

Learning to See Through Flare

Xiaopeng Peng¹, Heath Gemar², Erin Fleet², Kyle Novak², Abbie Watnik², Grover Swartzlander¹
¹Rochester Institute of Technology, ²U.S. Naval Research Laboratory

Abstract

Machine vision systems are susceptible to laser flare, where unwanted intense laser illumination blinds and distorts its perception of the environment through oversaturation or permanent damage to sensor pixels. We introduce NeuSee, the first computational imaging framework for high-fidelity sensor protection across the full visible spectrum. It jointly learns a neural representation of a diffractive optical element (DOE) and a frequency-space Mamba-GAN network for image restoration. NeuSee system is adversarially trained end-to-end on 100K unique images to suppress the peak laser irradiance as high as 10^6 times the sensor saturation threshold I_{sat} , the point at which camera sensors may experience damage without the DOE. Our system leverages heterogeneous data and model parallelism for distributed computing, integrating hyperspectral information and multiple neural networks for realistic simulation and image restoration. NeuSee takes into account open-world scenes with dynamically varying laser wavelengths, intensities, and positions, as well as lens flare effects, unknown ambient lighting conditions, and sensor noises. It outperforms other learned DOEs, achieving full-spectrum imaging and laser suppression for the first time, with a 10.1% improvement in restored image quality.

1. Introduction

Continuous advancements of laser technology have enabled the ready availability of low-cost, compact, and powerful lasers which, if misdirected toward an image sensor, may cause objectionable dazzle (e.g., sensor saturation and lens flare) or irreversible anomalies. As illustrated in Fig.1, lasers can disrupt vision and mislead the tracking system of unmanned aerial vehicles [1–3]. Adversarial laser attacks against the sensor of autonomous or robotic vehicles have been demonstrated to significantly compromise their safety and reliability [4–6]. Lasers also present risks to sensors in mixed reality devices (e.g., video see-through head-mounted displays). These devices may advance the development of eye protection goggles [7–10] by providing high-quality imaging capability that may otherwise not be possible. They



Figure 1. Illustration of sensor damage risks under the laser illumination (row 1) and applications of sensor protection (row 2), which includes mitigating disruption and improving vision performance of autonomous vehicles, robots, consumer cameras, as well as mixed reality devices for eye protection in scientific experiments, aviation, medical treatment, entertainment laser shows, and more.

are crucial in scientific experiments, aviation [11], law enforcement, manufacturing [12], medical treatment [13], and more. In addition, laser-induced damage to consumer camera sensors has been reported during entertainment events, such as laser shows [14].

The laser-induced saturation and damage of an imaging sensor depend on both the sensor and the laser characteristics. The damage thresholds of silicon-based imaging sensors are typically six to nine orders of magnitude higher than their saturation thresholds [15, 16]. To mitigate laser-related sensor risks and image degradation, optical techniques such as wavelength multiplexing [17, 18], non-linear optics [19, 20], holographic coatings [21], liquid crystals [22, 23], meta-materials [24, 25], integration time reduction [26], smoke obscurants [27]. Learned diffractive optical elements (DOEs) have also been explored, but their performance is typically constrained by limited field-of-view or spectral bandwidth [28–33]. None of these approaches has been found to simultaneously satisfy the desired bandwidth, response time, dynamic range, stability, and image quality characteristics.

In this work, we present NeuSee, a computational imaging framework that jointly learns a neural representation of a diffractive optical element (DOE) and a Mamba-GAN-based image restoration network operating in the frequency domain. NeuSee is trained end-to-end in an adversarial manner on 100K unique images, enabling it to suppress peak laser

irradiance up to 10^6 times the sensor saturation threshold I_{sat} , a level that could otherwise cause sensor damage without the use of a DOE. The framework comprises a learnable neural phase mask that attenuates laser light while preserving the transmission of the background scene. It also features a novel restoration network that corrects multiple image degradations, such as inpainting the saturated area, reducing image blur, and removing noise, resulting in a high-quality image. The contributions are summarized as follows.

