Towards Illumination-Aware Restoration of Metalens-Captured Images: A New Dataset and a Strong Baseline

Fen Fang[1], Xinan Liang[2, 3], Muli Yang[1*], Jionghong Zheng[1], Tobias W. W. Mass[4], Ying Sun[1], Xulei Yang[1*], Xuewu Xu[2, 3], Zhengguo Li[1]

[1]Institute for Infocomm Research (I^{2}R), A*STAR, Singapore
[2]Institute of Materials Research and Engineering (IMRE), A*STAR, Singapore
[3]National Semiconductor Translation and Innovation Center (NSTIC), Singapore
[4]MetaOptics Technologies, Singapore
{fangf, liangx, yangml, jzheng, suny, yangx, xux1, ezgli}@a-star.edu.sg, tobias@metaoptics.sg

Abstract

Metalenses offer compelling advantages such as lightweight and ultra-thin design, making them promising alternatives to conventional lenses. However, their widespread adoption is hindered by image quality degradation caused by chromatic and angular aberrations. To mitigate this, restoration processes are often necessary to recover high-quality RGB images from metalens-captured inputs. While recent deep learning-based restoration methods show promise, they typically (1) blur or distort peripheral regions, or (2) fail entirely under unseen illumination conditions. To advance metalens image restoration, we introduce **IlluMeta**—the first and largest real-world, illumination-aware metalens image dataset—captured across diverse lighting environments. In addition, we propose a novel end-to-end restoration framework that directs attention to challenging regions and adaptively adjusts to varying illuminations via reinforcement learning. Experiments show that our method can be applied in a plug-and-play manner to enhance existing models, significantly improving image restoration quality, especially under unseen lighting conditions, paving the way for broader real-world deployment of metalens technologies.

1 Introduction

Metalenses—ultra-thin, lightweight optical elements with subwavelength structures (Yu and Capasso 2014)—have recently attracted considerable attention for enabling compact imaging systems across applications such as smartphones (Martins et al. 2020; Wang et al. 2022a), UAVs, and AR/VR devices (Li et al. 2022). Despite these promising attributes, the adoption of metalens-based cameras in real-world scenarios remains limited due to significant image quality degradation caused by inherent chromatic and angular aberrations (Fan, Lin, and Su 2020). These artifacts are particularly severe in peripheral regions and are further compounded under non-uniform or low illumination, which frequently occurs in practical settings.

Recent advances in deep learning have shown promise in restoring high-quality images from metalens-captured inputs (Zamir et al. 2022b; Lee et al. 2024; Seo et al. 2024). However, as shown in Fig. 1, existing restoration models still suffer from two major limitations. First, they struggle to recover details in the outer regions of the image due to spatially-varying degradations caused by angular aberrations. These regions are typically under-optimized during training, as most models implicitly focus on the brighter and more informative image centers. Second, their performance drops significantly when tested under unseen illumination levels (ILs), because they are usually trained under fixed lighting conditions and fail to generalize across varying illumination profiles. These two challenges—*spatial degradation* and *illumination sensitivity*—are the main bottlenecks impeding robust deployment of metalens-based systems.

To address these limitations, we propose a two-pronged restoration framework that explicitly targets the above challenges. First, we introduce a Spatial-Aware Attention Scheduler that encourages the model to pay greater attention to peripheral regions during training. This design is motivated by the observation that angular aberrations introduce stronger degradations away from the image center, and that treating all regions equally in the loss function causes the model to neglect these harder regions. By weighting the reconstruction loss spatially, our method facilitates better structural fidelity across the entire image. Second, we propose a reinforcement learning (RL)-based Retraining-Free Illumination Adapter that adaptively modifies the illumination profile of the input image at inference time. Rather than retraining the restoration model for each lighting condition—which is both computationally expensive and hard to scale—we train a lightweight RL agent that learns to adjust image illumination on-the-fly to align with the pretrained model's preferences. This retraining-free, reward-driven strategy offers strong generalization without compromising efficiency.

A key enabler of our framework is the **IlluMeta** dataset—the first large-scale, real-world metalens image dataset captured under diverse illumination conditions. Existing datasets either lack real illumination variation, *e.g.*, DIV2K-Meta (Seo et al. 2024; Timofte et al. 2017) or are synthetic in nature, *e.g.*, DVD (Su et al. 2017), making it difficult to study illumination-aware restoration in realistic scenarios. In

*Corresponding authors.

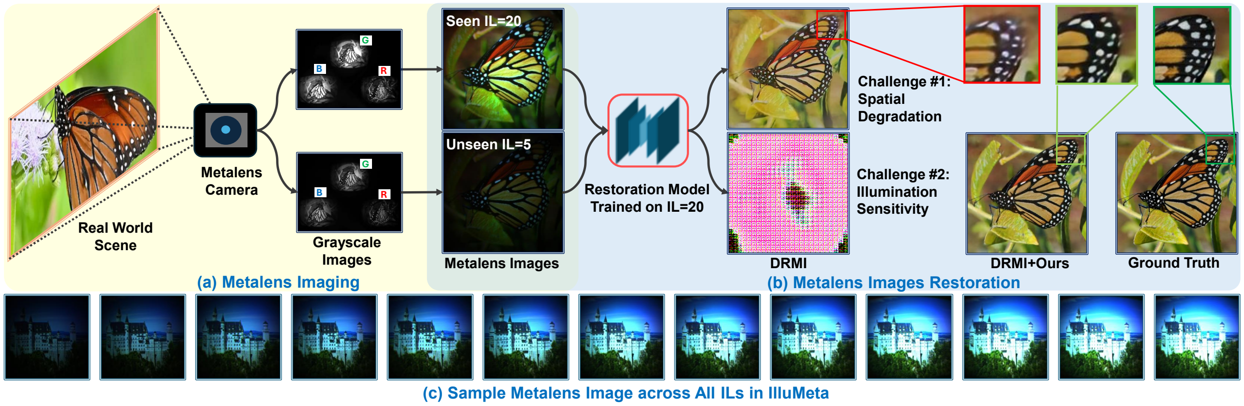

Figure 1: Overview of metalens imaging setup, restoration challenges, and IlluMeta dataset. **(a)** Metalens imaging setup: RGB images are reconstructed from monochrome channel captures of real-world scenes. **(b)** Two key challenges in restoring metalens images: #1 *Spatial degradation* in peripheral regions due to angular aberrations, as highlighted in the zoomed patches; #2 *Illumination sensitivity*: model trained under a specific illumination level (*e.g.*, IL=20) degrade significantly under unseen lighting conditions (*e.g.*, IL=5), as shown in the comparison between DRMI (Seo et al. 2024) and DRMI+Ours. **(c)** Sample images from IlluMeta under 13 distinct illumination levels (IL=2–50), demonstrating broad dynamic range and lighting diversity.

contrast, IlluMeta contains over 25,000 real-world metalens images, captured across 13 distinct illumination levels, and includes both daytime and nighttime scenes. This dataset enables the systematic evaluation of models under varying lighting conditions, encourages research into generalizable restoration techniques, and provides rich spatial degradation patterns for studying position-dependent image quality.

