulence degradation operator and generalizes very well on the real-world unseen degradation.

| Method | PiRN-Syn (Test) | | Heat Chamber | |
	PSNR (↑)	SSIM (↑)	PSNR (↑)	SSIM (↑)
TDRN [68]	19.48	0.5288	18.42	0.6424
MPRNet [42]	21.93	0.5819	18.12	0.6379
Uformer [65]	22.20	0.6133	18.68	0.6577
Restormer [70]	22.45	0.6274	19.12	0.6840
SwinIR [36]	22.67	0.6301	19.43	0.6901
TurbNet [41]	23.72	0.6749	19.76	0.6934
Ours [PiRN]	25.40	0.7198	20.54	0.7102
Ours [PiRN-SR]	**25.61**	**0.7204**	**20.59**	**0.7115**

Table 2. Performance comparison of our proposed method wrt. SOTA classical methods on our synthetic test set and real-world Heah Chamber[41] dataset.

Furthermore, we investigate the generalization capability of PiRN for different turbulence strengths (RQ2). Table 3 presents the results of our evaluation, which confirm the ability of our proposed simulator-integrated design (PiRN) to handle different levels of turbulence strength smoothly in our synthetic test set. PiRN consistently outperforms TurbNet by a margin of $+0.81, +0.77$, and $+1.12$ (PSNR), with significant benefits observed for strong degradation strength. The qualitative evaluation of the turbulence strength adaptation ability of PiRN on the real-world OTIS benchmark is also presented in Figure 3 (Row 2), which is consistent with our synthetic quantitative evaluation.

| Method | Weak | | Medium | | Strong | |
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
TDRN [68]	26.61	0.703	23.44	0.659	21.64	0.591
MPRNet [42]	27.01	0.716	24.76	0.672	22.98	0.620
Uformer [65]	27.76	0.721	25.30	0.687	23.54	0.633
Restormer [70]	27.91	0.729	25.68	0.689	23.56	0.638
SwinIR [36]	28.02	0.736	25.96	0.691	23.87	0.645
TurbNet [41]	28.31	0.759	26.07	0.738	23.90	0.649
Ours [PiRN]	29.12	0.763	26.84	0.747	25.02	0.662
Ours [PiRN-SR]	**29.15**	**0.766**	**26.86**	**0.747**	**25.07**	**0.665**

Table 3. Performance comparison of our proposed method across different turbulence strength wrt. other baselines.
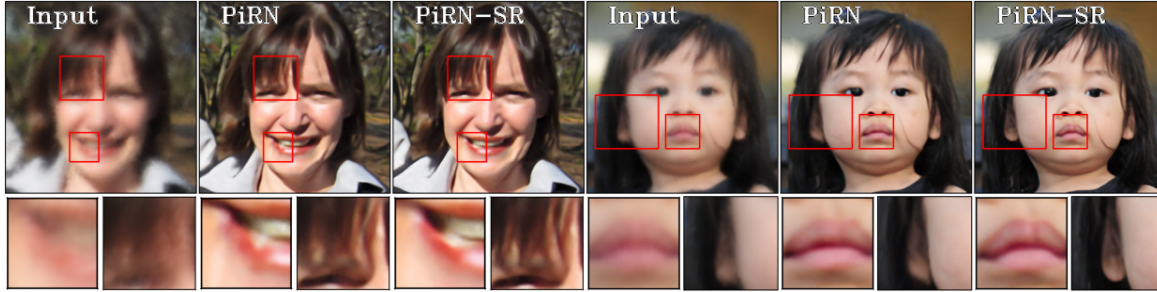
Figure 5. Qualitative performance benefits of using posterior diffusion sampling in improving the perceptual quality of the restored image. Our stochastic refinement module can be observed minimizing the gap between the rough restoration and real-world image distribution.
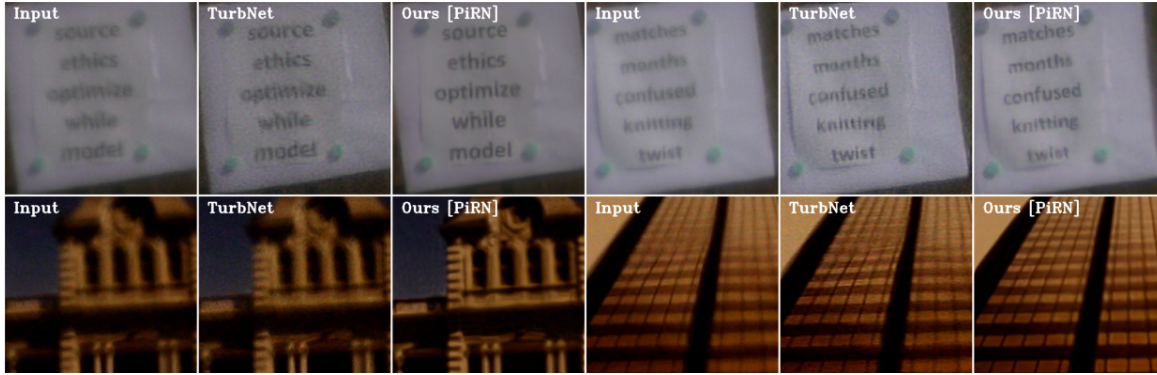