- We present the first framework that jointly learns a diffractive optical element (DOE) and image restoration to protect sensors from high-energy lasers across the full visible spectrum.
- The learned DOE achieves both high background-light throughput and laser suppression. A neural DOE representation with state-space GAN restoration and a two-stage learning strategy disentangles conflicting objectives.
- We curate a 100K high-resolution dataset with a physics-based synthesis pipeline, generating diverse training images across laser spectra, intensities, positions, ambient illumination levels, and noise.
- Our system outperforms existing sensor-protection DOEs in laser suppression ratio, imaging spectrum, laser spectral coverage, and image quality.

2. Related Work

End-to-End Learned Camera System. Computational imaging and photography are emerging areas that focus on improving and extending the capabilities of traditional imaging and camera systems using optical and computational methods. By altering light transmission at the pupil plane using an amplitude mask or a phase mask, the coded aperture approach has been investigated in many applications, such as coded exposure [34], achromatic imaging [35, 36], high dynamic range [37, 38], lens glare suppression [39, 40], extended depth of field [41–43], light field imaging [44, 45], lensless imaging [46], privacy-preserving imaging [47, 48], granular imaging [49, 50], hyperspectral imaging [51, 52], AR glasses [53–55] and more. Instead of relying solely on separately learned optics, recent advances in computational imaging have embraced end-to-end optics learning [56–59], where phase masks are co-optimized with image restoration algorithms. This approach integrates differentiable simulations of optical systems with task-driven neural reconstruction models, enabling greater adaptability and system-level optimization. Unlike traditional designs that primarily target optical aberrations in isolation, end-to-end methods consider the full imaging pipeline, tailoring both optics and algorithms to the specific task. The resulting phase masks learn to encode latent image information into PSFs, which is subsequently decoded through the reconstruction network.

Image Restoration. Learning a direct mapping from the sensor image s to the background scene radiance $\hat{b}$ using

DNN has been extensively studied for many low-level computer vision tasks [60], including image deblurring, denoising, deraining, dehazing, in- and outpainting, as low light, and many computational imaging applications [29, 61–63]. Vision transformers [64–66] seek to capture long-range image dependencies using self-attention mechanisms. Alternatively, the spatial gating unit [67] was introduced with the MLP to achieve a performance comparable to that of transformers [68]. Recent advances in state-space sequence models (SSMs or Mamba) [69] also demonstrated strong efficiency and effectiveness in vision tasks [70–72]. Additionally, changing the image basis through linear or non-linear transforms to better suit the orthogonally preconditioned optimizers was found to accelerate the convergence and improve the performance of the DNNs. One notable finding involves the use of FFT features to correct spectral bias and improve the learning of high-frequency functions from low dimension data [73–75]. Furthermore, by breaking down a complicated task into subproblems and solving them progressively, multistage and multiscale frameworks allow supervision and feature fusion in multiple restoration stages and multiple image scales [76, 77] and encourage the recovery of image details. Embedding of kernel functions [78] and image coordinates [79–81] into neural networks has also been introduced, respectively, to utilize the knowledge of system PSF and positional information of an image.

Generative models seek to learn the joint distribution $\Pr(s, b)$, which allows for more accurate recovery of missing data such as image saturation. Unlike VAEs [82] and flows [83], GANs do not rely on an explicit inference model. Instead, they learn the target distribution from an input distribution by seeking a Nash equilibrium between a generator and a discriminator through a minimax game. Diffusion models [84, 85] learn the implicit latent structure of a dataset by modeling the way in which data points diffuse through the latent space. Generative methods face the performance trilemma [86] of sampling quality, diversity, and speed. GAN models outperform many VAEs and flows in generating realistic looking images. While diffusion models can generate high-quality images with improved sample diversity [87, 88], their potential in real-time applications is limited due to the high cost of iterative sampling. The single-step distilled diffusion models remain underperforming in sample quality compared to GANs [89, 90].

3. Image Formation Model

Physics-based modeling enables accurate characterization of imaging systems and supports end-to-end optimization of optical elements and restoration algorithms, with efficient transfer from simulation to real data. We next outline an image formation model based on wave propagation.