Together, our proposed method and dataset offer a comprehensive solution to the longstanding challenges of metalens imaging: spatial restoration of peripheral degradations and generalization to unseen illumination levels.

In summary, our contributions are three-fold:

- We identify and tackle two core challenges in metalens image restoration: peripheral degradation and illumination generalization.
- We propose a novel framework that integrates sptial aware attention and reinforcement learning-based illumination aware controller, achieving retraining-free robustness under diverse conditions.
- We construct IlluMeta, the first real-world, illumination-diverse metalens dataset, establishing a new benchmark for restoration methods under practical settings.

2 Related Work

Image Aberration Correction. Metalens aberration correction has been tackled using both learning-based and hybrid methods. Early CNN-Wiener filtering approaches (Tseng et al. 2021) struggled with fine detail recovery due to limited receptive fields. Recent works enhance restoration by combining diffusion models with Wiener filtering (Chakravarthula et al. 2023; Ho, Jain, and Abbeel 2020; Wiener 1949), or by adopting lightweight CNNs for fast, blind color correction (Eboli, Morel, and Facciolo 2022). A more recent trend involves physics-informed networks that incorporate lens-specific priors to improve generalization (Gong et al. 2024).

Image Deblurring/Denoising. Deep networks have achieved strong performance in denoising (Zhang et al. 2023), deblurring (Nah et al. 2021; Tsai et al. 2022), and general restoration. Leading CNN-based models include MIRNetv2 (Zamir et al. 2022a), SFNet (Lee et al. 2019), HINet (Chen et al. 2021), and NAFNet (Chu, Chen, and Yu 2022), while transformer-based designs such as Restormer (Zamir et al. 2022b), Uformer (Wang et al. 2022b), and others (Zhang et al. 2018; Lecouat, Ponce, and Mairal 2020; Zamir et al. 2021; Chen et al. 2023; Sun et al. 2023; Liang et al. 2024) handle a wide range of degradations. In meta-optical systems, joint optics-restoration frameworks (Lin et al. 2021; Mansouree et al. 2020), including MetaFormer (Lee et al. 2024) and DRMI (Seo et al. 2024), have shown promising results. Additionally, diffusion-based approaches are increasingly adopted for their generative power and robustness (Saharia et al. 2022; Yu et al. 2024; Öždenizci and Legenstein 2023).

3 IlluMeta Dataset Construction

Metalens Camera Configuration. Our imaging system employs a custom-built, multi-metalens RGB camera[1] based on a single-layer metasurface design. The system supports a wide field of view (FOV) of 70° and operates over a broad visible wavelength range. Three metalenses, fabricated on a single glass substrate, are individually optimized for red ($\lambda_R = 635$ nm), green ($\lambda_G = 540$ nm), and blue ($\lambda_B = 460$ nm) wavelengths. All three share an identical focal length of 1.5 mm and are positioned in close spatial proximity to ensure optical alignment. These metalenses project their respective spectral components onto separate regions of a monochrome CMOS sensor (Thorlab CS165MU(/M)), enabling parallel acquisition of R, G, and B information.

[1]The RGB metalens & IoT camera module is manufactured by METAOPTICS TECHNOLOGIES PTE LTD.

Dataset	FoV	Category #	Image #	Illum. Level #
DIV2K-Meta	23^o	18	698	1
IlluMeta (Ours)	70^o	26	25,558	13

Table 1: Summary of the comparison between existing metalens dataset and our constructed dataset-IlluMeta. FoV indicates field of view of the metalens camera.

Data Acquisition Setup. As illustrated in Fig. 1(a), we began by collecting 1,966 high-resolution RGB images (3840×2160 pixels), which were displayed on a calibrated screen and serve as ground-truth references. The metalens camera was used to capture the central 2160×2160 region of each image. Each capture comprises three grayscale sub-images—corresponding to R, G, and B channels—acquired separately and later fused to synthesize a color image. The final reconstructed image is downsampled to 400×400 pixels for standardization across the dataset.

Preprocessing Pipeline. To generate a full-color metalens image, we first align the wavelength-specific regions (R, G, B) from the captured grayscale image (shown in Fig. 1(a)) with pixel-wise accuracy. These subregions are then assigned to the respective RGB channels of a digital color image to form the final output. However, due to the intrinsic chromatic and angular aberrations of metalens optics—as well as imperfections in the synthesis process—the resulting images often exhibit noticeable blurring and strong vignetting, especially toward the periphery of the field of view. These distortions necessitate advanced learning-based restoration methods to enhance overall image quality.

Key Dataset Characteristics. The proposed **IlluMeta** dataset exhibits several important properties:

- *Diverse content coverage:* The dataset includes a wide variety of objects and scene types captured under both daytime and nighttime conditions, across outdoor and indoor, facilitating robust model generalization to real-world scenarios.
- *Illumination variability:* To simulate different lighting environments, we captured image under 13 screen brightness levels: {2, 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 45, 50}, with a fixed camera exposure time-100 ms. This allows for systematic evaluation of illumination sensitivity.
- *Large-scale volume:* The dataset contains over 25,000 metalens images, making it suitable for training and benchmarking modern deep learning-based models.

Sample images from IlluMeta are illustrated in Fig. 1. A comparative overview between IlluMeta and the widely-used DIV2K-Meta dataset (Seo et al. 2024) is presented in Tab. 1, with representative metalens images from both datasets shown in Fig. 2. o the best of our knowledge, IlluMeta is the *first real-world metalens dataset* that systematically incorporates diverse illumination levels (ILs), capturing the inherent variability of real-world lighting conditions. This makes IlluMeta a valuable benchmark for assessing the robustness of image restoration algorithms under illumination changes.

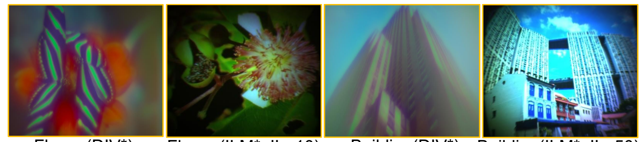

Figure 2: Comparison of Metalens Images between DIV2K-Meta (DIV*) and IlluMeta (ILM*) Datasets. While DIV2K-Meta exhibits uniform blur, IlluMeta presents more realistic degradations, such as non-uniform blur, peripheral distortions, and illumination-dependent variations, posing greater challenges for robust image restoration.

4 Our Approach

Our method addresses the challenge of metalens image restoration under varying illumination and edge degradation. It consists of two core components: (1) *Spatial-Aware Attention Scheduler* designed to enhance edge reconstruction quality, and (2) a *Retraining-Free Illumination Adapter* that dynamically adjusts image brightness during inference to improve generalization under diverse lighting conditions.