Figure 6. Qualitative performance of our proposed PiRN network on real-world Turbulence Text and Heat Chamber [39] datasets.

	Raw Input	SwinIR[36]	TurbNet[41]	**Ours [PiRN]**
AWDR(↑)	0.623	0.740	0.758	**0.776**
AD-LCS(↑)	5.076	7.002	7.314	**7.498**

Table 4. Performance comparison of state-of-art restoration baselines with respect to TurbNet on the Turbulence Text Dataset.

3.4. The importance of stochastic refinement

In this section, we investigate the effectiveness of posterior diffusion sampling in improving the perceptual quality of PiRN outputs (RQ3). Figure 5 visually demonstrates the benefits of recursively performing 15 denoising steps using the DDIM sampler on the restoration output of PiRN. It is evident that these inexpensive refinement iterations can considerably improve the perceptual quality of the PiRN outputs, making them more natural. To further quantify the perceptual improvements, we utilize popular perceptual metrics such as NIQE, NRQM, and LPIPS. Table 5 presents the performance of PiRN-SR compared to PiRN and other high-performing baselines. Our results indicate that integrating the diffusion sampler with the PiRN network significantly enhances the perceptual performance across all the above metrics on our PiRN-Syn test set compared to the state-of-the-art TurbNet baseline. Importantly, these benefits can be obtained in a plug-and-play fashion, depending on resource availability, without compromising the standard pixel-wise performance metrics such as PSNR and SSIM (Table 2). Specifically, PiRN-SR shows an improvement of ~ 1.00 (NIQE), ~ 0.643 (NRQM), and ~ 0.023 (LPIPIS)

over PiRN with only marginal computational cost.

	SwinIR[36]	TurbNet[41]	PiRN	PiRN-SR
NIQE (↓)	7.4914	7.0163	<u>6.5852</u>	**5.5847**
NRQM (↑)	3.6423	3.7790	<u>3.9810</u>	**4.6239**
LPIPIS (↓)	0.4585	0.4369	<u>0.4204</u>	**0.3978**

Table 5. Quantative perceptual performance benefits of PiRN-SR wrt. PiRN and other baselines on PiRN-Syn (test) dataset.