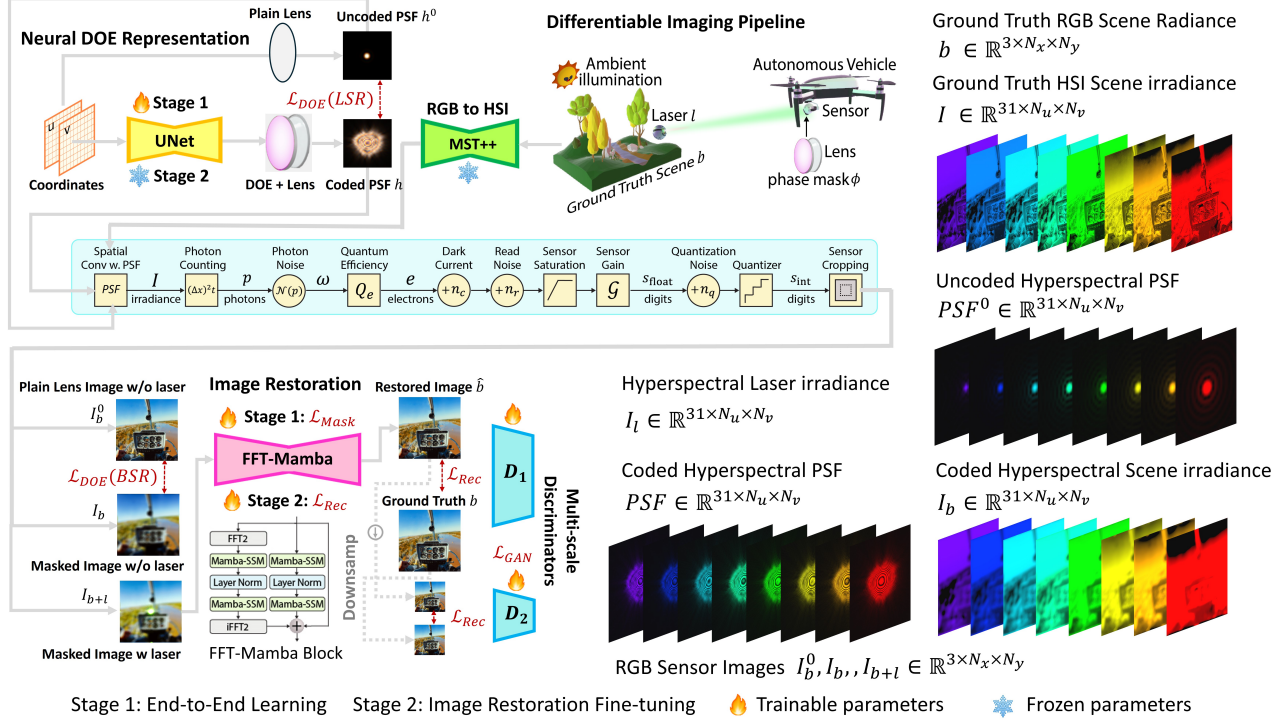


Figure 2. The proposed NeuSee system is a jointly learned imaging framework designed to protect sensors from laser dazzle across the full visible spectrum. Our system involves **1) A physics-based** differentiable imaging pipeline. To achieve accurate simulation of diverse scene and lasers, pre-trained hyperspectral (HSI) reconstruction net is used to transform each RGB radiance map to a 31-band HSI volume; **2) UNet-based neural representation** of the DOE (or phase mask); **3) A novel Mamba-GAN architecture** for image restoration, which makes use of image intensity and frequency information. The training of NeuSee includes two stages. **Stage 1** trains the DOE jointly with Mamba-GAN restoration with DOE loss and GAN loss. In **Stage 2**, the quality of the restored image is finetuned by freezing the DOE and training the Mamba-GAN with reconstruction and GAN loss.

3.1. RGB to HSI Irradiance

While proven effective in other computational imaging tasks [57, 58], ensuring continuous laser coverage is critical for our application. Consequently, simulating point spread functions (PSFs) and sensor responses using only three RGB channels is insufficient. To overcome this limitation, and given the scarcity of hyperspectral imaging (HSI) data, we employ a pre-trained spectral reconstruction network, G_{hsi} (MST++ Net [91]), to recover 31-channel scene irradiance $b(x, y, \lambda)$ in the visible spectrum (400–700 nm) from RGB ground-truth irradiance $b(x, y, 3)$. This approach enables us to leverage the broader availability of RGB ground-truth datasets. The transformation is formulated as:

