Learning Latent Transmission and Glare Maps for Lens Veiling Glare Removal

Xiaolong Qian[1,*] Qi Jiang[1,*] Lei Sun[2,†] Zongxi Yu[1] Kailun Yang[3] Peixuan Wu[1]
Jiacheng Zhou[1] Yao Gao[1] Yaoguang Ma[1] Ming-Hsuan Yang[4,5] Kaiwei Wang[1,†]
[1]Zhejiang University [2]INSAIT, Sofia University "St. Kliment Ohridski"
[3]Hunan University [4]University of California, Merced [5]Google DeepMind

(a) Formation of Aberration and Veiling Glare (b) A comparison of restoration performance

Figure 1. Compact optical systems suffer from residual aberrations and veiling glare (a), caused by design-induced blur and stray-light scattering from non-ideal surfaces and coatings. (b) Existing methods fail under this compound degradation: a Computational Aberration Correction (CAC) model [29] retrained on aberration-only data fails on unseen veiling glare, while a cascaded state-of-the-art dehazing model [49] introduces inconsistent artifacts. Our DeVeiler restores a clean image by jointly correcting the compound degradations.

Abstract

Beyond the commonly recognized optical aberrations, the imaging performance of compact optical systems—including single-lens and metalens designs—is often further degraded by veiling glare caused by stray-light scattering from non-ideal optical surfaces and coatings, particularly in complex real-world environments. This compound degradation undermines traditional lens aberration correction yet remains underexplored. A major challenge is that conventional scattering models (e.g., for dehazing) fail to fit veiling glare due to its spatial-varying and depth-independent nature. Consequently, paired high-quality data are difficult to prepare via simulation, hindering application of data-driven veiling glare removal models. To this end, we propose VeilGen, a generative model that learns to simulate veiling glare by estimating its underlying optical transmission and glare maps in an unsupervised manner from target images, regularized by Stable Diffusion (SD)-based priors. VeilGen enables paired dataset generation with realistic compound degradation of optical aberrations and veiling glare, while also providing the estimated latent optical transmission and glare maps to guide the veiling glare removal process. We further introduce DeVeiler, a restoration network trained with a reversibility constraint, which utilizes the predicted latent maps to guide an inverse process of the learned scattering model. Extensive experiments on challenging compact optical systems demonstrate that our approach delivers superior restoration quality and physical fidelity compared with existing methods. These suggest that VeilGen reliably synthesizes realistic veiling glare, and its learned latent maps effectively guide the restoration process in DeVeiler. All code and datasets will be publicly released at https://github.com/XiaolongQian/DeVeiler.

1. Introduction

Driven by the demand for compact and high-performance imaging systems in applications like Augmented Reality (AR/VR) [3, 27] and mobile photography [4, 6], compact optical systems such as single-lens [38] and metalens [46, 62] are becoming increasingly important. However, miniaturized optics inevitably introduce aberrations due to design trade-offs, while low-cost and structural constraints cause surface imperfections and non-ideal coatings, resulting in diffuse stray-light scattering manifested as *veiling glare* (see Fig. 1(a)). The resulting veiling glare manifests as a *widespread and diffuse veil* that reduces image contrast even under normal lighting conditions, in contrast to structured artifacts such as lens flare or ghosting, which typically occur when bright sources are in or near the field of view [24]. Consequently, residual aberra-

[*] Equal contribution. [†] Corresponding authors.

tion and veiling glare frequently coexist, forming a *compound degradation* that severely limits image quality in compact systems. While Computational Aberration Correction (CAC) [2, 4, 5, 14, 16, 20, 21, 26, 28] can mitigate spatial-varying aberrations, it cannot recover the contrast loss induced by diffuse scattering. Moreover, hardware suppression of veiling glare remains difficult, as compact designs leave little room for baffles or advanced coatings. These factors collectively highlight the urgent need for a computational framework capable of modeling and restoring this coupled optical degradation.

Addressing this compound degradation is non-trivial, as the absence of realistic training data remains a key bottleneck. Physically-accurate simulation of veiling glare demands both complete optomechanical models and computationally prohibitive non-sequential ray-tracing, making large-scale data generation impractical [18, 39]. This data scarcity constrains existing paradigms. Classical approaches to veiling glare removal, such as deconvolution-based methods [12, 43] or image decomposition [65], often rely on restrictive assumptions like spatial invariance and tend to fail in low-texture scenes. Modern learning-based methods that rely on simplified 2D synthesis models [42] suffer from a significant domain gap. Moreover, methods designed for visually similar degradations are fundamentally incompatible. Flare removal algorithms [8, 10, 31, 45] are ineffective as they target structured artifacts, not the diffuse nature of veiling glare. Dehazing methods [15, 49, 51, 56] are similarly unsuitable, as their atmospheric scattering models are incompatible with the *depth-independent* physics of glare originating within the lens assembly.

To address this data challenge, recent Stable Diffusion (SD)-based generative models [36, 49, 58] have been explored, yet they often operate as "black-boxes", lacking physical grounding. In contrast, we introduce VeilGen, a physics-informed SD-based model that embeds the formation principles of veiling glare directly into the generative process (Fig. 2). In our setting, paired data are only available for aberration correction, while unpaired degraded images from the compact systems exhibit veiling glare on top of residual aberrations, with physical components that are difficult to measure precisely. To bridge this gap, VeilGen incorporates a Latent Optical Transmission and Glare Map Predictor (LOTGMP) that estimates the two key latent components, the transmission and glare maps, within the diffusion process. The latent maps are injected via the Veiling Glare Imposition Module (VGIM) to modulate features, enabling VeilGen to synthesize realistic paired data (see §3.3).

Building upon our generative foundation, we propose DeVeiler, a restoration network trained with a reversibility constraint (Fig. 3(b)). Unlike pure blind-learning approaches, which produce an averaged, suboptimal removal for spatial-varying veiling glare, DeVeiler is structurally guided to learn an inverse mapping of the degradation. However, directly using VeilGen during restoration training is impractical because its multi-step diffusion sampling is computationally expensive. We therefore distill the generative process into a lightweight forward model that preserves its behavior and latent-map conditioning, and use it to supervise the training process of DeVeiler (see §3.4). Its core component, the Veiling Glare Compensation Module (VGCM), uses internally estimated latent maps to modulate image features, thereby precisely reversing the forward glare imposition process (see §3.5). This allows DeVeiler to remove glare based on its underlying physical causes instead of spurious statistical correlations, achieving superior restoration results, as shown in Fig. 1(b). We verify our framework on two challenging optical systems: a large-aperture single-lens and a meta-refractive hybrid lens, both of which exhibit significant compound degradations. Extensive experimental results show that our approach achieves state-of-the-art results, outperforming existing methods. This validates our framework, where a learned forward scattering model provides powerful supervision to ensure its inverse restoration network can effectively utilize the shared latent maps to remove veiling glare. In a nutshell, our contributions are summarized as follows:

- We propose VeilGen, a novel physics-informed generative model. It features a Latent Optical Transmission and Glare Map Predictor (LOTGMP) to estimate physical maps and a Veiling Glare Imposition Module (VGIM) that uses these maps to guide a diffusion process for synthesizing realistic compound optical degradations.
- We introduce DeVeiler, a restoration network trained with a reversibility constraint to approximate the inverse of the learned degradation. Supervision from a lightweight distilled forward model enforces consistency between the forward VGIM and the inverse feature modulation in the Veiling Glare Compensation Module (VGCM).
- Extensive experiments demonstrate that our method achieves state-of-the-art performance in the challenging task of joint aberration and veiling glare removal.