3.5. Benefits of PiRN/PiRN-SR for ad-hoc downstream applications

In this section, we investigate the potential benefits of our proposed restoration network for ad-hoc downstream tasks (RQ4). We adopted the idea of using the performance of high-level vision tasks, namely text recognition task [41] and face detection [71], as an evaluation metric to validate the necessity of restoration. *Firstly*, for the text recognition task, Figure 6 (Row 1) illustrates the qualitative benefits of our PiRN network with respect to SOTA baseline TurbNet. For quantitative evaluation, we used Average Word Detection Ratio (AWDR) and Average Detected Longest Common Subsequence (AD-LCS) metrics introduced in [41] for publicly available OCR detection, and recognition algorithms [56, 62]. Table 4 presents the performance gain by PiRN over the real turbulence degraded text images and their restored version by various state-of-the-art methods. OCR algorithms achieve massive improvements of +0.153 (AWDR) and +2.422 (AD-LCS) when used on images re-
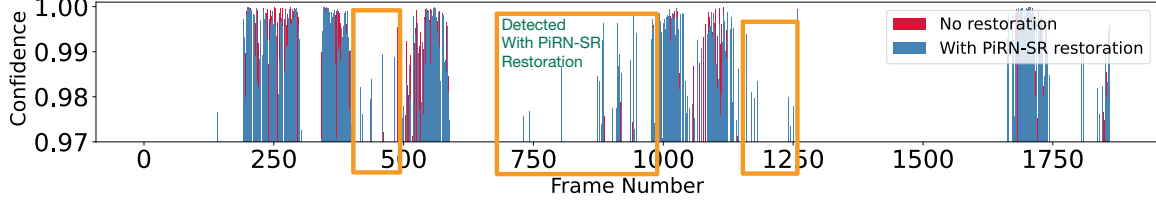
Figure 7. Performance of ad-hoc MTCNN-based face detector [71] on a real-world video captured under atmospheric turbulence (camera distance 400m) [15] wrt. its restored version from PiRN-SR. Each line indicates if a person's face has been detected in that frame with a given confidence score. It can be observed that PiRN-SR restoration significantly helps in improving the performance of MTCNN.
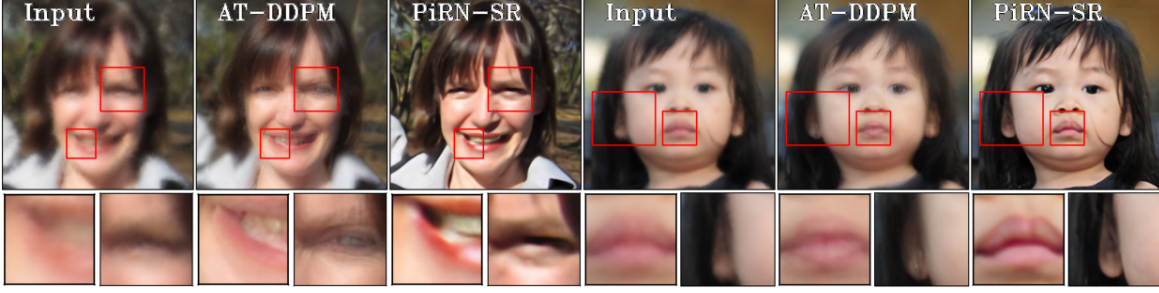

Figure 8. Qualitative performance comparison of AT-DDPM with respect to our proposed framework for atmospheric turbulence mitigation. It can be clearly observed that AT-DDPM hallucinate significantly in comparison with our PiRN-SR output which is assisted by the restoration output of a physics-integrated restoration network (PiRN).

stored by PiRN compared to being used directly on real images from our proposed test dataset. *Secondly*, for the face detection task, we sampled 15 videos captured from varying distances (200m, 400m, 500m) under turbulence from [15]. Figure 7 presents the performance of ad-hoc MTCNN-based face detector [71] on video frames (distance 400m). It can be clearly observed that our PiRN-SR restored videos significantly help in improving the performance capability of MTCNN (+0.034 gain in F1-score). Table 6 illustrates a detailed F1-score comparison of ad-hoc MTCNN detector (at a confidence threshold ≥0.98) on our sampled video subset with and without restoration using recent baselines.