$$b_{\lambda}(x, y, 31) = G_{hsi}(b(x, y, 3)) \quad (1)$$

3.2. Neural Representation of the DOE

In this study, the incident laser is characterized as a plane wavefront at the entrance pupil. To achieve sensor protection, we introduce a learnable diffractive optical element (DOE) at the entrance pupil of the system, the height of which is

written as a mapping of the pupil Cartesian coordinates of pupil plane (u, v) by a phase representation net G_{rep} :

$$h_{DOE}(u, v) = G_{rep}(\text{concat}(u, v)) \quad (2)$$

here we use an 8-layer UNet as the phase representation net. Assuming the DOE has a wavelength dependent refractive index $\Delta n(\lambda)$, the phase distribution of the DOE is given by:

$$\phi_{DOE}(u, v) = \frac{2\pi}{\lambda} \cdot \Delta n(\lambda) \cdot h_{DOE}(u, v) \quad (3)$$

3.3. System Point Spread Function

The spectral irradiance distribution in the focal plane (x, y) is described by wavelength-dependent pupil function and point spread function (PSF):

$$U_{\lambda}(x, y, \lambda) = \frac{e^{ikf}}{i\lambda f} \iint A(u, v) e^{i\phi(u, v)} e^{ik(xu+yv)/f} du dv \quad (4a)$$

$$PSF_{\lambda}(x, y, \lambda) = \left| U(x, y, \lambda) \right|^2 \quad (4b)$$

Table 1. Physical Parameters

Parameter	Symbol	Value
Background wavelength	λ_b	400-700 nm
Laser central wavelength	λ_l	400-700 nm
Effective focal length	f	0.11 m
Exposure time	t	0.1 sec
Aperture diameter	W_a	11 mm
Quantum efficiency	Q_e	0.56
Sensor gain	$\mathcal{G}$	0.37
Full well capacity	e_{sat}	$25500e^-$
Read noise (mean)	μ_r	$390e^-$
Read noise (std.)	σ_r	$10.5e^-$
Dark current	σ_c	$0.002e^-$
Bit depth per channel	bpc	16
Pupil pitch	$\Delta u = \Delta v$	$3.74 \mu\text{m}$
Sensor pitch	$\Delta x = \Delta y$	$2.9 \mu\text{m}$
Pupil resolution	$N_u \times N_v$	2160×2160
Sensor resolution	$N_x \times N_y$	2048×2048

where f is the focal length of the imaging lens, $k = 2\pi/\lambda$ is the wave number, and the integration is made over the interior of a circular aperture $A(u, v)$ of radius R .

To accelerate the computation of Eq. 4, we make use of the efficient Scaled Fresnel method [92] that relates the Fourier transform of the field in the pupil with that at the sensor planes when the two domains have different pitch of the pixels. When a camera captures a scene with a strong light source in presence, the resulting image may exhibit lens flare artifacts that caused by light scattering. These flares can manifest in various forms, such as halos, streaks, color bleeding, and haze. They are typically caused by dirt, scratches, windshield dents, grease, or smudges, or a mixture of them. Inspired by previous work [93–95], we also simulate these lens flare effects as shown in Fig. 4.

3.4. Sensor Image

Irradiance Distribution. A shift-invariant imaging system integrates the radiance distribution over the solid angle that is extended by the aperture through spatial convolution $*$, resulting in an irradiance map in the image plane. It is assumed that the background illumination has a visible wavelength λ_b and the laser has a narrow wavelength λ_l . For the phase-coded system, we represent the irradiance distributions in the sensor plane of the background scene and a laser spot:

$$I_b(x, y, \lambda) = \mathcal{F}^{-1}(\mathcal{F}(b_\lambda) \cdot U_\lambda) \cdot T_b(\lambda) \quad (5a)$$

$$I_l(x, y, \lambda) = \delta(x - \Delta l_x, y - \Delta l_y) \cdot PSF(x, y, \lambda) \cdot T_l(\lambda) \quad (5b)$$

where $b_\lambda(x, y)$ is the ground truth radiance map of the background scene, $T_b(\lambda)$ is the simulated CIE daylight spectral