4.1 Spatial-Aware Attention Scheduler

Following typical restoration framework, *e.g.*, DRMI (Seo et al. 2024), we adopt a hybrid loss that combines pixel-wise PSNR loss L_{psnr} quantifying the image fidelity loss and adversarial loss L_{Adv} encouraging more perceptually realistic image generation in the Fourier domain. The overall training objective is: $L(\theta_{Res}) = L_{psnr}(\theta_{Res}) + \lambda L_{Adv}(\theta_{Res}, \theta_{Dis})$, where θ_{Res} and θ_{Dis} are the parameters of the restoration model and discriminator, and λ balances the two losses.

However, in metalens images, severe angular aberration leads to non-uniform degradation—especially in peripheral regions—posing challenges for standard convolutional models to generalize across spatially varying distortions. Although DRMI integrates positional embeddings to model angular dependencies, it remains limited in reconstructing high-fidelity details at the edges. To address this, we introduce a spatially varying attention map at the patch level. For a given pixel (x, y), the attention magnitude is defined as

$$\mathrm{Att}(x, y) = k\frac{\sqrt{(x-x_c)^2 + (y-y_c)^2}}{\text{patch\_size}} + 1\,, \qquad (1)$$

where (x_c, y_c) is the patch center, and k controls the attention scale. The modified loss becomes

$$L(\theta_{Res}) = \mathrm{Att} \cdot L_{psnr}(\theta_{Res}) + \lambda L_{Adv}(\theta_{Res}, \theta_{Dis})\,. \quad (2)$$

Empirically, we observe that emphasizing edge regions improves peripheral reconstruction but may degrade central fidelity if we fix k as 1. To mitigate this, we introduce a schedule-aware modulation by choosing a training-dependent function $k(e)$:

$$k(e) = \begin{cases} 0\,, & \text{if } e \le e_T \\ \frac{e}{e_{\max}} - \frac{1}{3}\,, & \text{otherwise} \end{cases}\,, \qquad (3)$$

where e is the current epoch, $e_T = \frac{1}{3}e_{\max}$ defines the onset of attention, and $e_{\max}$ is the total number of training epochs.

4.2 Retraining-Free Illumination Adapter

Deep restoration models trained on metalens data are highly sensitive to illumination conditions. Significant deviation in ILs between training and test samples often leads to degraded restoration quality, particularly under underexposed or overexposed conditions. While retraining the model with diverse ILs might seem a solution, it typically reduces performance on any specific IL due to increased data diversity. For instance, our experiments show a drop in PSNR from 24.20 to 22.71 when training across multiple ILs.

To overcome this, we propose a reinforcement learning (RL)-based module that dynamically adjusts the illumination of metalens images at inference time—without altering the restoration network. We adopt RL because (1) it enables adaptive decision-making based on environmental feedback without requiring ground truth—ideal for illumination adjustment in metalens images, where even humans struggle to determine the correct level; and (2) training the RL agent is significantly more efficient than retraining the restoration model, while achieving better performance, as shown in Fig. 7. Unlike traditional lenses, illumination in metalens images naturally attenuates from center to edges. Thus, IL adjustment must account for spatial variation. We define an adjustment function as

$$\hat{V}(i,j) = \begin{cases} \alpha V(i,j)\,, & \text{if } \alpha \leq 1 \\ \alpha V(i,j)\left(1 - \frac{D}{C(\beta+1)}\right)\,, & \text{otherwise} \end{cases}, \tag{4}$$

where $V(i,j)$ and $\hat{V}(i,j)$ represent the original and adjusted pixel values, respectively; D denotes the distance from pixel (i,j) to the image center; and C controls the strength of attenuation. The adjustment is governed by two parameters: α, which handles global scaling, and β, which provides spatial compensation to mitigate attenuation pattern distortion caused by overflow. We model this adjustment as a Markov Decision Process (MDP), with the key elements shown in Fig. 3 and detailed below.

State and Environment. The state for RL is the metalens image and the environment is the pretrained restoration model.

Action Space. The actions correspond to discrete estimation values for α and β. We define 15 discrete actions for each of a^α and a^β. The actions range from 0 to 15 with step size 1. With the estimated a^α, the value of α can be computed as $\alpha = \alpha_{base} + a^\alpha \times m$, where α_{base} and m are the base illumination scaling factor and coefficient of the effect of action a^α on the factor. The value of β corresponds to the predicted a^β itself.

Policy Network. The policy network consists of five convolutional layers, taking an input feature map of size 64 $\times$ 64 $\times$3 and producing an output feature map of size 8$\times$8. This is followed by a fully connected layer and two output branches: predicting actions for α and β respectively.

Reward Function. We use PSNR as the reward , hence the reward function is defined as

$$R(s, a^\alpha, a^\beta) = 10 \cdot \log_{10}(\frac{255^2}{\text{MSE}}), \tag{5}$$

where MSE (Mean Squared Error) is the average of the squared differences between corresponding pixels in the ground truth image x and restored image $\hat{x}$.

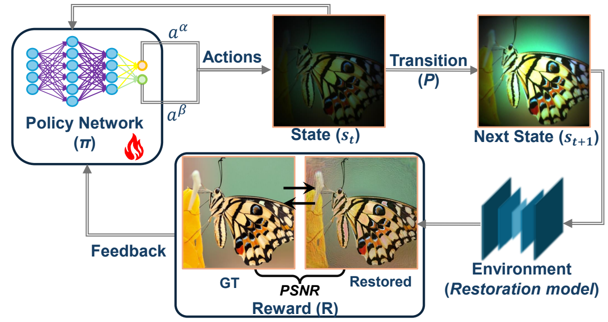

Figure 3: Architecture of the RL agent for illumination level control in metalens image restoration. At each time step t, given the current state (s_t), the policy network π selects actions a^α and a^β. The next state s_{t+1} is computed via Eq. (4), and subsequently passed to the environment, represented by a pretrained restoration model, which generates the corresponding restored image. This output is compared against the ground truth (GT) to calculate the PSNR severing as the reward R to update the policy.