3.6. Comparison of PiRN-SR with AT-DDPM

We compare the performance of our proposed approached PiRN-SR with AT-DDPM, which performs knowledge distillation to transfer class prior information from a network trained for image super-resolution to the network for removing turbulence degradation. During inference, rather than starting from pure Gaussian noise, AT-DDPM begins with noised turbulence degraded images for speed-up in inference times. Figure 8 illustrate the performance comparison of AT-DDPM with respect to our proposed framework for atmospheric turbulence mitigation. It can be clearly observed that AT-DDPM hallucinate significantly in comparison with our PiRN-SR output which is conditioned on high-quality restoration, the output of a physics-integrated restoration network (PiRN). Moreover, instead of completely relying on the diffusion network for complex turbulence restoration, our divide-and-

conquer strategy suggests PiRN focus on handling turbulence strength variations and facilitating stochastic refinement (SR) with 10-20 denoising steps to mitigate the gap between the rough restoration and real-world image distribution. Table 7 presents the performance comparison (F1-score) of MTCNN-based face detection AT-DDPM w.r.t. our proposed method and SwinIR. It can be observed that AT-DDPM, as a stochastic method gets lower performance than the deterministic method SwinIR.

	No Restoration	SwinIR [36]	TurbNet[41]	Ours (PiRN-SR)
200m	0.5810	0.5866	0.5923	**0.6104**
400m	0.4987	0.4985	0.4996	**0.5093**
500m	0.4331	0.4006	0.4510	**0.4648**

Table 6. Performance comparison of ad-hoc MTCNN-based face detector [71] on real-world videos captured under atmospheric turbulence [15] from varying distances. We present the F1-score (at a confidence threshold ≥ 0.98, the higher the better) of the ad-hoc MTCNN detector on videos with and without restoration using recent baselines.

	No Restoration	SwinIR	AT-DDPM Restoration	Ours
400m	0.4987	0.4985	0.4562	0.5093
500m	0.4331	0.4006	0.3997	0.4648

Table 7. Performance comparison (F1-score) of MTCNN face detector on videos from [15] with varying distances.

4. Related Work

Turbulence forward model. The simulation of atmospheric imaging can be roughly classified along a spectrum, with pure numerical optics on one side and vision-based simulation on the other. Optics simulations most often come

in the form of split-step simulation [22, 3, 52, 55], which numerically propagates waves through a set of random phase screens that represent the atmosphere's spatially varying index of refraction. This method is the most accurate but suffers from a slow speed, preventing the development of large datasets. Computer vision simulations have been utilized [74, 32, 4], in which pixels are displaced according to some simple local correlations, followed by an invariant Gaussian blur. Despite the fast speed, the domain gap between the simulation and reality is large, limiting its utility for data-dependent restoration or other downstream tasks.

Recently, Zernike-based turbulence simulation methods have been proposed [8, 40, 9]. This method could match the statistics of optics-based simulation and enjoy realistic visual quality while keeping a real-time data synthesis speed. It has been applied to turbulence mitigation [72, 41] to facilitate the generalization capability of those models.

Learning-based turbulence mitigation. Image restoration for atmospheric turbulence has been studied since the 1990s [50]. Because of the scarcity of data, most methods over the decades are model-based [16, 20, 27, 49, 2, 17, 23, 74, 39]. Learning-based methods have been recently proposed, and a detailed introduction is provided in section 1. Besides deterministic and stochastic, learning-based turbulence mitigation can also be classified into single-frame [41, 34, 45, 26, 6, 44] and multi-frame [28, 47, 37, 1, 18, 72] ones.

Diffusion-based stochastic restoration sampler. Diffusion or score-based models have been widely used for image restoration tasks as strong posterior estimators. In degradation problems with a known forward model, the unconditional diffusion model can be directly used without any re-training [30, 59, 14, 13, 63, 58]. In those works, forward and reverse models of the degradation functions can be inserted at each denoising step to contract the degraded and clean image manifold. Another line of work aims to solve real-world blind restoration problems without the precise forward model [54, 53, 66], it usually requires training a conditional diffusion model with degraded images or latent maps. Diffusion-based restoration sampler could achieve SOTA pixel-wise precision while outperforming deterministic models in perceptual quality by a large margin.