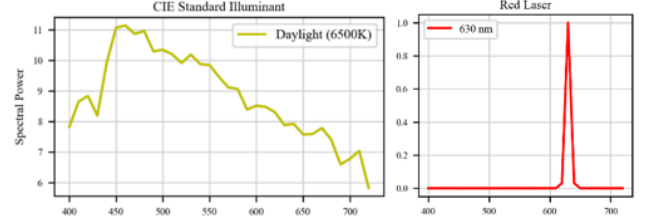


Figure 3. Spectral profile of scene illumination and a red laser.

curve shown in Fig. 3, $T_l(\lambda)$ represents the spectral profile of the laser characterized by a Gaussian distribution centered at the wavelength λ_l and having a full width at half maximum bandwidth $\Delta\lambda_{\text{FWHM}} = 10$ nm. The laser is considered a Dirac Delta function, which targets the sensor at a normal $\vec{n}_l = (n_u, n_v)$ with respect to the optical axis. Its footprint shift on the focal plane is thus given by $(\Delta l_x, \Delta l_y) = f \cdot \vec{n}_l$. $\mathcal{F}$ and $\mathcal{F}^{-1}$ represents respectively the Fourier and its inverse transform, and $\delta(\cdot)$ represents the Dirac delta function, where we assume the values of $b_\lambda(x, y)$ are normalized to the range $[0, 1]$. By replacing the coded PSF h with the uncoded PSF h_0 in Eq. 5, the irradiance maps of the background scene and the laser are defined as I_{b0} and I_{l0} respectively for an unprotected system.

Sensor Saturation. For a given wavelength, the irradiance value that saturates a sensor is expressed as:

$$I_{\text{sat}}(\lambda) = e_{\text{sat}} \cdot \frac{\hbar \cdot c}{\lambda \cdot t \cdot (\Delta x)^2 \cdot Q_e} \quad (6)$$

where e_{sat} , Q_e , and Δx are respectively the full well capacity, quantum efficiency, and pixel pitch of the sensor, $\hbar = 6.63 \cdot 10^{-34}$ [J · s] is Planck's constant, $c = 3 \cdot 10^8$ [m/s] is the speed of light in vacuum, and t is the exposure time. For an unprotected system, let us express the peak irradiances of the background scene and the laser spot, respectively, as $I_{b0, \text{peak}}$ and $I_{l0, \text{peak}}$, which are proportionate to the irradiance saturation value:

$$I_{b, \text{peak}} = \alpha_b \cdot BSR \cdot I_{\text{sat}}(\lambda_b) \quad (7a)$$

$$I_{l, \text{peak}} = \alpha_l \cdot LSR \cdot I_{\text{sat}}(\lambda_l) \quad (7b)$$

where α_b and α_l are respectively the strength of the background illumination and the laser. The laser saturates a single pixel when $\alpha_l = 1$ and multiple pixels when $\alpha_l > 1$. Sensor damage may occur at $\alpha_l > 10^6$. For an optical system protected by a pupil plane phase mask the corresponding peak irradiance values of the background scene and laser are respectively scaled by a background suppression value BSR and the laser suppression value LSR . The values $BSR \sim 1$ and $LSR \ll 1$ lower the risk of sensor saturation and damage while maintaining the transmission rate of the scene irradiance.

Photon to Electron. Photons arrival at a sensor has a Poisson distribution, the rate of which is determined by the

image irradiance I , the pixel pitch Δx , wavelength λ , and the integration time t :

$$p(\lambda) = \frac{(I_b + I_l) \cdot \lambda \cdot t \cdot (\Delta x)^2}{h \cdot c} \quad (8)$$

The hyperspectral photons are then converted to RGB space via the following equation

$$p = \begin{bmatrix} p_R \\ p_G \\ p_B \end{bmatrix} = k \sum_{i=1}^N S(\lambda_i) I(\lambda_i) \begin{bmatrix} \bar{x}(\lambda_i) \\ \bar{y}(\lambda_i) \\ \bar{z}(\lambda_i) \end{bmatrix} \Delta \lambda \quad (9)$$

where $N = 31$

$$\begin{cases} S(\lambda_i) : \text{spectral reflectance or radiance at wavelength } \lambda_i, \\ I(\lambda_i) : \text{spectral power distribution (illuminant) at } \lambda_i, \\ \bar{x}(\lambda_i), \bar{y}(\lambda_i), \bar{z}(\lambda_i) : \text{CIE 1931 color matching functions,} \\ k : \text{normalization constant, } k = \frac{100}{\sum_{i=1}^N I(\lambda_i) \bar{y}(\lambda_i) \Delta \lambda}, \\ \Delta \lambda : \text{wavelength sampling interval.} \end{cases}$$