We adopt DQN (Mnih et al. 2015) as base RL algorithm, the policy π optimization process can be estimated as

$$Q^\pi(s_t, a_t^\alpha, a_t^\beta) = E[R_t | s_t, a_t^\alpha, a_t^\beta]. \tag{6}$$

The policy optimization problem is then converted to action-value (Q-value) function optimization, namely $Q(\theta) \rightarrow Q(\theta^*)$. The parameter set θ can be learned via the policy network π by minimizing the MSE loss between the current Q-value $Q(s_t, a_t^\alpha, a_t^\beta; \theta)$ and target Q-value $\hat{y}_t$.

$$\begin{aligned} Q(s_t, a_t^\alpha, a_t^\beta; \theta) =& V^s(s;\theta^f,\theta^s) + V^a(s, a^\alpha, a^\beta; \theta^f, \theta^{a^\alpha}, \theta^{a^\beta}) \\ & - \frac{1}{N}\sum_{\hat{a}\in A} V^{\hat{a}^i}(s, \hat{a}^\alpha, \hat{a}^\beta; \theta^f, \theta^{\hat{a}^\alpha}, \theta^{\hat{a}^\beta}), \end{aligned} \tag{7}$$

$$\hat{y}_t = r_t + \gamma \max_{a_{t+1}, a \in \alpha,\beta} Q(s_t, a_{t+1}^\alpha, a_{t+1}^\beta; \theta), \tag{8}$$

where V^s represents the state value, V^a is the action value, θ^f is the parameters of common portion (purple part of policy network in Fig.3) of feature extractor. θ^s and θ^a are parameters of state value and action (output branches). $N = |A|$ is the cardinality of actions. The loss used to optimize the parameters of policy network is defined as:

$$L(\theta) = [\hat{y}_t - Q(s_t, a_t^\alpha, a_t^\beta; \theta)]^2. \tag{9}$$

After training the policy network, the optimal parameters θ^* is obtained, the optimal policy π^* then can be estimated by $Q(s, a^\alpha, a^\beta; \theta^*)$. Hence, the optimal actions can be predicted by:

$$\{a^\alpha, a^\beta\} = \underset{a^\alpha \in A_t, a^\beta \in A_t}{\arg\max}\; Q(S_t, a_{t-1}^\alpha, a_{t-1}^\beta; \theta^*). \tag{10}$$

Discussion. Our method can be applied in a plug-and-play manner to enhance existing models, significantly improving image restoration quality, especially under unseen lighting conditions. Specifically, the Spatial-Aware Attention Scheduler can be applied to other restoration models, such as NAFNet (Chu, Chen, and Yu 2022), HINet (Chen et al. 2021), and SFNet (Lee et al. 2019), as PSNR loss is one of the most commonly used objectives in image restoration tasks. Moreover, the Retraining-Free Illumination Adapter is model-agnostic and can be integrated with different restoration models by simply replacing the model component within the RL architecture, as illustrated in Fig. 3.

5 Experiments

Datasets. We evaluated our method on two datasets: (1) DIV2K-Meta (Seo et al. 2024), a widely used benchmark with 698 high-resolution images (1280×800), using 628 for training and 70 for evaluation; and (2) IlluMeta, our proposed dataset detailed in Sec. 3, using the first 1,000 of 1,966 images per illumination level (IL) for training and the remaining 966 for testing. The DIV2K-Meta dataset, with fixed illumination, provides a controlled setting to evaluate our Spatial-Aware Attention Scheduler. In contrast, the IlluMeta dataset, with diverse illumination levels, enables comprehensive assessment of both the Scheduler and the Retraining-Free Illumination Adapter under varying ILs.

Baselines. We evaluate our method by applying it to four representative baseline models—DRMI (Seo et al. 2024), NAFNet (Chu, Chen, and Yu 2022), HINet (Chen et al. 2021), and SFNet (Lee et al. 2019)—on both datasets. For each model, we compare performance before and after integrating our approach to assess its general applicability and improvement potential.

Experimental Setup. We design two experimental setups to evaluate the effectiveness, data efficiency, and generalization capability of our method.

Setup 1: Comparison with Joint Learning. As previously observed, training on a single IL often yields higher PSNR than joint training across multiple ILs. Here, we compare our method to a joint learning baseline. The baseline model (Model-A) is trained on a combined ILs (e.g., ILs {20, 30, 40}) and evaluated on all 13 ILs. In contrast, our method trains a model (Model-B) on a single IL (e.g., IL 30), then trains a RL agent using a subset (e.g., 50%) of images from the combined ILs, with Model-B as the environment. During inference, test images are first adapted by the RL agent and then restored by Model-B. Both approaches use the same ILs for training and testing, ensuring fair comparison.

Setup 2: Generalization Across ILs. To assess the adaptability of our method in more realistic scenarios, we allow the RL agent to access training subsets from additional ILs beyond those seen by the restoration model. This setup evaluates the agent's ability to enhance restoration performance under previously unseen illumination conditions. We conduct experiments on two restoration models—one trained on a single IL and another on multiple ILs—to demonstrate the robustness and general applicability of our approach.

Metrics. Consistent with previous works (Seo et al. 2024; Chu, Chen, and Yu 2022), we use PSNR and SSIM to assess the fidelity of image restoration, and LPIPS to evaluate the perceptual quality of the restored images by utilizing pre-trained deep learning networks (*e.g.*, AlexNet).

Method	PSNR↑ *Wh*	PSNR↑ *Per*	SSIM↑ *Wh*	SSIM↑ *Per*	LPIPS↓ *Wh*	LPIPS↓ *Per*
Raw Matelens	$14.72^{\pm1.33}$	$12.65^{\pm1.47}$	$0.43^{\pm0.16}$	$0.39^{\pm0.17}$	$0.79^{\pm0.11}$	$0.84^{\pm0.11}$
SFNet (2019)	$18.22^{\pm1.73}$	$13.62^{\pm1.93}$	$0.57^{\pm0.13}$	$0.51^{\pm0.15}$	$0.52^{\pm0.10}$	$0.64^{\pm0.13}$
SFNet+Ours	$\mathbf{18.47}^{\pm1.68}$	$\mathbf{14.13}^{\pm2.81}$	$\mathbf{0.59}^{\pm0.15}$	$\mathbf{0.54}^{\pm0.12}$	$\mathbf{0.50}^{\pm0.12}$	$\mathbf{0.58}^{\pm0.11}$
HINet (2021)	$21.36^{\pm2.33}$	$15.84^{\pm2.38}$	$0.64^{\pm0.12}$	$0.57^{\pm0.14}$	$0.46^{\pm0.10}$	$0.51^{\pm0.11}$
HINet+Ours	$\mathbf{21.62}^{\pm2.40}$	$\mathbf{16.08}^{\pm2.40}$	$\mathbf{0.65}^{\pm0.13}$	$\mathbf{0.62}^{\pm0.13}$	$\mathbf{0.44}^{\pm0.11}$	$\mathbf{0.49}^{\pm0.13}$
NAFNet (2022)	$21.69^{\pm2.38}$	$16.05^{\pm2.50}$	$0.64^{\pm0.12}$	$0.59^{\pm0.13}$	$0.44^{\pm0.10}$	$0.52^{\pm0.10}$
NAFNet+Ours	$\mathbf{21.98}^{\pm2.39}$	$\mathbf{16.52}^{\pm2.48}$	$\mathbf{0.67}^{\pm0.12}$	$\mathbf{0.63}^{\pm0.14}$	$\mathbf{0.43}^{\pm0.10}$	$\mathbf{0.51}^{\pm0.09}$
DRMI (2024)	$22.09^{\pm2.42}$	$16.56^{\pm2.58}$	$0.69^{\pm0.14}$	$0.65^{\pm0.14}$	$0.43^{\pm0.10}$	$0.51^{\pm0.10}$
DRMI+Ours	$\mathbf{22.17}^{\pm2.03}$	$\mathbf{17.46}^{\pm2.34}$	$\mathbf{0.70}^{\pm0.19}$	$\mathbf{0.66}^{\pm0.13}$	$\mathbf{0.43}^{\pm0.08}$	$\mathbf{0.49}^{\pm0.10}$

Table 2: Performance comparison of baseline methods on the DIV2K-Meta dataset, before and after applying our method. Results are reported as $mean^{std}$ on the *Wh (whole)* image or in the *Per (peripheral)* area.