5. Conclusion

This paper proposes Physics-integrated Restoration Network (PiRN) to handle stochastic degradation by atmospheric turbulence. PiRN brings the physics-based simulator directly into the training process of a DL restoration paradigm to help the network disentangle the stochasticity from degradation and improve generalization to multiple physical attributes of turbulence. In parallel to improving fidelity, our extended framework PiRN-SR demonstrates how a carefully trained conditional diffusion model can be used

as a plug-and-play refiner to generate high perceptual quality restoration results from turbulence-degraded input images at marginal inference overhead. Next, we aim to develop a theoretical understanding to explain the generalization benefits identified by our design.

Acknowledgments

References

[1] N. Anantrasirichai. Atmospheric turbulence removal with complex-valued convolutional neural network, 2022. Available online: `https://arxiv.org/abs/2204.06989`.

[2] N. Anantrasirichai, A. Achim, N. G. Kingsbury, and D. R. Bull. Atmospheric turbulence mitigation using complex wavelet-based fusion. *IEEE Transactions on Image Processing*, 22(6):2398 – 2408, Jun. 2013.

[3] J. P. Bos and M. C. Roggemann. Technique for simulating anisoplanatic image formation over long horizontal paths. *Optical Engineering*, 51(10):101704, 2012.

[4] Wai Ho Chak, Chun Pong Lau, and Lok Ming Lui. Subsampled turbulence removal network. *Mathematics, Computation and Geometry of Data*, 1:1 – 33, 2021.

[5] Stanley H. Chan. Tilt-then-blur or blur-then-tilt? clarifying the atmospheric turbulence model. *IEEE Signal Processing Letters*, 29:1833–1837, 2022.

[6] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12299–12310, 2021.

[7] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pages 17–33. Springer, 2022.

[8] Nicholas Chimitt and Stanley H Chan. Simulating anisoplanatic turbulence by sampling correlated zernike coefficients. In *2020 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2020.

[9] Nicholas Chimitt, Xingguang Zhang, Zhiyuan Mao, and Stanley H. Chan. Real-time dense field phase-to-space simulation of imaging through atmospheric turbulence. *IEEE Transactions on Computational Imaging*, 8:1159–1169, 2022.

[10] Jun-Ho Choi, Huan Zhang, Jun-Hyuk Kim, Cho-Jui Hsieh, and Jong-Seok Lee. Evaluating robustness of deep image super-resolution against adversarial attacks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[11] Jun-Ho Choi, Huan Zhang, Jun-Hyuk Kim, Cho-Jui Hsieh, and Jong-Seok Lee. Deep image destruction: Vulnerability of deep image-to-image models against adversarial attacks. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 1287–1293. IEEE, 2022.

[12] Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6059–6069, 2023.

[13] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023.

[14] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.

[15] David Cornett, Joel Brogan, Nell Barber, Deniz Aykac, Seth Baird, Nicholas Burchfield, Carl Dukes, Andrew Duncan, Regina Ferrell, Jim Goddard, et al. Expanding accurate person recognition to new altitudes and ranges: The briar dataset. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 593–602, 2023.

[16] Mauricio Delbracio and Guillermo Sapiro. Removing camera shake via weighted fourier burst accumulation. *IEEE Transactions on Image Processing*, 24(11):3293 – 3307, 2015.

[17] Douglas R. Droege, Russell C. Hardie, Brian S. Allen, Alexander J. Dapore, and Jon C. Blevins. A real-time atmospheric turbulence mitigation and super-resolution solution for infrared imaging systems. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIII*. Proc. SPIE 8355, 2012.

[18] Brandon Y Feng, Mingyang Xie, and Christopher A Metzler. Turbugan: An adversarial learning approach to spatially-varying multiframe blind deconvolution with applications to imaging through turbulence. *IEEE Journal on Selected Areas in Information Theory*, 2023.

[19] D. L. Fried. Optical resolution through a randomly inhomogeneous medium for very long and very short exposures. *Journal of Optical Society of America*, 56(10):1372 – 1379, 1966.