According to the central limit theorem, the Poisson distribution may be approximated by a Gaussian distribution, which was found to be a better characterization of our sensor in practice. The Gaussian distributed photon is given by $\omega \sim \mathcal{N}(c_1 \cdot \mu_p, c_2 \cdot \sigma_p)$, where its mean and standard deviation are written respectively as the modulated mean (μ_p) and standard deviation (σ_p) of the photon arrival rate p , and c_1 and c_2 are the modulation coefficients. Given a quantum efficiency Q_e , the collected photons are converted to electrons: $e = Q_e \cdot \omega$, followed by noise corruptions and the digitization process:

$$s = \text{crop} \left(\min \left(s_{\text{sat}}, \left\lfloor \mathcal{G} \cdot \min(e_{\text{sat}}, e + n_d + n_c) + n_q \right\rfloor \right) \right) \quad (10)$$

where unwanted electrons generated by other factors are modeled as additive dark current n_c and read noise n_r . The dark current has a Poisson distribution $n_c \sim \mathcal{P}(\mu_c)$, and the read noise is Gaussian distributed $n_r \sim \mathcal{N}(\mu_r, \sigma_r)$. The mean values μ_c, μ_r and the standard deviation σ_r of the noise are obtained through sensor calibration.

Digitization. Electrons are converted into an array of integer digital counts that represents the image recorded by the sensor. The total number of electrons that exceeds the full well capacity of the sensor e_{sat} is clipped. Electrons are then amplified by a sensor gain $\mathcal{G}$, producing an array of floating points. Uniformly distributed quantization noise $n_q \sim \mathcal{U}(-0.5, 0.5)$ is added to these digits, and floating-point digital values are then quantized to integer

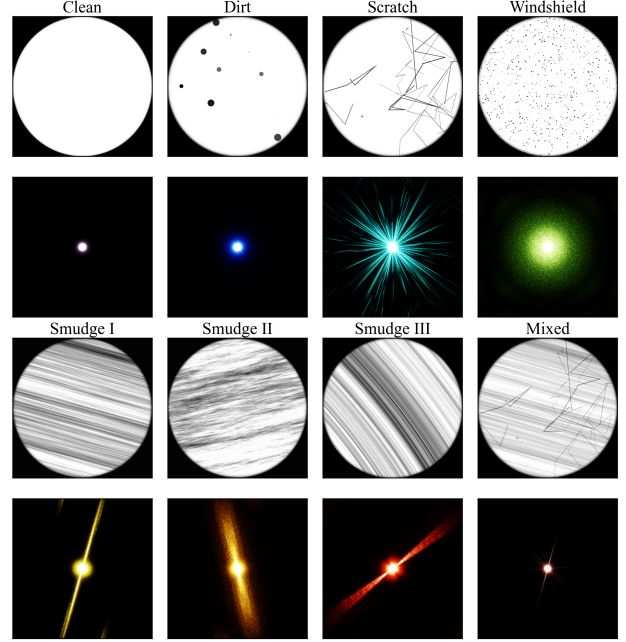


Figure 4. Lens flare effects under laser illumination at various wavelengths, caused by lens imperfections such as dirt, scratches, windshield dents, grease, smudges, or their combinations.

digital counts. The upper limit of the digital counts is determined by the bit depth per channel (bpc) of the sensor, where $s_{\text{sat}} = 2^{\text{bpc}} - 1$. The size of the image recorded by the sensor is determined by the finite size (W_s, H_s) of that sensor. Consider that a radiance map in the object plane has a size (W_o, H_o) and the system PSF has a width W_h , the size of the image formed in the focal plane is given by ($W_o + W_h, H_o + W_h$). Boundary areas that exceed the sensor size are cropped. The values of the physical parameters used in the simulation match the experiment (see Table 1).