Implementation Details. To better capture spatially variant degradations, we apply two pre-processing steps: (1) randomly crop 256×256 patches from full-resolution images (1280×800 for DIV2K-Meta, 400×400 for IlluMeta); and (2) for DRMI (Seo et al. 2024), add position embeddings via 1×1 convolution using patch coordinates. During evaluation, full-resolution images are used after applying TLC (Chu et al. 2022) to reduce statistical inconsistencies.

All results presented in the following sections for each setup are obtained using five-fold cross-validation, with $\alpha_{base} = 0.4$ and $m = 0.25$ used to compute α in Eq. 4.

Results on DIV2K-Meta. Tab. 2 compares four baseline methods on the DIV2K-Meta dataset, before and after applying our method. All baselines show improved performance on both the full image (left) and the outer $\frac{1}{4}$ regions (right). While overall gains are modest, improvements in the outer regions are notable, for instance, DRMI achieves a 0.9dB PSNR increase. These results validate the effectiveness of our Spatial-Aware Attention Scheduler.

Results on IlluMeta-Setup 1. Tab. 3 reports the performance and data efficiency of our method applied to four baseline models under Setup 1. Several observations can be made: *First*, models trained on a single IL (e.g., IL 30) outperform those jointly trained on ILs 20, 30, and 40 when evaluated on the seen IL. For example, DRMI achieves 23.7dB PSNR on IL 30 vs. 22.6dB on average with joint training. *Second*, using more ILs during training improves generalization. For instance, DRMI's average PSNR increases from 16.4 (IL 30 only) to 16.8 (ILs 20–40), and NAFNet's from 17.3 to 19.7 (highlighted in yellow). *Finally*, our method further improves performance using the same ILs. With restoration model trained on IL 30 and the RL agent trained on subsets from ILs 20–40, the model adapts effectively to varying ILs. DRMI improves PSNR from 16.8dB to 18.8dB, and NAFNet from 19.7dB to 19.9dB. These results highlight the ability of our approach to improve both data efficiency and restoration quality under varying illumination conditions.

	ILs	20	30	40	2	5	8	10	12	15	25	35	45	50	Avg.
		DRMI (Seo et al. 2024)													
PSNR↑	Orig.	19.1	23.7	20.6	9.20	10.7	12.0	12.9	14.1	15.7	20.8	22.1	17.8	16.5	16.4
	JT	22.8	22.4	22.8	7.37	7.83	7.95	8.82	12.6	18.8	21.8	22.0	22.4	21.0	16.8
	Ours	**22.7**	**23.4**	**22.6**	**9.86**	**11.9**	**14.1**	**15.7**	**17.5**	**20.6**	**21.9**	**21.9**	**22.2**	**20.8**	**18.8**
SSIM↑	Orig.	0.78	0.80	0.78	0.34	0.47	0.57	0.62	0.60	0.67	0.78	0.78	0.76	0.74	0.67
	JT	0.79	0.79	0.78	0.17	0.19	0.24	0.31	0.43	0.71	0.78	0.78	0.77	0.75	0.58
	Ours	**0.79**	**0.79**	**0.78**	**0.40**	**0.56**	**0.66**	**0.71**	**0.67**	**0.78**	**0.78**	**0.78**	**0.77**	**0.75**	**0.71**
LPIPS↓	Orig.	0.26	0.22	0.25	0.48	0.47	0.45	0.45	0.40	0.38	0.24	0.24	0.27	0.35	0.34
	JT	0.23	0.23	0.23	0.54	0.53	0.53	0.50	0.30	0.26	0.24	0.24	0.23	0.24	0.33
	Ours	**0.23**	**0.23**	**0.23**	**0.30**	**0.24**	**0.26**	**0.24**	**0.29**	**0.25**	**0.24**	**0.24**	**0.23**	**0.23**	**0.25**
		HINet (Chen et al. 2021)													
PSNR↑	Orig.	20.4	26.0	20.6	8.76	10.3	11.9	12.9	14.1	16.1	24.1	23.1	18.5	17.2	17.2
	JT	24.3	24.0	24.2	10.3	12.7	15.1	16.7	18.2	21.8	22.0	24.0	23.5	22.8	19.9
	OO	**23.8**	**23.5**	**23.6**	**11.8**	**15.4**	**18.3**	**20.6**	**20.8**	**23.4**	**23.5**	**23.2**	**23.0**	**22.3**	**21.0**
SSIM↑	Orig.	0.82	0.85	0.84	0.28	0.44	0.57	0.63	0.61	0.75	0.84	0.84	0.81	0.80	0.70
	JT	0.86	0.86	0.86	0.43	0.60	0.68	0.71	0.76	0.83	0.85	0.86	0.86	0.85	0.77
	Ours	**0.84**	**0.84**	**0.84**	**0.53**	**0.71**	**0.77**	**0.79**	**0.71**	**0.83**	**0.84**	**0.85**	**0.85**	**0.84**	**0.79**
LPIPS↓	Orig.	0.24	0.22	0.24	0.49	0.46	0.44	0.42	0.38	0.33	0.24	0.25	0.29	0.30	0.33
	JT	0.22	0.22	0.22	0.48	0.45	0.41	0.40	0.37	0.23	0.24	0.22	0.22	0.22	0.30
	Ours	**0.23**	**0.23**	**0.23**	**0.36**	**0.30**	**0.28**	**0.26**	**0.26**	**0.25**	**0.23**	**0.23**	**0.24**	**0.24**	**0.26**