[20] G. Gilles and S. Osher. Wavelet burst accumulation for turbulence mitigation. *Journal of Electronic Imaging*, 25(3):033003, 2016.

[21] Jérôme Gilles and Nicholas B Ferrante. Open turbulent image set (OTIS). *Pattern Recognition Letters*, 86:38 – 41, 2017.

[22] R. C. Hardie, J. D. Power, D. A. LeMaster, D. R. Droege, S. Gladysz, and S. Bose-Pillai. Simulation of anisoplanatic imaging through optical turbulence using numerical wave propagation with new validation analysis. *Optical Engineering*, 56(7):071502, 2017.

[23] Russell C. Hardie, Michael A. Rucci, Alexander J. Dapore, and Barry K. Karch. Block matching and wiener filtering approach to optical turbulence mitigation and its application to simulated and real imagery with quantitative error analysis. *Optical Engineering*, 56(7):071503, 2017.

[24] Michael Hirsch, Suvrit Sra, Bernhard Schölkopf, and Stefan Harmeling. Efficient filter flow for space-variant multiframe blind deconvolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 607 – 614, 2010.

[25] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[26] Matthew A. Hoffmire, Russell C. Hardie, Michael A. Rucci, Richard Van Hook, and Barry K. Karch. Deep learning for anisoplanatic optical turbulence mitigation in long-range imaging. *Optical Engineering*, 60(3), 2021.

[27] Claudia S. Huebner. Turbulence mitigation of short exposure image data using motion detection and background segmentation. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIII*. Proc. SPIE 8355, May 2012.

[28] D. Jin, Y. Chen, Y. Lu, J. Chen, P. Wang, Z. Liu, S. Guo, and X. Bai. Neutralizing the impact of atmospheric turbulence on complex scene imaging via deep learning. *Nature Machine Intelligence*, 3:876 – 884, 2021.

[29] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[30] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, 2022.

[31] Svetlana L. Lachinova, Mikhail A. Vorontsov, Vadim V. Dudorov, Valeriy V. Kolosov, and Michael T. Valley. Anisoplanatic imaging through atmospheric turbulence: brightness function approach. In *Atmospheric Optics: Models, Measurements, and Target-in-the-Loop Propagation*. Proc. SPIE 6708, 2007.

[32] C. P. Lau, Y. H. Lai, and L. M. Lui. Restoration of atmospheric turbulence-distorted images via RPCA and quasiconformal maps. *Inverse Problems*, Mar. 2019.

[33] Chun Pong Lau, Hossein Souri, and Rama Chellappa. Atfacegan: Single face image restoration and recognition from atmospheric turbulence. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 32–39, 2019.

[34] C. P. Lau, H. Souri, and R. Chellappa. Atfacegan: Single face semantic aware image restoration and recognition from atmospheric turbulence. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(2):240 – 251, Feb. 2021.

[35] Nianyi Li, Simron Thapa, Cameron Whyte, Albert W. Reed, Suren Jayasuriya, and Jinwei Ye. Unsupervised non-rigid image distortion removal via grid deformation. In *IEEE/CVF International Conference on Computer Vision*, pages 2522 – 2532, Oct. 2021.

[36] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021.

[37] Feng Liu, Ryan Ashbaugh, Nicholas Chimitt, Najmul Hassan, Ali Hassani, Ajay Jaiswal, Minchul Kim, Zhiyuan Mao, Christopher Perry, Zhiyuan Ren, et al. Farsight: A physics-driven whole-body biometric system at large distance and altitude. *arXiv preprint arXiv:2306.17206*, 2023.

[38] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[39] Z. Mao, N. Chimitt, and S. H. Chan. Image reconstruction of static and dynamic scenes through anisoplanatic turbulence. *IEEE Transactions on Computational Imaging*, 6:1415 – 1428, Oct. 2020.

[40] Zhiyuan Mao, Nicholas Chimitt, and Stanley H. Chan. Accelerating atmospheric turbulence simulation via learned phase-to-space transform. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14739–14748, 2021.