4. Image Restoration

Here we introduce a Mamba-GAN for image restoration in frequency space, where the background scene radiance $\hat{b}$ is restored from the sensor image s :

$$\hat{b} = G_{\text{res}}(s) \quad (11)$$

The image restoration generator G_{res} and phase representation net G_{rep} are adversarially trained with a multiscale discriminator $D = \{D^L | L = 0, 1\}$. As shown in Fig. 2, the architecture of G_{res} consists of 8 layers FFT-Mamba blocks.

To better recover image details, a coarse-to-fine architecture is established. The restored image $\hat{b}$ is downsampled by antialiased bicubic interpolation and refined by the generator, producing estimated radiance maps $\{\hat{b}^L\}$, where $L = 0, 1$ represent coarse and fine scales respectively. The discriminators $D = \{D^L | L = 0, 1\}$ determine whether the estimate is real or false on each scale. The adversarial objective at

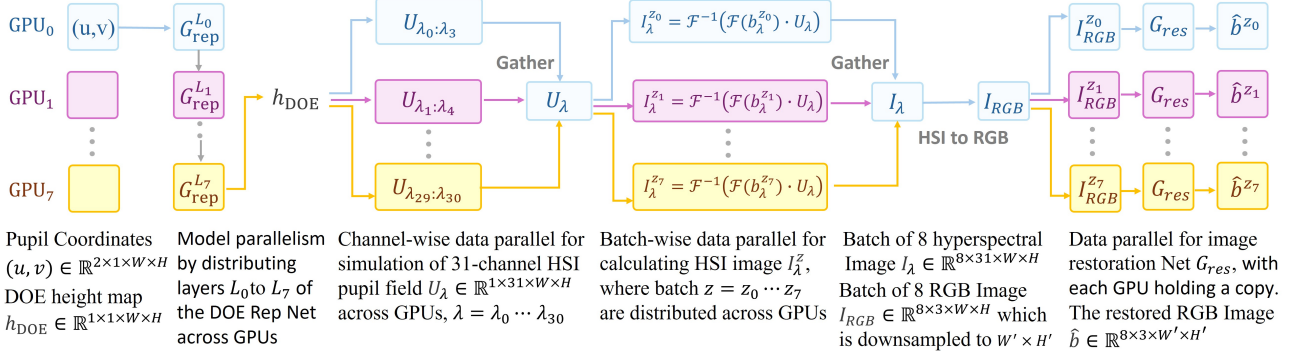


Figure 5. Illustration of heterogeneous data and model parallelism for distributed end-to-end training of the NeuSee system.

this stage is written as:

$$\mathcal{L}_{GAN,G} = -\lambda_{ADV} \cdot \sum_L \mathbb{E}[D^L(\hat{b}^L)] \quad (12a)$$

$$\begin{aligned} \mathcal{L}_{GAN,D} = & \lambda_{ADV} \cdot \sum_L \mathbb{E}[D_2^L(b^L)] - \mathbb{E}[D^L(\hat{b}^L)] \\ & - \lambda_{GP} \cdot \sum_L \mathbb{E}[(\|\nabla_{\tilde{b}^L} D^L(\tilde{b}^L)\|_2 - 1)^2] \end{aligned} \quad (12b)$$

where $\tilde{b}$ is sampled uniformly along a straight line between a pair of estimated and ground truth radiance maps $\hat{b}$ and b . $\lambda_{ADV} = 0.1$ and $\lambda_{GP} = 1$.

The DOE (or phase mask) learning objective consists of two terms. the laser suppression loss $\mathcal{L}_{DOE}(LSR) = \sum I_l(\lambda)/I_{l0}(\lambda)$ encourages minimizing the laser suppression ratio, while the background suppression loss $\mathcal{L}_{DOE}(BSR) = \sum I_{b0}(\lambda)/I_b(\lambda)$ encourages maximizing the transmission rate of non laser-induced light, such as background scene:

$$\mathcal{L}_{DOE} = \mathcal{L}_{DOE}(LSR) + \mathcal{L}_{DOE}(BSR) \quad (13)$$