	ILs	20	30	40	2	5	8	10	12	15	25	35	45	50	Avg.
		NAFNet (Chu, Chen, and Yu 2022)													
PSNR↑	Orig.	20.5	25.8	20.5	9.07	10.6	12.1	13.1	14.5	16.3	24.2	23.1	18.5	17.2	17.3
	JT	24.4	24.6	24.3	8.01	12.0	14.1	15.7	17.9	21.6	23.8	24.1	23.9	22.5	19.7
	Ours	**24.2**	**24.8**	**24.5**	**9.30**	**12.4**	**14.5**	**16.0**	**18.1**	**21.5**	**23.6**	**23.4**	**23.6**	**22.8**	**19.9**
SSIM↑	Orig.	0.81	0.85	0.83	0.31	0.46	0.57	0.63	0.63	0.75	0.84	0.84	0.81	0.79	0.71
	JT	0.85	0.85	0.85	0.33	0.57	0.67	0.70	0.71	0.83	0.85	0.85	0.85	0.84	0.75
	Ours	**0.85**	**0.85**	**0.85**	**0.35**	**0.58**	**0.68**	**0.71**	**0.71**	**0.83**	**0.85**	**0.85**	**0.85**	**0.85**	**0.76**
LPIPS↓	Orig.	0.23	0.20	0.23	0.51	0.48	0.45	0.42	0.35	0.31	0.20	0.20	0.28	0.30	0.32
	JT	0.20	0.20	0.20	0.53	0.45	0.41	0.39	0.28	0.21	0.20	0.20	0.20	0.21	0.29
	Ours	**0.20**	**0.20**	**0.20**	**0.49**	**0.44**	**0.40**	**0.39**	**0.28**	**0.21**	**0.20**	**0.20**	**0.20**	**0.21**	**0.28**
		SFNet (Lee et al. 2019)													
PSNR↑	Orig.	15.0	18.9	15.7	9.17	10.6	12.0	13.0	14.2	15.4	17.6	16.5	14.8	14.1	14.4
	JT	17.1	16.9	17.0	9.59	11.6	13.7	14.9	16.2	16.8	17.1	16.5	15.5	14.8	15.3
	OO	**18.5**	**18.0**	**17.1**	**9.83**	**12.1**	**14.3**	**15.6**	**17.3**	**18.2**	**18.3**	**17.3**	**16.6**	**16.7**	**16.1**
SSIM↑	Orig.	0.55	0.59	0.57	0.26	0.35	0.42	0.45	0.49	0.51	0.56	0.57	0.55	0.54	0.49
	JT	0.56	0.55	0.55	0.31	0.35	0.41	0.44	0.61	0.64	0.66	0.63	0.47	0.45	0.51
	Ours	**0.65**	**0.62**	**0.56**	**0.33**	**0.47**	**0.56**	**0.59**	**0.63**	**0.64**	**0.63**	**0.58**	**0.54**	**0.54**	**0.56**
LPIPS↓	Orig.	0.39	0.26	0.38	0.48	0.47	0.44	0.41	0.38	0.37	0.28	0.32	0.38	0.39	0.38
	JT	0.31	0.28	0.28	0.46	0.44	0.42	0.38	0.30	0.30	0.29	0.28	0.28	0.29	0.33
	Ours	**0.28**	**0.27**	**0.28**	**0.45**	**0.43**	**0.40**	**0.33**	**0.27**	**0.28**	**0.27**	**0.27**	**0.27**	**0.28**	**0.31**

Table 3: Performance comparison between three settings: **(a)** the original restoration model (Orig.) trained on IL 30, **(b)** a jointly trained (JT) model using ILs 20, 30, 40, and **(c)** the original model combined with our method (Ours). Cells with a gray background indicate illumination levels (ILs) used to train the restoration model; blue background denotes ILs used exclusively to train the RL agent; cells without background represent ILs unseen by both the restoration model and the RL agent.

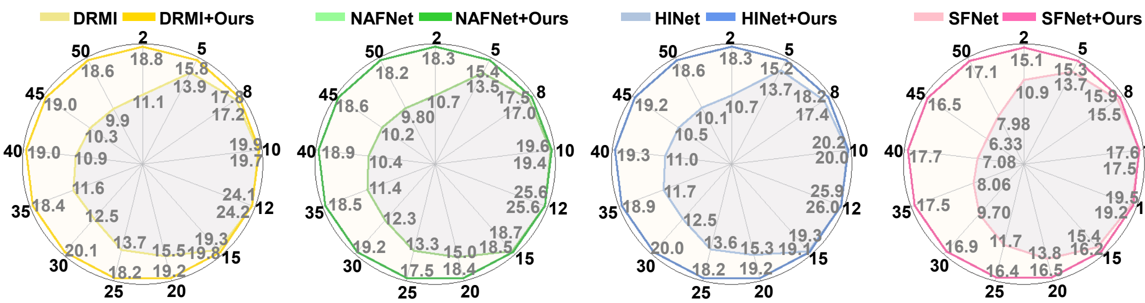

Figure 4: Comparison of PSNR (dB) between baseline models and their counterparts integrated with our method across all illumination levels (ILs) in Setup-2. The restoration models are trained on IL 12, the RL agents are trained on a subset of ILs {2, 10, 20, 30, 40, 50}.

DRMI	M1	M2	PSNR↑ *Wh*	PSNR↑ *Per*	SSIM↑ *Wh*	SSIM↑ *Per*	LPIPS↓ *Wh*	LPIPS↓ *Per*
√			16.8	14.4	0.58	0.49	0.33	0.40
√	√		17.3	15.3	0.61	0.51	0.32	0.38
√		√	18.2	16.5	0.69	0.62	0.27	0.31
√	√	√	**18.8**	**16.6**	**0.71**	**0.63**	**0.25**	**0.29**

Table 4: Ablation study on the impact of Attention Scheduler (M1) and the Illumination Adapter (M2) in our method on the Wh (whole) image or in the Per (peripheral) area

		Raw Metalens			**DRMI**			**DRMI+Ours**		
Object	GT	2	5	10	2	5	10	2	5	10
Person	0.80	0.51	0.56	0.61	0.03	0.02	0.70	0.58	0.68	0.76
Car	0.82	0.56	0.58	0.63	0.19	0.19	0.70	0.58	0.69	0.71

Table 5: Comparison of AP for "person" and "car" detection across ground truth (GT), metalens, restored images by DRMI and restored ones by DRMI combined our method. The restoration model was trained on IL 30, while the RL agent in our method was trained on ILs { 20, 30, 40}.

Results on IlluMeta-Setup 2. To further assess the generalization of our method, we conduct experiments where the RL agent accesses training subsets from a wider range of ILs, allowing it to better adapt illumination and attenuation parameters to those used in the restoration model's training.

PSNR comparisons between baseline models and their counterparts augmented with our method (based on the IL-12 model) are shown in Fig. 4. As shown, our method consistently improves performance across all baselines, especially on ILs that differ significantly from the training IL (e.g., ILs 40, 45, 50). Improvements on similar ILs (e.g., ILs 8, 10, 15) are smaller. These results highlight the method's robustness to unseen and distribution-shifted ILs.

Visualization. Fig. 5 shows two visual comparisons of restoration results between DRMI and our method under Setup 1. It is clear that directly applying the restoration model alone produces bad or even unrecognizable results, whereas our method significantly improves image quality.