[41] Z. Mao, A. Jaiswal, Z. Wang, and S. H. Chan. Single frame atmospheric turbulence mitigation: A benchmark study and a new physics-inspired transformer model. In *Proc. European Conference on Computer Vision*, 2022.

[42] Armin Mehri, Parichehr B Ardakani, and Angel D Sappa. Mprnet: Multi-path residual network for lightweight image super resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2704–2713, 2021.

[43] Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, and Jaejun Yoo. Reliable fidelity and diversity metrics for generative models. In *International Conference on Machine Learning*, pages 7176–7185. PMLR, 2020.

[44] Nithin Gopalakrishnan Nair, Kangfu Mei, and Vishal M Patel. At-ddpm: Restoring faces degraded by atmospheric turbulence using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3434–3443, 2023.

[45] N. G. Nair and V.M. Patel. Confidence guided network for atmospheric turbulence mitigation. In *IEEE International Conference on Image Processing*, pages 1359 – 1363, 2021.

[46] Nithin Gopalakrishnan Nair and Vishal M. Patel. Confidence guided network for atmospheric turbulence mitigation. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 1359–1363, 2021.

[47] Robert Nieuwenhuizen and Klamer Schutte. Deep learning for software-based turbulence mitigation in long-range imaging. In *Artificial Intelligence and Machine Learning in Defense Applications*. Proc. SPIE 11169, 2019.

[48] R. J. Noll. Zernike polynomials and atmospheric turbulence. *Journal of Optical Society of America*, 66(3):207 – 211, Mar. 1976.

[49] O. Oreifej, X. Li, and M. Shah. Simultaneous video stabilization and moving object detection in turbulence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):450 – 462, Feb. 2013.

[50] J Primot, Gi Rousset, and JC Fontanella. Deconvolution from wave-front sensing: a new technique for compensating turbulence-degraded images. *JOSA A*, 7(9):1598–1608, 1990.

[51] Shyam Nandan Rai and CV Jawahar. Removing atmospheric turbulence via deep adversarial learning. *IEEE Transactions on Image Processing*, 31:2633–2646, 2022.

[52] Michael C. Roggemann, Byron M. Welsh, Dennis Montera, and Troy A. Rhoadarmer. Method for simulating atmospheric turbulence phase effects for multiple time slices and anisoplanatic conditions. *Applied Optics*, 34(20):4037 – 4051, Jul. 1995.

[53] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022.

[54] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[55] J. D. Schmidt. *Numerical simulation of optical wave propagation: With examples in MATLAB*. SPIE Press, 2010.

[56] Baoguang Shi, Xiang Bai, and Cong Yao. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:2298–2304, 2015.

[57] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015.

[58] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023.

[59] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022.

[60] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[61] V. I. Tatarski. *Wave Propagation in a Turbulent Medium*. New York: Dover Publications, 1961.

[62] Zhi Tian, Weilin Huang, Tong He, Pan He, and Yu Qiao. Detecting text in natural image with connectionist text proposal network. In *European Conference on Computer Vision*, 2016.

[63] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations*, 2023.

[64] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17683–17693, June 2022.

[65] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17683–17693, 2022.

[66] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16293–16303, 2022.

[67] Tete Xiao, Mannat Singh, Eric Mintun, Trevor Darrell, Piotr Dollár, and Ross Girshick. Early convolutions help transformers see better. *Advances in Neural Information Processing Systems*, 34:30392–30400, 2021.

[68] Rajeev Yasarla and Vishal M. Patel. Learning to restore images degraded by atmospheric turbulence using uncertainty. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 1694–1698, 2021.

[69] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5728–5739, June 2022.

[70] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022.

[71] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10):1499–1503, 2016.

[72] X. Zhang, Z. Mao, N. Chimitt, and S. H. Chan. Imaging through the atmosphere using turbulence mitigation transformer. Available online: `https://arxiv.org/abs/2207.06465`.

[73] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(6):1452–1464, 2018.

[74] X. Zhu and P. Milanfar. Removing atmospheric turbulence via space-invariant deconvolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):157–170, Jan. 2013.