2. Related Work

Computational Aberration Correction. Computational Aberration Correction (CAC) employs algorithms to compensate residual optical aberrations. Recent learning-based approaches [4, 5, 20, 33, 34, 37, 62] have surpassed traditional optimization-based methods [13, 41, 54, 61] in performance. However, their dependence on simulated training data leads to a persistent sim-to-real gap and limited generalization across lenses. Efforts toward universal CAC frameworks [14, 21, 26] improve cross-lens adaptability by leveraging diverse lens libraries, yet they primarily target aberration-induced blur. The concurrent veiling glare from diffuse scattering remains underexplored, constraining per-

Figure 2. **Overall architecture of the proposed VeilGen.** In Stage I, VeilGen is trained to synthesize compound degradations by using a Latent Optical Transmission and Glare Map Predictor (LOTGMP) to estimate latent maps. These maps then guide the diffusion process via the Veiling Glare Imposition Module (VGIM). $Z_{de}^{\mathcal{T}}$ denotes the target degraded latent representation. Z_t denotes the noisy latent representation at timestep t of the forward diffusion process. Z_{null} represents an all-zero latent representation. $txt^{\mathcal{S}/\mathcal{T}}$ denotes the text prompts for the source and target domains, respectively.

formance under compound degradations.

Veiling Glare Removal. Research on glare-related artifacts generally falls into two categories. Lens flare removal [8–10, 31, 45, 47, 57, 66], targets structured, localized artifacts such as bright spots and streaks. In contrast, veiling glare manifests as a diffuse degradation that reduces image contrast, making flare-oriented approaches ineffective. Prior work relied on deconvolution using the Glare Spread Function (GSF) [43] or on image decomposition [64], yet these methods impose restrictive assumptions, often require multiple captures, or fail in low texture scenes. More recent learning-based techniques are limited by the lack of realistic paired data, as existing synthesis strategies [42] remain overly simplified. Moreover, previous studies typically handle veiling glare and aberration as independent degradations, whereas our framework jointly addresses both within a unified computational model.

Domain Adaptation in Low-level Vision. Domain Adaptation (DA) is key to mitigating domain shifts in low-level vision tasks [7, 25, 50, 53], yet generic approaches struggle when the shift arises from compound degradations that couple optical aberrations and veiling glare. This has motivated specialized DA methods for CAC that adapt pretrained models to new lenses [21, 22]. A promising direction is target-domain data synthesis via generative models [36, 49, 58], with SD-based frameworks such as HazeGen [49] showing strong potential. However, for optical degradations with well-defined physical properties, a black-box approach may be suboptimal. To this end, our work introduces a physics-informed generative model that explicitly estimates latent maps for veiling glare. These maps serve dual roles: conditioning an SD-based generator to produce realistic degradations and guiding the restoration network to better remove veiling glare.

3. Methodology

3.1. Problem Formulation

This work aims to restore a latent clean image I_c from a single degraded observation I_{de} captured by a compact optical system. The degradations central to this work are twofold: (i) spatial-varying aberration blur and (ii) veiling glare. We model the forward degradation process in two sequential steps. Following the widely adopted patch-wise convolution model [26, 33, 59, 60], the aberration image patch I_{ab}^p is obtained by convolving each clean patch I_c^p with its corresponding Point Spread Function (PSF) K^p, representing local aberration characteristics. This aberration patch I_{ab}^p is then sequentially degraded by veiling glare according to scattering model, formulated as a combination of attenuation and additive light [42, 65]: each patch I_{ab}^p is modulated by a transmission map T^p and combined with a glare map I_g^p, where T^p denotes local contrast attenuation and I_g^p represents the veiling glare. For each color channel (omitted

for brevity), the overall degradation model is expressed as:

$$I_{de}^p = \underbrace{(I_c^p \otimes K^p)}_{I_{ab}^p} \cdot T^p + I_g^p. \quad (1)$$

The restoration of I_c^p from I_{de}^p thus constitutes a highly ill-posed blind inverse problem. The numerous unknowns in the degradation model make optimization-based methods with generic priors ineffective for this problem. This motivates a data-driven approach capable of learning the inverse mapping directly from data. However, its success fundamentally depends on the availability of high-quality paired training data (I_{de}, I_c), which is rarely accessible in practice. Accordingly, our work focuses on two key objectives: (1) generating paired data that accurately captures the target degradation characteristics and (2) designing a restoration network that effectively leverages it. Formally, the overall objective is to learn optimal network parameters θ via

$$\theta^* = \arg\min_{\theta} \mathcal{L}(f_\theta(I_{de}), I_c), \quad (2)$$

where $\mathcal{L}$ denotes an appropriate loss function.

3.2. Overall Framework

Our framework addresses compound degradations through a three-stage, physics-informed pipeline that unifies data synthesis and restoration via shared latent priors. In Stage I, VeilGen, an SD-based generator, synthesizes realistic compound degradations. Its LOTGMP predicts latent transmission and glare maps, which guide the Veiling Glare Imposition Module (VGIM) to modulate image features and produce realistic degradations (Fig. 2). In Stage II, VeilGen is distilled into a lightweight Distilled Degradation Net (DDN) that preserves its degradation behavior while remaining efficient enough to supervise restoration training (Fig. 3(a)). In Stage III, the restoration network, DeVeiler, is trained to invert the forward degradation. It achieves this by leveraging the predicted latent maps via its Veiling Glare Compensation Module (VGCM) to invert the scattering process of the VGIM (Fig. 3(b)).

3.3. Stage I: Realistic Degradation Generation

To overcome data scarcity, we introduce VeilGen, an SD-based generative model that synthesizes realistic paired data (Fig. 2(a)). Built upon DiffBIR [30] with a fixed Stable Diffusion backbone and IRControlNet, VeilGen adapts the standard denoising paradigm [17] and incorporates a novel mechanism for modeling veiling glare.

Physics-informed Hybrid Training. To bridge a source domain $\mathcal{S}$ with paired, glare-free data and a target domain $\mathcal{T}$ with unpaired compound degradations, we propose a physics-informed hybrid training paradigm inspired by recent work in domain adaptation [49]. Since the

Figure 3. **The distillation and restoration networks.** (a) The Distilled Degradation Net (DDN), trained in Stage II, models the forward degradation, using VGIM to apply the veiling glare prior. (b) The DeVeiler, trained in Stage III, is designed to reverse this process: it first removes the veiling glare using the VGCM, and then feeds the intermediate result into its main bottleneck to correct the aberrations.

spatial-varying transmission and glare maps governing veiling glare are difficult to obtain directly, the diffusion model alone cannot capture the underlying physics. To address this, we introduce the Latent Optical Transmission and Glare Map Predictor (LOTGMP), which infers these latent maps during denoising, providing physically meaningful guidance to the generative process. Unlike conventional end-to-end degradation learning, VeilGen steers generation using unified latent maps $c_{vg} = (z_{trans}, z_{glare})$:

$$c_{vg} = \begin{cases} (\mathbf{1}, \mathbf{0}) & \text{if in source domain,} \\ \text{LOTGMP}(z_t, z_{de}^{\mathcal{T}}, t) & \text{otherwise.} \end{cases} \quad (3)$$

For the source domain, we use fixed maps $(\mathbf{1}, \mathbf{0})$ corresponding to an identity transformation (no veiling glare). For the target domain, we introduce LOTGMP, a lightweight, time-aware network that predicts the latent maps from a noisy latent z_t, a target degraded latent $z_{de}^{\mathcal{T}}$, and the timestep t. The predicted maps are injected via VGIM (Fig. 2(b)), mirroring the forward degradation model (Eq. 1). The overall training objective is a weighted combination of losses from both domains, each conditioned on a domain-specific text prompt:

$$\mathcal{L}_{gen} = p\, \mathcal{L}_{\mathcal{S}} + (1-p)\, \mathcal{L}_{\mathcal{T}}. \quad (4)$$

The source loss, $\mathcal{L}_{\mathcal{S}} = \mathbb{E}[\|\epsilon - \epsilon_\theta(z_t^{\mathcal{S}}, c^{\mathcal{S}}, c_{txt}^{\mathcal{S}}, c_{vg}, t)\|_2^2]$, is conditioned on both the clean image c^s and a text prompt

Algorithm 1 Degradation-Clean Paired Data Generation

Require: Denoising model ϵ_θ, VAE encoder $\mathcal{E}$ and decoder $\mathcal{D}$, LOTGMP $\mathcal{P}$, I_c: a clean image, $I_{de}^{\mathcal{T}}$: a target-domain degraded image, w: mixture coefficient

$\mathbf{z}_{de}^{\mathcal{T}} \leftarrow \mathcal{E}(I_{de}^{\mathcal{T}})$

▷ Set a timestep t^* in $\mathcal{P}$

$\boldsymbol{\epsilon}_{t^*} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