Computation Efficiency Analysis. As mentioned earlier, our RL agent is trained on subsets of ILs unseen by the restoration model. In Tab. 3 and Fig. 4, this subset includes

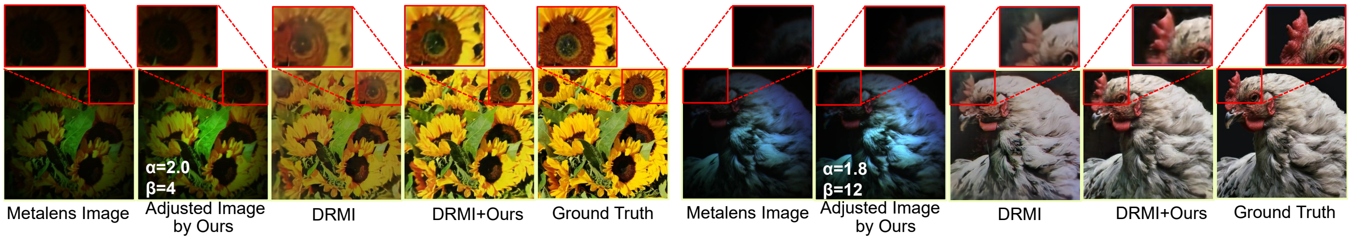

Figure 5: Comparison of two restoration examples using DRMI and DRMI enhanced with our method. The adjustment factors α and β, estimated by the RL agent, adapt the illumination levels (ILs) of metalens images for improved restoration. The original DRMI model is trained on ILs $\{20, 30, 40\}$, whereas our method uses a model trained only on IL 30, along with an RL agent trained on $\{20, 30, 40\}$. Both test images are from IL 8, which is unseen by both the restoration model and the RL agent.

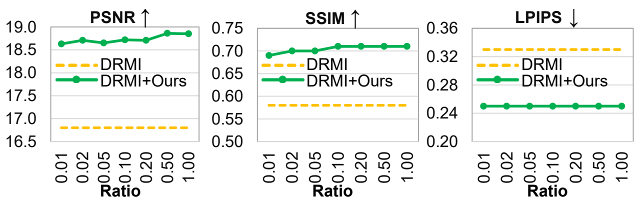

Figure 6: Data efficiency comparison of our method. Yellow lines represent joint training on ILs $\{20, 30, 40\}$, while green lines show our method using a DRMI model trained on IL 30 and an RL agent trained on subsets of data randomly sampled (*e.g.*, 1%) from the combined ILs.

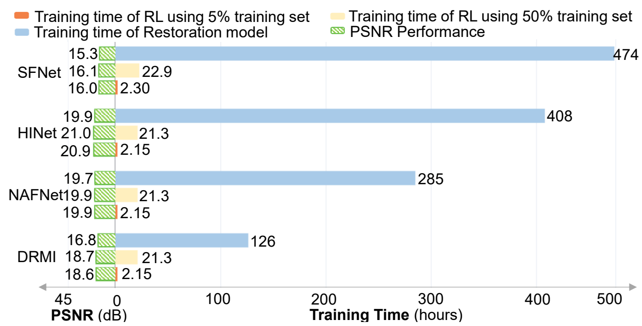

Figure 7: Training cost comparison in Setup-1.The restoration model is initially trained on IL 30. Blue bars represent retraining with additional ILs 20 and 40, while orange and yellow bars show our method using an RL agent trained on 5% and 50% of their data, respectively. This highlights the data and compute efficiency of our approach.

50% of training data from training set of each IL. To assess data efficiency, we train RL agents using varying data ratios (1%, 2%, 5%, 10%, 20%, 50% and 100%) from each IL and evaluate the restored images using the pretrained model. Fig. 6 shows that even with just 1% of data from unseen ILs, the RL agent can effectively adjust images, yielding PSNR comparable to full-data training.

We also compare training costs between two approaches: retraining the restoration model with unseen ILs vs. training our RL agent using only 5% and 50% of unseen IL data, as shown in Fig. 7. Results show that our method requires significantly less additional training time for unseen ILs.

Ablation Study. Our method includes two modules: the Spatial-Aware Attention Scheduler and the Retraining-Free Illumination Adapter, evaluated via ablation studies. Tab. 4 shows average performance across all ILs using different combinations with DRMI under Setup-1. Each module individually improves restoration: the Attention Scheduler boosts PSNR by 0.5dB overall and 0.9dB in Peripheral regions, while the Illumination Adapter alone adds 1.4dB. Combining both yields the best results, highlighting their complementary strengths. Additional validation of the Attention Scheduler on DIV2K-Meta is shown in Tab. 2.

Performance on Object Detection. We further evaluate our method on the downstream task of object detection. A total of 185 "person" and 52 "car" instances were selected from both ground truth and metalens images captured under ILs 2, 5 and 10. Detection was performed using the YOLOv8x (Jocher, Chaurasia et al. 2023) model pretrained on the MSCOCO dataset (Lin et al. 2014). Tab. 5 reports the average precision (AP) for both categories. Ground truth images yield the highest AP (0.8), while our method consistently outperforms both direct detection and DRMI restoration, particularly under unseen ILs (2 and 5). These results highlight the effectiveness of our approach in enhancing structural fidelity for downstream detection tasks.

6 Conclusions

In this work, we introduce a new dataset—**IlluMeta** specifically designed to facilitate the study of metalens image restoration under diverse illumination levels (ILs). To tackle the unique challenges posed by this setting, we propose a novel framework comprising two key components: (1) a Spatial-Aware Attention Scheduler that encourages the restoration model to prioritize structurally challenging regions, and (2) a reinforcement learning-based Retraining-Free Illumination Adapter that dynamically enhances performance across varying and unseen ILs. Extensive experiments confirm the strong effectiveness and generalization of our method, particularly under illumination conditions not seen during training. Our work advances research on metalens image restoration under diverse illumination conditions, bridging the gap toward real-world metalens applications.

Acknowledgments

This work was supported by the Agency for Science, Technology and Research (A*STAR) under its GAP project (Grant No. I24D1AG062) and its MTC Programmatic Funds (Grant No. M23L7b0021).

References

Chakravarthula, P.; Sun, J.; Li, X.; Lei, C.; Chou, G.; Bijelic, M.; Froesch, J.; Majumdar, A.; and Heide, F. 2023. Thin On-Sensor Nanophotonic Array Cameras. *ACM Transactions on Graphics*, 42(6): 1–18.

Chen, L.; Lu, X.; Zhang, J.; Chu, X.; and Chen, C. 2021. HINet: Half Instance Normalization Network for Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 182–192.

Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; and Dong, C. 2023. Activating More Pixels in Image Super-Resolution Transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22367–22377.

Chu, X.; Chen, L.; ; Chen, C.; and Lu, X. 2022. Improving Image Restoration by Revisiting Global Information Aggregation. In *ECCV*, 53–71.

Chu, X.; Chen, L.; and Yu, W. 2022. NAFSSR: Stereo Image Super-Resolution Using NAFNet. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1239–1248.