▷ Add noise to $\mathbf{z}_{de}^{\mathcal{T}}$

$\mathbf{z}_{t^*} \leftarrow \sqrt{\bar{\alpha}_{t^*}}\mathbf{z}_{de}^{\mathcal{T}} + \sqrt{1-\bar{\alpha}_{t^*}}\boldsymbol{\epsilon}_{t^*}$

$c_{vg} \leftarrow \mathcal{P}(\mathbf{z}_{t^*}, \mathbf{z}_{de}^{\mathcal{T}}, t^*)$

$\mathbf{z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

for $t = T, \ldots, 1$ **do**

▷ Blended noise prediction

$\hat{\epsilon} = w\ \epsilon_\theta(\mathbf{z}_t, \mathcal{E}(I_c), \varnothing, t) + (1-w)\ \epsilon_\theta(\mathbf{z}_t, \varnothing, c_{vg}, t)$

▷ Sampling step

$\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

$\mathbf{z}_{t-1} \leftarrow \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{z}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\hat{\epsilon}\right) + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}(1-\alpha_t)}\epsilon$

end for

return $\mathcal{D}(\mathbf{z}_0)$

$c_{txt}^{\mathcal{S}}$ describing only aberrations. The target loss, $\mathcal{L}_{\mathcal{T}} = \mathbb{E}[\|\epsilon - \epsilon_\theta(z_t^{\mathcal{T}}, \varnothing, c_{txt}^{\mathcal{T}}, c_{vg}, t)\|_2^2]$, is guided by both a text prompt $c_{txt}^{\mathcal{T}}$ describing the full compound degradations and the latent maps c_{vg}. The expectation $\mathbb{E}$ is taken over all relevant variables for each domain, and $p \in [0, 1]$ is a balancing hyperparameter.

Degradation-Clean Paired Data Generation. Once trained, VeilGen is employed to generate the paired dataset for training our restoration network in an offline process (Algorithm 1). We use the standard notation from DDPM [17], where $\beta_t \in (0, 1)$ is the variance schedule, $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$ is the cumulative product. To generate each degraded image I_{de}, we first extract latent maps c_{vg} from an unpaired target-domain image $I_{de}^{\mathcal{T}}$ via the frozen LOTGMP. These maps then guide a full diffusion sampling process that applies the degradation to a clean image I_c. This sampling process incorporates the blended strategy from [49], mixing conditional and unconditional predictions to enhance realism and diversity. This process produces a high-fidelity paired dataset in which each degraded image contains realistic compound degradations, including aberration and veiling glare.

3.4. Stage II: Forward Model Distillation

Our objective is to use the forward degradation from VeilGen as supervision for the restoration network in Stage III. However, invoking VeilGen during training is computationally prohibitive due to its multi-step diffusion sampling. We therefore distill its behavior into a lightweight Distilled Degradation Network (DDN), which serves as an efficient forward model (Fig. 3(a)). Given a clean image I_c and latent maps c_{vg}, the DDN is trained to regress the degraded output I_{de} synthesized by VeilGen:

$$\mathcal{L}_{distill} = \|\text{DDN}(I_c, c_{vg}) - \text{VeilGen}(I_c, c_{vg})\|_1 . \quad (5)$$

After training, this frozen network provides an efficient forward degradation model for imposing the consistency constraint in Stage III.

3.5. Stage III: Reversible Restoration

To tackle the spatial-varying veiling glare that hinders blind restoration, we develop DeVeiler, a reversible restoration network that predicts latent maps to invert the underlying degradation. The central component is the Veiling Glare Encoder (VG-Enc), which predicts latent maps $\hat{c}_{vg} = \text{VG-Enc}(I_{de})$ representing the spatial distribution of glare-related parameters. Simply using the predicted maps $\hat{c}_{vg}$ in a one direction manner is ineffective due to a fundamental domain mismatch. Latent maps optimized for conditioning iterative diffusion sampling do not transfer well to guiding the restoration process (see in Tab. 5). To bridge this gap, we introduce a reversibility-guided training paradigm that provides indirect but powerful supervision by leveraging the frozen DDN as a forward degradation model to enforce cycle consistency (Fig. 3(b)). This constraint drives $\hat{c}_{vg}$ to acquire meaningful physical interpretability, as it must parameterize two mutually inverse operations. Within the frozen DDN, the VGIM module uses $\hat{c}_{vg}$ to impose forward modulation, applying the glare degradation to clean features. Symmetrically, in DeVeiler, the VGCM applies an inverse modulation using the same latent maps $\hat{c}_{vg}$ to remove the glare components from degraded features. This structural coupling of VGIM and VGCM encourages the restoration to be a consistent inverse of the learned forward degradation, rather than a loosely correlated statistical mapping.

Optimization Objective. DeVeiler is trained end-to-end with a composite objective that balances image reconstruction fidelity and reversibility consistency:

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \lambda_{\text{rev}}\mathcal{L}_{rev}. \quad (6)$$

Here, $\mathcal{L}_{rec}$ combines pixel-wise L1 and perceptual losses to ensure high visual fidelity between the restored and ground-truth (GT) images. The reversibility loss $\mathcal{L}_{\text{rev}}$ enforces reversibility by constraining the estimated latent maps $\hat{c}_{vg}$ to reproduce the observed degradation through the frozen DDN:

$$\mathcal{L}_{rev} = \|\text{DDN}(I_c, \hat{c}_{vg}) - I_{de}\|_1. \quad (7)$$

Two-Phase Training. To effectively address the domain gap, we propose a two-phase training strategy. In Phase I (pre-training), DeVeiler is first trained on the source domain to establish a solid baseline for aberration correction. In Phase II (fine-tuning), it is further adapted using a hybrid dataset that mixes synthetic paired data generated by

Methods	Screen-SL			Screen-MRL			Overhead
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	Latency(s)↓
Single Degradation Pipeline							
Fast two-step[ECCV22] [11]	16.62 (73.5%)	0.539 (35.3%)	0.674 (61.3%)	14.94 (78.3%)	0.610 (22.3%)	0.504 (40.3%)	0.390
SwinIR[ICCVW21] [29]	18.18 (62.0%)	0.686 (6.27%)	0.298 (12.4%)	19.34 (40.2%)	0.722 (3.32%)	0.354 (15.0%)	0.782
NAFNET[ECCV22] [1]	18.75 (56.7%)	0.684 (6.58%)	0.363 (28.1%)	18.91 (45.8%)	0.723 (3.18%)	0.377 (20.2%)	0.353
DiffBIR[ECCV24] [30]	17.95 (63.9%)	0.621 (17.4%)	0.398 (34.4%)	18.70 (48.4%)	0.625 (19.4%)	0.412 (26.9%)	66.13
Cascaded Pipeline							
SwinIR[ICCVW21] [29]+RIDCP*[CVPR23] [56]	20.06 (41.4%)	0.695 (4.89%)	0.385 (32.2%)	18.66 (48.8%)	0.692 (7.80%)	0.456 (34.0%)	1.240
SwinIR[ICCVW21] [29]+DiffDehaze*[CVPR25] [49]	19.31 (50.7%)	0.642 (13.6%)	0.347 (24.8%)	18.28 (53.1%)	0.603 (23.7%)	0.392 (23.2%)	92.54
SwinIR[ICCVW21] [29]+Flare7K++*[TPAMI2024] [10]	21.67 (15.1%)	0.723 (0.86%)	0.297 (12.1%)	20.74 (17.4%)	0.745 (0.13%)	0.336 (10.4%)	0.990
Domain Adaptation Pipeline							
CycleGAN[ICCV17] [67]	18.20 (61.8%)	0.558 (30.6%)	0.649 (59.8%)	17.80 (58.0%)	0.623 (19.7%)	0.485 (37.9%)	0.208
UCL-Dehaze[TIP24] [51]	18.55 (58.6%)	0.581 (25.5%)	0.554 (52.9%)	18.69 (48.5%)	0.688 (8.43%)	0.451 (33.3%)	0.208
DiffDehaze[CVPR25] [49]	18.87 (55.4%)	0.657 (11.0%)	0.435 (40.0%)	18.55 (50.1%)	0.645 (15.7%)	0.426 (29.3%)	91.76
QDMR[OLT2025] [22]	18.45 (59.5%)	0.681 (7.05%)	0.291 (10.3%)	20.67 (18.7%)	0.725 (2.90%)	0.315 (4.44%)	0.428
DeVeiler (Ours)	**22.38** (0.00%)	**0.729** (0.00%)	**0.261** (0.00%)	**21.57** (0.00%)	**0.746** (0.00%)	**0.301** (0.00%)	0.387

Table 1. **Quantitative comparison on the `Screen-Compound` domain (SL and MRL systems).** * denotes using the pretrained model. Latency(s) is computed on images with the size of 1280×1920 with an NVIDIA A100 GPU. **Red** and blue indicate the best and the second best performance, respectively. The numbers in parentheses denote the improvement rate of our method compared to others.