Eboli, T.; Morel, J.-M.; and Facciolo, G. 2022. Fast Two-Step Blind Optical Aberration Correction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 13666 of *Lecture Notes in Computer Science*, 693–708. Springer.

Fan, C.; Lin, C.; and Su, G. J. 2020. Ultrawide-angle and high-efficiency metalens in hexagonal arrangement. *Scientific Reports*, 10.

Gong, J.; Yang, R.; Zhang, W.; Suo, J.; and Dai, Q. 2024. A Physics-Informed Low-Rank Deep Neural Network for Blind and Universal Lens Aberration Correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 24861–24870. IEEE Computer Society.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 6840–6851. Curran Associates, Inc.

Jocher, G.; Chaurasia, A.; et al. 2023. YOLOv8: Ultralytics Official Implementation. https://github.com/ultralytics/ultralytics. Accessed: 2025-07-23.

Lecouat, B.; Ponce, J.; and Mairal, J. 2020. Fully Trainable and Interpretable Non-Local Sparse Models for Image Restoration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 238–254.

Lee, B.; Kim, Y.; Jo, Y.; Kim, H.; Park, H.; Kim, Y.; Mandal, D.; Chakravarthula, P.; Kim, I.; and Park, E. 2024. MetaFormer: High-Fidelity Metalens Imaging via Aberration Correcting Transformers. *arXiv preprint arXiv:2412.04591*. Available at https://arxiv.org/abs/2412.04591.

Lee, J.; Kim, D.; Ponce, J.; and Ham, B. 2019. SFNet: Learning Object-Aware Semantic Correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2278–2287.

Li, Z.; Pestourie, R.; Park, J.; Huang, Y.; Johnson, S.; and Capasso, F. 2022. Inverse design enables large-scale high-performance meta-optics reshaping virtual reality. *Nature Communications*, 13.

Liang, J.; Cao, J.; Fan, Y.; Zhang, K.; Ranjan, R.; Li, Y.; Timofte, R.; and Gool, L. V. 2024. VRT: A Video Restoration Transformer. *IEEE Transactions on Image Processing*, 33: 2171–2182.

Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft COCO: Common Objects in Context. *CoRR*, abs/1405.0312.

Lin, Z.; Roques-Carmes, C.; Pestourie, R.; Soljačić, M.; Majumdar, A.; and Johnson, S. G. 2021. End-to-end nanophotonic inverse design for imaging and polarimetry. *Nanophotonics*, 10(3): 1177–1187.

Mansouree, M.; Kwon, H.; Arbabi, E.; McClung, A.; and Faraon, A. 2020. Multifunctional 2.5D metastructures enabled by adjoint optimization. *Optica*, 7(1): 77–84.

Martins, A.; Li, K.; Li, J.; Liang, H.; Conteduca, D.; Borges, B.; Krauss, T.; and Martins, E. 2020. On Metalenses with Arbitrarily Wide Field of View. *ACS Photonics*, 7(8): 2073–2079.

Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A. A.; Veness, J.; Bellemare, M. G.; Graves, A.; Riedmiller, M.; Fidjeland, A. K.; Ostrovski, G.; et al. 2015. Human-level control through deep reinforcement learning. *Nature*, 518(7540): 529–533.

Nah, S.; Son, S.; Lee, S.; Timofte, R.; and Lee, K. M. 2021. NTIRE 2021 Challenge on Image Deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 149–165.

Öždenizci, O.; and Legenstein, R. 2023. Restoring Vision in Adverse Weather Conditions with Patch-Based Denoising Diffusion Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Saharia, C.; Chan, W.; Chang, H.; Lee, C. A.; Ho, J.; Salimans, T.; Fleet, D. J.; and Norouzi, M. 2022. Palette: Image-to-Image Diffusion Models. In *Proceedings of the ACM SIGGRAPH Conference*. ArXiv:2111.05826.

Seo, J.; Jo, J.; Kim, J.; Kang, J.; Kang, C.; Moon, S.-W.; Lee, E.; Hong, J.; Rho, J.; and Chung, H. 2024. Deep-learning-driven end-to-end metalens imaging. *Advanced Photonics*, 6: 066002.

Su, S.; Delbracio, M.; Wang, J.; Sapiro, G.; Heidrich, W.; and Wang, O. 2017. Deep Video Deblurring for Hand-Held Cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 237–246.

Sun, L.; Dong, J.; Tang, J.; and Pan, J. 2023. Spatially-Adaptive Feature Modulation for Efficient Image Super-Resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 13190–13199.

Timofte, R.; Agustsson, E.; Van Gool, L.; Yang, M.-H.; Zhang, L.; Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K. M.; et al. 2017. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 126–135.

Tsai, F.; Peng, Y.; Lin, Y.; Tsai, C.; and Lin, C. 2022. Stripformer: Strip Transformer for Fast Image Deblurring. In Avidan, S.; Brostow, G.; Cissé, M.; Farinella, G. M.; and Hassner, T., eds., *European Conference on Computer Vision (ECCV)*, volume 13679 of *Lecture Notes in Computer Science*, 146–162. Springer.

Tseng, E.; Colburn, S.; Whitehead, J.; Huang, L.; Baek, S.-H.; Majumdar, A.; and Heide, F. 2021. Neural nano-optics for high-quality thin lens imaging. *Nature Communications*, 12: 6493.

Wang, F.; Geng, G.; Wang, X.; Li, J.; Bai, Y.; Li, J.; Wen, Y.; Li, B.; Sun, J.; and Zhou, J. 2022a. Visible Achromatic Metalens Design Based on Artificial Neural Network. *Advanced Optical Materials*, 10(21): 2101842.

Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022b. Uformer: A General U-Shaped Transformer for Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17683–17693.

Wiener, N. 1949. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. Cambridge, MA: MIT Press. ISBN 0-262-23002-X, 0-262-25719-X. Originally circulated as a classified report in 1942.

Yu, F.; Gu, J.; Li, Z.; Hu, J.; Kong, X.; Wang, X.; He, J.; Qiao, Y.; and Dong, C. 2024. Scaling Up to Excellence: Practicing Model Scaling for Photo-Realistic Image Restoration In the Wild. *arXiv preprint arXiv:2401.13627*.

Yu, N.; and Capasso, F. 2014. Flat optics with designer metasurfaces. *Nature Materials*, 13(2): 139–150.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.; and Shao, L. 2021. Multi-Stage Progressive Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14821–14831.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.; and Shao, L. 2022a. Learning Enriched Features for Fast Image Restoration and Enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022b. Restormer: Efficient Transformer for High-Resolution Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5718–5729.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2018. Simple baselines for image restoration. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*.

Zhang, Y.; Li, Y.; Zhang, Y.; and Timofte, R. 2023. NTIRE 2023 Challenge on Image Denoising: Methods and Results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 188–197.