VeilGen with the original source-domain pairs. This hybrid approach is critical, acting as a powerful regularizer: it allows the network to learn the new compound degradation from the synthetic pairs, while the source domain data preserves the foundational aberration correction and prevents overfitting to the synthetic-only distribution.

4. Experiments and Results

4.1. Experiment Settings

Datasets. We verify our method on two distinct, independently evaluated optical systems: a large-aperture single lens (SL) and a metasurface-refractive hybrid lens (MRL). We first constructed a source domain of paired, aberration-only data. Lacking the precise optical parameters required for simulation, we follow prior work [3, 62] and adopt a screen-capture setup to re-capture clean images from the DIV2K dataset [44], yielding 170 (SL) and 125 (MRL) pairs. We then collected two distinct Target Domains for adaptation, adhering to a strict "Source + one Target" setup. The `Screen-Compound` domain used the same screen-capture method but with artificial lighting to induce compound degradation, providing GT pairs for 50 training and 42/25 (SL/MRL) test images. The `Realworld-Compound` domain consists of real-world captures exhibiting natural compound degradation with no GT, providing 50 training and 51/11 (SL/MRL) test images. Finally, we use 500 clean images from the Flickr2K dataset [44] to generate 500 corresponding target-domain degraded pairs. All data across both systems are captured at a resolution of 1280×1920.

Implementation Details. VeilGen builds upon pretrained Stable Diffusion v2-1 [40] and is optimized with AdamW [32] at a learning rate of 1×10^{-5} for 9k iterations with a batch size of 16. The probability parameter p is set to 0.3 and the mixture coefficient w to 0.85, consistent with common practices [49]. t^* is set to 0. DeVeiler is trained in two phases using Adam [23]: Phase I pre-trains on source-domain pairs, and Phase II fine-tunes on a hybrid dataset of 500 synthetic pairs and source pairs. Each encoder and decoder consists of three groups of ResBlocks, while the bottleneck is constructed with RSTB layers from SwinIR [29]. The balance weight λ_{rev} is set to 1.0. All experiments are conducted on an NVIDIA A100 GPU. More details are delivered in the supplementary material.

Methods	Realworld-SL	Realworld-MRL
Single Degradation Pipeline		
Fast two-step [11]	0.356/2.942/5.821	0.388/2.687/5.352
SwinIR [29]	0.424/3.518/5.710	0.374/3.191/6.696
NAFNET [1]	0.374/3.598/5.535	0.398/3.238/5.512
DiffBIR [30]	0.404/3.476/5.770	0.367/2.761/6.251
Cascaded Pipeline		
SwinIR [29]+RIDCP* [56]	0.386/3.755/5.810	0.399/3.290/6.502
SwinIR [29]+DiffDehaze* [49]	0.573/3.679/6.141	0.437/3.542/5.230
SwinIR [29]+Flare7K++* [10]	0.459/3.528/5.622	0.387/3.319/6.336
Domain Adaptation Pipeline		
CycleGAN [67]	0.451/2.961/5.300	0.418/3.319/**3.836**
UCL-Dehaze [51]	0.297/2.763/8.322	0.343/2.963/5.644
DiffDehaze [49]	0.406/3.982/6.476	0.428/3.476/6.802
QDMR [22]	0.405/3.864/4.773	0.376/3.337/5.509
DeVeiler (Ours)	**0.607/3.987/4.448**	**0.440/3.586**/5.296

Table 2. **Quantitative comparison (CLIPIQA↑/Q-Align↑/NIQE↓) on the `Realworld-Compound` domain (SL and MRL systems).** * denotes using the pretrained model.

4.2. Comparison with State-of-the-Art Methods

We compare against three pipelines in Tab. 1 and Tab. 2. First, we evaluate single-degradation baselines, which include the blind CAC method Fast two-step [11] as well as well-known image restoration models (SwinIR [29], NAFNet [1], DiffBIR [30]) re-trained on source domain pairs. Second, cascaded pipelines combine the re-trained

Figure 4. **Visual results on `Screen-Compound` domain.** Zoom in for the best view.

Figure 5. **Visual results on `Realworld-Compound` domain.** Zoom in for the best view.

SwinIR with pre-trained dehazing (RIDCP [56], DiffDehaze [49]) or flare removal (Flare7K++ [10]) models to validate the distinct nature of veiling glare. Finally, the Domain Adaptation (DA) methods are all retrained on our source and target domains for a fair comparison. This includes CycleGAN [67], UCL-Dehaze [51], the aberration-specific method QDMR [22], and a re-trained DiffDehaze [49].

Results on `Screen-Compound` Domain. We first conduct a quantitative evaluation on the `Screen-Compound` domain, which provides paired GT. We employ full-reference metrics: PSNR and SSIM [52] to measure fidelity, and LPIPS [63] to assess perceptual quality. As shown in Tab. 1, our DeVeiler significantly outperforms all baselines across all metrics. Single degradation pipeline, trained only on the source domain, fails to generalize to the unseen veiling glare. Cascaded pipelines also prove insufficient. We use the strongest single-degradation model, SwinIR, as the CAC base, followed by pre-trained dehazing or flare removal models. These naive cascades yield unsatisfactory results: dehazing models severely harm color fidelity, while the flare removal model shows negligible perceptual gains. This confirms the distinct nature of veiling glare compared to these other degradations. Furthermore, the DA pipeline produces sub-optimal results due to its physically-agnostic designs. In contrast, DeVeiler leverages the latent veiling glare maps (c_{vg}) via the mutually inverse VGIM/VGCM structure, enabling it to precisely reverse the degradation.

Visual comparisons in Fig. 4 on both SL and MRL systems confirm these quantitative findings. For example, in the first row, the results of DeVeiler consistently restore high color fidelity, and in the second row, they successfully remove severe veiling glare while preserving fine details. In contrast, competing methods exhibit clear color shifts, residual glare, or loss of detail.

Results on `Realworld-Compound` Domain. We further evaluate the performance of DeVeiler on the challenging `Realworld-Compound` domain using both SL and MRL systems. As no GT is available, we employ three widely-used no-reference IQA metrics: CLIPIQA [48], Q-Align [55], and NIQE [35]. Tab. 2 shows that DeVeiler achieves excellent performance across the majority of metrics, verifying its generalizability in handling complex, real-world degradations. In addition, we provide visual comparisons in Fig. 5. DeVeiler consistently restores fine details and color fidelity while introducing minimal artifacts, whereas other methods struggle with residual glare and artifacts. Notably, the first row example highlights that aberration and veiling glare are a more pervasive challenge in compact systems than severe flare or ghosting, which are situational. These experimental results demonstrate the ro-

LOTGMP	SD-based prior	PSNR↑	SSIM↑	LPIPS↓
✗	✗	20.82	0.708	0.273
✓	✗	21.39	0.708	0.268
✓	✓	**21.56**	**0.712**	**0.264**

Table 3. Ablation study on the proposed VeilGen.

Methods	PSNR↑	SSIM↑	LPIPS↓
CycleGAN	18.85	0.706	0.326
Degradation Transfer	17.64	0.686	0.353
HazeGen	20.82	0.709	0.273
VeilGen (Ours)	**21.56**	**0.712**	**0.264**

Table 4. Ablation study on data generation.

Figure 6. Visual comparison of data generation methods.

bustness of DeVeiler in realistic scenarios. We refer the reader to the supplementary material for more visual results.

4.3. Ablation Studies

To analyze the effectiveness of each component, we conduct ablation studies on the `Screen-Compound` SL dataset, maintaining consistent settings for fair comparison.

Ablation Study for VeilGen. We evaluate the impact of the LOTGMP module by training DeVeiler (*w/o* VGCM) on data generated from different VeilGen variants. As shown in Tab. 3, removing LOTGMP or discarding its SD-based prior input (z_t) consistently degrades performance, indicating that the SD-guided latent input is essential for adaptive prior learning (c_{vg}) and realistic data synthesis.

Effectiveness of VeilGen. We compare VeilGen with competitive data generation methods (CycleGAN [67], DegradationTransfer [4], HazeGen [49]). For a fair comparison, an identical restoration network (DeVeiler (*w/o* VGCM)) is trained on the data generated by each method. As illustrated in Tab. 4, training the restoration network on data from VeilGen yields superior performance, indicating that its generated pairs more accurately capture the target degradation characteristics. The visual comparison in Fig. 6 further shows that VeilGen synthesizes realistic compound degradations while other methods struggle to reproduce the complex degradations.

Ablation Study for DeVeiler. We next ablate the method for using the latent maps (c_{vg}) in DeVeiler (Tab. 5). The results show that naive unidirectional injection (via Concat or VGCM) provides no benefit. We attribute this failure to a fundamental domain mismatch: diffusion-optimized latent maps are ill-suited for conventional restoration networks. This mismatch is resolved by our reversible paradigm, where the bidirectional VGIM/VGCM structure enforces pathway consistency and unlocks performance gains.

Methods	PSNR↑	SSIM↑	LPIPS↓
Baseline (*w/o* Latent maps)	21.56	0.712	0.264
Concat (Unidirectional)	21.12	0.671	0.317
VGCM (Unidirectional)	20.83	0.712	0.264
Concat (Bidirectional)	20.65	0.662	0.329
VGIM/VGCM (Bidirectional)	**22.38**	**0.729**	**0.261**

Table 5. Ablation study on the LOTGMP usage.

Figure 7. Visual ablation of core components.

Figure 8. Interpretability of the reversible mechanism.

Visual Ablation of Core Components. Fig. 7 shows the cumulative effect of our core components. The source-only baseline fails to remove veiling glare. Adding LOTGMP improves data generation (consistent with Tab. 3), and activating the full bidirectional DeVeiler further eliminates residual artifacts, producing results closest to the GT.

Analysis of the VGIM/VGCM Mechanism. Fig. 8(b)-(f) visualizes internal feature maps of DeVeiler. The network accurately estimates latent components, identifying low-transmittance and high-glare regions in the input. After modulation by the VGCM, activations in the glare-affected regions are effectively suppressed. This verifies that DeVeiler learns an interpretable restoration process, where the VGCM module specifically targets and removes the glare component, rather than performing a black-box mapping.

5. Conclusion

In this paper, we address the problem of compound degradation involving co-existing aberration and veiling glare in

compact optical systems. Our physics-informed framework unifies data generation (VeilGen) and restoration (DeVeiler) through a shared latent map realized via bidirectional modulation. Experimental results show that our method performs favorably compared to other methods. We hope our work can inspire the development of high-quality imaging in compact optics.

References

[1] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, 2022. 6

[2] Liqun Chen, Yuxuan Li, Jun Dai, Jinwei Gu, and Tianfan Xue. A physics-informed blur learning framework for imaging systems. In *CVPR*, 2025. 2

[3] Qikai Chen, Jiacheng Zhou, Sijie Pian, Jingang Xu, Xingyi Li, Bihua Li, Chentao Lu, Zhuning Wang, Qi Jiang, Shanhe Qin, Hantao Zhan, Benhao Zhang, Xu Liu, Kaiwei Wang, and Yaoguang Ma. Hybrid meta-optics enabled compact augmented reality display with computational image reinforcement. *ACS Photonics*, 2024. 1, 6, 11

[4] Shiqi Chen, Huajun Feng, Keming Gao, Zhihai Xu, and Yueting Chen. Extreme-quality computational imaging via degradation framework. In *ICCV*, 2021. 1, 2, 8

[5] Shiqi Chen, Huajun Feng, Dexin Pan, Zhihai Xu, Qi Li, and Yueting Chen. Optical aberrations correction in postprocessing using imaging simulation. *ACM Transactions on Graphics*, 2021. 2, 13

[6] Shiqi Chen, Jingwen Zhou, Menghao Li, Yueting Chen, and Tingting Jiang. Mobile image restoration via prior quantization. *Pattern Recognition Letters*, 2023. 1

[7] Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. PSD: Principled synthetic-to-real dehazing guided by physical priors. In *CVPR*, 2021. 3

[8] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7K: A phenomenological nighttime flare removal dataset. In *NeurIPS*, 2022. 2, 3

[9] Yuekun Dai, Yihang Luo, Shangchen Zhou, Chongyi Li, and Chen Change Loy. Nighttime smartphone reflective flare removal using optical center symmetry prior. In *CVPR*, 2023.

[10] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, Yihang Luo, and Chen Change Loy. Flare7K++: Mixing synthetic and real datasets for nighttime flare removal and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2, 3, 6, 7, 15

[11] Thomas Eboli, Jean-Michel Morel, and Gabriele Facciolo. Fast two-step blind optical aberration correction. In *ECCV*, 2022. 6

[12] K. Faulkner, C. J. Kotre, and M. Louka. Veiling glare deconvolution of images produced by X-ray image intensifiers. In *ICIPA*, 1989. 2

[13] D. A. Fish, A. M. Brinicombe, E. R. Pike, and J. G. Walker. Blind deconvolution by means of the Richardson–Lucy algorithm. *Journal of the Optical Society of America A*, 1995. 2

[14] Jin Gong, Runzhao Yang, Weihang Zhang, Jinli Suo, and Qionghai Dai. A physics-informed low-rank deep neural network for blind and universal lens aberration correction. In *CVPR*, 2024. 2

[15] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011. 2

[16] Felix Heide, Mushfiqur Rouf, Matthias B. Hullin, Björn Labitzke, Wolfgang Heidrich, and Andreas Kolb. High-quality computational imaging through simple lenses. *ACM Transactions on Graphics*, 2013. 2

[17] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020. 4, 5

[18] Matthias Hullin, Elmar Eisemann, Hans-Peter Seidel, and Sungkil Lee. Physically-based real-time lens flare rendering. In *ACM SIGGRAPH*, 2011. 2

[19] The MathWorks Inc. MATLAB version: 9.13.0 (R2022b), 2022. 11

[20] Qi Jiang, Shaohua Gao, Yao Gao, Kailun Yang, Zhonghua Yi, Hao Shi, Lei Sun, and Kaiwei Wang. Minimalist and high-quality panoramic imaging with PSF-aware transformers. *IEEE Transactions on Image Processing*, 2024. 2

[21] Qi Jiang, Yao Gao, Shaohua Gao, Zhonghua Yi, Xiaolong Qian, Hao Shi, Kailun Yang, Lei Sun, and Kaiwei Wang. OmniLens: Towards universal lens aberration correction via LensLib-to-specific domain adaptation. *arXiv preprint arXiv:2409.05809*, 2024. 2, 3

[22] Qi Jiang, Zhonghua Yi, Shaohua Gao, Yao Gao, Xiaolong Qian, Hao Shi, Lei Sun, JinXing Niu, Kaiwei Wang, Kailun Yang, and Jian Bai. Representing domain-mixing optical degradation for real-world computational aberration correction via vector quantization. *Optics & Laser Technology*, 2025. 3, 6, 7, 15

[23] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 6, 14

[24] Yousef Kotp and Marwan Torki. Toward flare-free images: A survey. *arXiv preprint arXiv:2310.14354*, 2023. 1

[25] Lerenhan Li, Yunlong Dong, Wenqi Ren, Jinshan Pan, Changxin Gao, Nong Sang, and Ming-Hsuan Yang. Semi-supervised image dehazing. *IEEE Transactions on Image Processing*, 2020. 3

[26] Xiu Li, Jinli Suo, Weihang Zhang, Xin Yuan, and Qionghai Dai. Universal and flexible optical aberration correction using deep-prior based deconvolution. In *ICCV*, 2021. 2, 3

[27] Zhaoyi Li, Peng Lin, Yao-Wei Huang, Joon-Suh Park, Wei Ting Chen, Zhujun Shi, Cheng-Wei Qiu, Ji-Xin Cheng, and Federico Capasso. Meta-optics achieves RGB-achromatic focusing for virtual reality. *Science Advances*, 2021. 1

[28] Zongling Li, Fanjiao Tan, Rongshuai Zhang, and Qingyu Hou. Optical attention mechanism for high-resolution computational imaging. *Optica*, 2025. 2

[29] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *ICCVW*, 2021. 1, 6, 15

[30] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. DiffBIR: Toward blind image restoration with generative diffusion prior. In *ECCV*, 2024. 4, 6

[31] Qu Lishen, Liu Zhihao, Pan Jinshan, Zhou Shihao, Shi Jinglei, Chen Duosheng, and Yang Jufeng. FlareX: A physics-informed dataset for lens flare removal via 2D synthesis and 3D rendering. In *NeurIPS*, 2025. 2, 3

[32] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2019. 6, 14

[33] Jun Luo, Yunfeng Nie, Wenqi Ren, Xiaochun Cao, and Ming-Hsuan Yang. Correcting optical aberration via depth-aware point spread functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2, 3

[34] Debabrata Mandal, Zhihan Peng, Yujie Wang, and Praneeth Chakravarthula. Enabling high-quality in-the-wild imaging from severely aberrated metalens bursts. *arXiv preprint arXiv:2510.10083*, 2025. 2

[35] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind"' image quality analyzer. *IEEE Signal Processing Letters*, 2013. 7

[36] Long Peng, Wenbo Li, Renjing Pei, Jingjing Ren, Jiaqi Xu, Yang Wang, Yang Cao, and Zheng-Jun Zha. Towards realistic data generation for real-world super-resolution. In *ICLR*, 2025. 2, 3

[37] Yifan Peng, Qilin Sun, Xiong Dun, Gordon Wetzstein, Wolfgang Heidrich, and Felix Heide. Learned large field-of-view imaging with thin-plate optics. *ACM Transactions on Graphics*, 2019. 2

[38] Xiaolong Qian, Qi Jiang, Yao Gao, Shaohua Gao, Zhonghua Yi, Lei Sun, Kai Wei, Haifeng Li, Kailun Yang, Kaiwei Wang, and Jian Bai. Towards single-lens controllable depth-of-field imaging via depth-aware point spread functions. *IEEE Transactions on Computational Imaging*, 2025. 1

[39] Ramesh Raskar, Amit Agrawal, Cyrus A Wilson, and Ashok Veeraraghavan. Glare aware photography: 4d ray sampling for reducing glare effects of camera lenses. In *ACM SIGGRAPH*, 2008. 2

[40] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022. 6, 14

[41] Christian J. Schuler, Michael Hirsch, Stefan Harmeling, and Bernhard Schölkopf. Blind correction of optical aberrations. In *ECCV*, 2012. 2

[42] Alexey Valeryevich Shoshin and Evgeny Alexandrovich Shvets. Veiling glare removal: synthetic dataset generation, metrics and neural network architecture. *Computer Optics*, 2021. 2, 3

[43] Eino-Ville Talvala, Andrew Adams, Mark Horowitz, and Marc Levoy. Veiling glare in high dynamic range imaging. *ACM Transactions on Graphics*, 2007. 2, 3

[44] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, Lei Zhang, Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, Kyoung Mu Lee, et al. NTIRE 2017 challenge on single image super-resolution: Methods and results. In *CVPRW*, 2017. 6, 13, 14

[45] Shr-Ruei Tsai, Wei-Cheng Chang, Jie-Ying Lee, Chih-Hai Su, and Yu-Lun Liu. LightsOut: Diffusion-based outpainting for enhanced lens flare removal. In *ICCV*, 2025. 2, 3

[46] Ethan Tseng, Shane Colburn, James Whitehead, Luocheng Huang, Seung-Hwan Baek, Arka Majumdar, and Felix Heide. Neural nano-optics for high-quality thin lens imaging. *Nature Communications*, 2021. 1

[47] Florin Vasluianu, Zongwei Wu, and Radu Timofte. SFNet - A spatial-frequency domain neural network for image lens flare removal. In *ICIP*, 2024. 3

[48] Jianyi Wang, Kelvin C. K. Chan, and Chen Change Loy. Exploring CLIP for assessing the look and feel of images. In *AAAI*, 2023. 7

[49] Ruiyi Wang, Yushuo Zheng, Zicheng Zhang, Chunyi Li, Shuaicheng Liu, Guangtao Zhai, and Xiaohong Liu. Learning hazing to dehazing: Towards realistic haze generation for real-world image dehazing. In *CVPR*, 2025. 1, 2, 3, 4, 5, 6, 7, 8, 14, 15

[50] Wei Wang, Haochen Zhang, Zehuan Yuan, and Changhu Wang. Unsupervised real-world super-resolution: A domain adaptation perspective. In *ICCV*, 2021. 3

[51] Yongzhen Wang, Xuefeng Yan, Fu Lee Wang, Haoran Xie, Wenhan Yang, Xiao-Ping Zhang, Jing Qin, and Mingqiang Wei. UCL-Dehaze: Toward real-world image dehazing via unsupervised contrastive learning. *IEEE Transactions on Image Processing*, 2024. 2, 6, 7

[52] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004. 7

[53] Yunxuan Wei, Shuhang Gu, Yawei Li, Radu Timofte, Longcun Jin, and Hengjie Song. Unsupervised real-world image super resolution via domain-distance aware training. In *CVPR*, 2021. 3

[54] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. MIT press Cambridge, MA, 1949. 2

[55] Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, Qiong Yan, Xiongkuo Min, Guangtao Zhai, and Weisi Lin. Q-Align: Teaching LMMs for visual scoring via discrete text-defined levels. In *ICML*, 2024. 7

[56] Rui-Qi Wu, Zheng-Peng Duan, Chun-Le Guo, Zhi Chai, and Chongyi Li. RIDCP: Revitalizing real image dehazing via high-quality codebook priors. In *CVPR*, 2023. 2, 6, 7

[57] Yicheng Wu, Qiurui He, Tianfan Xue, Rahul Garg, Jiawen Chen, Ashok Veeraraghavan, and Jonathan T. Barron. How to train neural networks for flare removal. In *ICCV*, 2021. 3

[58] Tao Yang, Peiran Ren, Xuansong Xie, and Lei Zhang. Synthesizing realistic image restoration training pairs: A diffusion approach. *arXiv preprint arXiv:2303.06994*, 2023. 2, 3

[59] Xinge Yang, Qiang Fu, Mohamed Elhoseiny, and Wolfgang Heidrich. Aberration-aware depth-from-focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 3

[60] Xinge Yang, Chuong Nguyen, Wenbin Wang, Kaizhang Kang, Wolfgang Heidrich, and Xiaoxing Li. Efficient depth- and spatially-varying image simulation for defocus deblur. In *ICCVW*, 2025. 3

[61] Tao Yue, Jinli Suo, Jue Wang, Xun Cao, and Qionghai Dai. Blind optical aberration correction by exploring geometric and visual priors. In *CVPR*, 2015. 2

[62] Jianing Zhang, Jiayi Zhu, Feiyu Ji, Xiaokang Yang, and Xiaoyun Yuan. Degradation-modeled multipath diffusion for tunable metalens photography. *ICCV*, 2025. 1, 2, 6, 11, 13

[63] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 7

[64] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 3

[65] Zheng Zhang, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. Single image veiling glare removal. *Journal of Modern Optics*, 2018. 2, 3

[66] Jie Zhu and Sungkil Lee. PBFG: A new physically-based dataset and removal of lens flares and glares. In *ICCV*, 2025. 3

[67] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *ICCV*, 2017. 6, 7, 8

This supplementary material is organized as follows. First, §A details the procedures for data acquisition and usage across different domains. Next, §B presents a detailed motivation analysis regarding the learning of latent veiling glare maps. §C provides further implementation details, including training configurations and hyperparameters. In §D, we discuss the broader societal impacts of our work, along with its limitations and potential future directions. Finally, §E demonstrates the high data efficiency of our domain adaptation framework and presents additional visual comparisons.

A. Detailed Data Collection and Usage

This section details the three-domain dataset structure. We evaluate our method on two distinct optical systems: the large-aperture single lens (SL) and the metasurface-refractive hybrid lens (MRL), as visualized in Fig. 9(a)-(b). A summary of acquisition setups, illumination conditions, and dataset splits is provided in Tab. 6. To illustrate the degradation components, Fig. 10 shows a comparison among the clear image, the aberration-only capture (Source domain), and the compound case with additional veiling glare (`Sreen-Compound` domain). This reveals that veiling glare degrades the image through global contrast reduction and color shifts, compounding the intrinsic optical blur. While identical scenes are presented here for direct comparison, we ensure strictly disjoint scene contents across all dataset splits in our experiments to prevent data leakage. All procedures described below are conducted independently for each system, resulting in two parallel datasets.

A.1. Data Acquisition Setup

We employ two distinct data acquisition setups: a controlled monitor-based setup for the Source and `Screen-Compound` domains, and an in-the-wild capture setup for the `Realworld-Compound` domain.

Monitor-based Setup. We use a screen-capture setup [3, 62] for both the Source and `Screen-Compound` domains. This approach is chosen because precise lens parameters for accurate simulation are not available. In this setup, a high-resolution monitor displays the GT images, and the optical system is placed on a controlled mounting setup to capture the screen. The illumination conditions are carefully controlled to distinguish the two domains:

- **Source Domain:** The environment is kept completely dark to avoid any ambient light, ensuring that the captured images contain only intrinsic optical aberrations without veiling glare.
- **`Screen-Compound` Domain:** Under the same geometric alignment, we introduce an external light source (Fig. 9(c)) to generate compound degradation. This source is placed off the optical axis and remains outside the field of view. This configuration simulates the common real-world scenario where invisible external light induces diffuse veiling glare. We employ this standardized lighting setup to establish a controlled benchmark. Despite lacking the variability of natural environments, this setup provides the essential paired data required for full-reference validation. In contrast, the `Realworld-Compound` domain (detailed below) encompasses random, diverse lighting conditions to verify the method's generalization capability.

In-the-wild Capture Setup. We collect diverse real-world scenes in uncontrolled environments to form the `Realworld-Compound` domain. These images are captured under naturally varying illumination, where the intensity and distribution of light are not regulated. As a result, the collected data reflect the actual imaging behavior of the optical system in everyday scenarios in which glare arises naturally.

Geometric Alignment. To ensure precise pixel-level correspondence for paired data, a two-stage geometric correction process is applied independently for each system.

- **Lens Distortion Correction:** Intrinsic parameters are calibrated using the MATLAB Camera Calibrator [19]. Chessboard images captured from multiple viewpoints allow computation of a distortion map that removes nonlinear distortions such as barrel or pincushion effects.
- **Affine Alignment:** Residual planar misalignment between the sensor and the monitor is corrected using an affine transformation. We extract N corner correspondences $\{(x_i, y_i) \rightarrow (x'_i, y'_i)\}$ from a high-definition chessboard displayed on the monitor. The transformation is modeled directly in homogeneous coordinates as:

$$\begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} = \mathbf{M}_{\text{aff}} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & t_x \\ m_{21} & m_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

Domain	Acquisition Setup	Lighting Condition	GT Avail.	Data Quantity (SL / MRL) Training	Data Quantity (SL / MRL) Testing	Usage Setting
Source	Monitor-based	Dark	✓	170 / 125	—	**Train:** Supervised (Aberration Only)
`Screen-Compound`	Monitor-based	Artificial light	✓†	50 / 50	42 / 25	**Train:** Unpaired Adaptation (GT Discarded) **Test:** Full-Ref. Metrics & Visual Comp.
`Realworld-Compound`	In-the-wild	Uncontrolled	×	50 / 50	51 / 11	**Train:** Unpaired Adaptation **Test:** No-Ref. Metrics & Visual Comp.

† GTs are available but explicitly discarded during training to ensure strictly unpaired adaptation.

Table 6. **Summary of Datasets and Usage Setting.** We evaluate our method on two independent optical systems: Large-aperture Single Lens (SL) and Metasurface-Refractive Lens (MRL). The table details the acquisition setup, lighting conditions, data splits, and specific usage for each domain.

(a) SL System (b) MRL System (b) Screen-Compound Domain Capture Setup

Figure 9. (a) A large-aperture single-lens (SL) system. (b) A metasurface–refractive hybrid-lens (MRL) system. (c) The capture setup for the `Screen-Compound` domain.

Figure 10. Visual Decomposition of Compound Degradation.

The affine parameters are estimated by minimizing the sum of squared reprojection errors:

$$\min_{m_{ij},t} \sum_{i=1}^{N} \left\| \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} - \begin{bmatrix} x'_i \\ y'_i \end{bmatrix} \right\|^2$$

This correction is strictly applied to the Source and `Screen-Compound` domains to ensure high-fidelity alignment for quantitative metrics. Conversely, for the `Realworld-Compound` domain, geometric correction is intentionally omitted. By operating directly on uncorrected RGB images containing native distortion, we explicitly validate the robustness of the model in practical, calibration-free deployment scenarios.

A.2. Domain Definitions and Usage

We define three domains. The Source domain establishes baseline capabilities for aberration synthesis and correction. For unpaired adaptation, each target domain (`Screen-Compound` and `Realworld-Compound`) is treated separately, ensuring the model adapts to each domain independently.

Source Domain (Paired, Aberration-only).

- **Acquisition:** Using the aligned screen-capture setup, clean images from the DIV2K dataset [44] are re-captured in a dark environment. This ensures that only the intrinsic spatial-varying aberrations are recorded.
- **Data Split:** This domain yields 170 pairs for the SL system and 125 pairs for the MRL system.
- **Usage:** These pairs provide fully supervised supervision for modeling and correcting spatial-varying aberrations.

Target Domain: `Screen-Compound` (Unpaired Training, Paired Testing). This domain introduces controlled veiling glare and serves a dual purpose for training and evaluation.

- **Acquisition:** The process follows the Source domain setup, except that a single external light source is directed toward the lens (Fig. 9), inducing compound degradation from both aberrations and veiling glare.
- **Data Split:** We collect 50 training images and 42 test images for the SL system (50 training, 25 test for MRL). All images are captured with corresponding GT pairs.
- **Usage:** The dataset is strictly split:
  - **Training (Unpaired):** The 50 degraded images form the target domain, with GTs explicitly discarded to ensure unpaired adaptation.
  - **Testing (Paired):** Held-out pairs are reserved exclusively for full-reference quantitative evaluation.

Target Domain: `Realworld-Compound` (Unpaired). This domain serves to adapt the model to uncontrolled, realistic scenarios and evaluate its generalization capabilities.

- **Acquisition:** Diverse real-world scenes featuring natural, complex compound degradation are captured using independent optical systems.
- **Data Split:** No GT is available for these images. The dataset comprises 50 unpaired training images and 51 test images for the SL system (50 training, 11 test for MRL).
- **Usage:** Unpaired training images are used for adaptation; test images are used for qualitative comparison and quantitative evaluation via no-reference IQA metrics.

B. Motivation Analysis

Our framework addresses compound degradation arising from optical aberrations and veiling glare. Although we aim to mitigate both artifacts, our methodology prioritizes the modeling of the veiling glare component. This section outlines the rationale for this focus.

Baseline for Aberration Correction. Existing literature demonstrates that supervised networks trained on paired datasets effectively handle spatially varying aberrations [5, 62]. We follow this data-driven strategy by constructing a paired source domain to train the backbone network. This approach yields robust performance for aberration correction. Consequently, our work focuses on the removal of veiling glare, a more challenging and under-constrained problem, rather than the incremental improvements of aberration correction.

Data Scarcity in Veiling Glare. Unlike optical aberrations, acquiring paired data for real-world veiling glare is infeasible due to the sensitivity to dynamic lighting conditions. Moreover, high-fidelity simulation is restricted by the inaccessibility of proprietary opto-mechanical structures and the prohibitive computational cost of non-sequential ray tracing. This lack of GT results in an under-constrained degradation model (Eq. 1 in the main text). While the source domain constrains the PSF term K^p, the transmission map T^p and the glare map I_g^p remain unsupervised. Consequently, general domain adaptation methods yield suboptimal performance without explicit physical guidance.

Targeted Design for Glare Modeling and Removal. To bridge the supervision gap, we introduce VeilGen for data synthesis and DeVeiler for image restoration. Both components leverage the proposed LOTGMP prior to ensure physical consistency. In VeilGen, the prior guides the generation of plausible transmission and glare maps in latent space. This synthesis is validated in two ways: the resulting degradations are more realistic (Fig. 6 in the main text), and a network trained on this synthetic data achieves superior restoration performance (Tab. 3 in the main text). Similarly, DeVeiler incorporates the prior during restoration. The VG-Enc module estimates latent glare and transmission, which

the VGCM module then uses to selectively suppress activations in glare-affected regions, a process confirmed by our feature analysis (Fig. 8 in the main text). This dual application of the prior effectively addresses the challenges of data scarcity and model ambiguity. Future work could explore a unified architecture that jointly learns to correct both aberration and glare.

C. More Implementation Details

VeilGen Training. In stage I, the VeilGen is initialized from the pre-trained Stable Diffusion v2-1 [40]. We employ AdamW [32] with a constant learning rate of 1×10^{-5} for both the main diffusion backbone and the LOTGMP module. The model is trained for $9k$ iterations with a batch size of 16. Following standard practices in hybrid-domain diffusion learning [49], we set the probability parameter p to 0.3 during training and the mixture coefficient w to 0.85 during inference. We set the reference timestep to $t^* = 0$ to extract latent maps from the fully denoised state, ensuring estimation precision. To balance generation quality and efficiency, we employ a 10-step sampling strategy. Using the Flickr2K [44] dataset as the source, we synthesize 500 high-resolution images (1280×1920). This process requires approximately 3 hours on a single NVIDIA A100 GPU.

Distillation to DDN. In stage II, we distill the trained VeilGen for $25k$ iterations using the Adam optimizer [23] with a cosine annealing learning rate schedule, starting from 2×10^{-4} and decaying to 1×10^{-7}. The batch size is 8, and the input patch size is 256×256. This distillation preserves VeilGen's physically grounded degradation behavior while producing an efficient forward model suitable for supervision in Stage III.

DeVeiler Training. In stage III, DeVeiler is trained end-to-end in two phases. In phase I (pre-training), we train on paired source-domain data for $100k$ iterations using Adam [23], with a learning rate that decays from 2×10^{-4} to 1×10^{-7} via cosine annealing. In phase II (fine-tuning), we adapt the model on a mixed dataset of 500 generated pseudo-pairs and the original source pairs for $5k$ iterations. The learning rate decays from 5×10^{-5} to 1×10^{-7} with cosine annealing. Both phases use a batch size of 8, a 256×256 patch size, and random horizontal/vertical flips for augmentation. The loss weight λ_{rev} is set to 1.0. The fine-tuning phase is highly efficient, completing in approximately 40 minutes on an NVIDIA A100 GPU.

Text prompts. We utilize distinct text prompts for the source and target domains. The source domain prompt focuses solely on aberration: *a photograph with spatial-varying PSF blur, optical aberrations, defocus, and chromatic fringing.* For the target domains, which include the `Screen-Compound` and `Realworld-Compound`, the prompt is expanded to describe the full compound degradation: *a photograph with spatial-varying PSF blur, optical aberrations, defocus, chromatic fringing, and noticeable stray light with veiling glare.*

Figure 11. **Data Efficiency Analysis.** PSNR performance on the `Screen-Compound` test set with varying numbers of unpaired target training images (N). Note that $N = 0$ denotes the non-adapted Source-only baseline.

Baseline Implementations. For fair comparison, we use official implementations, retraining learning-based models on our dataset with default settings.

D. Discussion

D.1. Societal Impacts

This work alleviates the inherent trade-off between optical miniaturization and image quality. While aberration correction is established, veiling glare remains a bottleneck for compact optical systems. By addressing this compound degradation, we facilitate high-performance imaging in hardware with strict spatial limits, such as medical endoscopes, autonomous drones, and mobile devices. Beyond specific applications, our framework demonstrates a strategy for inverse problems in data-scarce domains. Instead of purely black-box approaches, we incorporate physical models into the latent space to synthesize realistic training pairs from unpaired data. This physics-aware generation strategy offers a scalable solution for other fields lacking GT, including underwater imaging and astronomy.

D.2. Limitations and Future Work

Limitations. Although our framework delivers favorable restoration results, slight color deviations may persist under intense veiling glare (see Fig. 13). This is fundamentally due to the ambiguity in overexposed regions, where the additive glare merges with the saturated scene signal. Regarding data acquisition, the source domain is constructed using images with a fixed scene depth. Although aberrations are theoretically depth-dependent, real-world performance indicates that this approximation is acceptable for compact optical systems. From a modeling perspective, the current

framework learns implicit light representations rather than explicit parameters (*e.g.*, 3D position). While sufficient for effective restoration, explicit parameterization could offer additional controllability in future iterations.

Future Work. To address these limitations, future research could incorporate explicit light-source modeling for color recovery and extend the source domain to include multiple scene depths. Additionally, we plan to investigate frequency-domain strategies for improved glare separation and adapt advanced generative backbones for efficient one-step synthesis.

E. More Results

Analysis of Target Data Efficiency We analyze the impact of target training set size ($N \in \{0, 5, 15, 25, 50\}$) on restoration performance. To decouple the adaptation strategy from the specific architecture of DeVeiler, we employ a standard SwinIR [29] backbone. Figure 11 indicates that a subset of $N = 5$ yields substantial improvements over the baseline ($N = 0$). Furthermore, performance converges at $N = 15$, reaching parity with the full dataset ($N = 50$). These results verify the capability of the framework for robust few-shot adaptation with limited data.

More Visual Results. To further verify the effectiveness of our method, we present more visual comparison results between the proposed DeVeiler and other advanced methods across both the `Screen-Compound` and `Realworld-Compound` domains. Specifically, the results captured by the SL and MRL systems are shown for the `Screen-Compound` domain in Fig. 12 and Fig. 13, and for the `Realworld-Compound` domain in Fig. 14 and Fig. 15, respectively.

In these scenarios, the input images exhibit compound degradation, where light sources introduce veiling glare, leading to a noticeable reduction in contrast and color shifts alongside intrinsic optical aberrations. Baseline aberration correction methods [29], trained on the source domain, show limited generalization to the unseen veiling glare. Cascaded pipelines employing dehazing models [49] can improve global contrast but may result in the smoothing of fine textures. Similarly, cascaded flare removal approaches [10], typically designed for localized artifacts, are less effective in addressing the spatially diffusive nature of veiling glare. Furthermore, general domain adaptation methods [22, 49], without explicit physical modeling for glare, may exhibit color deviations. In contrast, DeVeiler leverages the latent veiling glare maps to achieve favorable results across diverse scenarios, preserving structural details and recovering color fidelity. In regions of intense illumination, the input signal approaches saturation (Fig. 13). While signal loss hinders full recovery, DeVeiler minimizes color deviations compared to competing methods.

Figure 12. **Visual results on the `Screen-Compound` domain captured by the SL system.** The method is shown at the bottom of each case. Zoom in for the best view.

Figure 13. **Visual results on the `Screen-Compound` domain captured by the MRL system.** The method is shown at the bottom of each case. Zoom in for the best view.

Figure 14. **Visual results on `Realworld-Compound` domain captured by the SL system.** The method is shown at the bottom of each case. Zoom in for the best view.

Figure 15. **Visual results on `Realworld-Compound` domain captured by the MRL system.** The method is shown at the bottom of each case. Zoom in for the best view.