The reconstruction objective consists of two terms. The first term is given by the Charbonnier L_1 difference between the estimated and the ground truth, where the fine-scale ground truth image is downsampled from the coarse-scale ground truth image using an anti-aliasing bicubic method [96]. In this method, the high-frequency components that cause aliasing artifacts are filtered by a low-pass cubic kernel. The second term is a FFT objective is expressed as the sum of absolute difference between the Fourier transforms of the ground truth radiance map and estimated radiance map at fine- and coarse scales: radiance pyramids:

$$\mathcal{L}_{REC} = \sum_{L=0,1} \sqrt{|b^L - \hat{b}^L|^2 + \epsilon} + |\mathcal{F}(b^L) - \mathcal{F}(\hat{b}^L)| \quad (14)$$

4.1. Two-Stage Training

Suppressing laser irradiation while maintaining background light transmission presents two conflicting objectives. A

common strategy in multi-objective learning is to jointly train a single network across tasks, exploiting shared representations for greater efficiency and performance than training tasks independently. While this can yield positive transfer, disparities in task difficulty and objective, data distribution, or optimization dynamics often lead to gradient interference and representation bias, reducing overall accuracy [97]. To address this issue, we propose a two-stage learning method that first learns shared features through joint training of the neural DOE and image restoration neural net and then fine-tunes parameters specific to image restoration while freezing the neural DOE.

In the **first stage**, we jointly train the DOE (phase mask) and image restoration networks, keeping both the phase representation net and restoration net parameters learnable. We optimize the mask objective and adversarial term with gradient accumulation, which provides a more stable update direction when handling noisy gradients from wide laser strength variations:

$$\min_{G_{rep}, G_{res}} \left(\left(\max_{D_1, D_2} \mathcal{L}_{GAN} \right) + \mathcal{L}_{DOE} \right) \quad (15)$$

In the **second stage**, we fine-tune the image restoration network with the phase net frozen, updating only restoration weights via mini-batch optimization of reconstruction objective and adversarial terms. Gradients are reset each iteration to avoid inaccuracies from accumulation:

$$\min_{G_{rep}, G_{res}} \left(\left(\max_{D_1, D_2} \mathcal{L}_{GAN} \right) + \mathcal{L}_{Rec} \right) \quad (16)$$

In each stage, the generators and the discriminators are trained in an alternating manner. Denote $L_{GAN} = L_{GAN,G}$ and $L_{GAN} = L_{GAN,D}$ as the adversarial objectives for generators and discriminators respectively, the generators seek to minimize the adversarial and reconstruction objectives, while the discriminators D_1 and D_2 aim to maximize only the adversarial terms. Using only the generators at the inference stage, the radiance of the scene is restored in an end-to-end manner.

5. Experiment Setup

Deep learning systems rely on large-scale and high-quality data to achieve optimal results [98, 99]. A set of 100K unique color images of versatile contents and 4K resolution (3860×2160) is collected[100] as RGB input scene radiance b for training, and another 1K testing images are collected for testing. During training, coded sensor images are numerically simulated in an online manner. To match the resolution of our laboratory camera sensor (2048×2048), each image is randomly cropped. The simulated sensor images are then downsampled to 256×256 using antialiased bicubic interpolation [96] to reduce the computational cost of neural network training. Laser strengths α_l are randomly sampled from 100K predetermined values, which are uniformly distributed in the range of $[0, 2e6]$. The incident angles of laser $\vec{n}_l = (n_u, n_v)$ are normally sampled, with the "3-sigma" (three times the standard deviation) set to $0.36 \cdot [f/W_s, f/H_s]$ along each axis, where f is the focal length, and W_s and H_s are the sensor width and height respectively. The models are trained with various noise strengths, where the dark current is normally sampled with a standard deviation equal to half its mean $\mu_c = 0.002e^-$. The read noise is uniformly sampled with $\mu_r \in [350, 400]e^-$ and $\sigma_r \in [10, 11]e^-$. The Gaussian-distributed photon noise has coefficients uniformly sampled $c_1 \in [0, 25\%]$ and $c_2 \in [0.9, 1.1]$. A variety of background illumination strengths $\alpha_b \in [0.3, 0.7]$ are considered. Exposure times t are normally sampled with a mean of 0.1 seconds, and the standard deviation is 0.1 times the mean.

5.1. Model Training

Our NeuSee system was trained on eight A100-80G GPUs. To accommodate the imaging pipeline with high-resolution hyperspectral data ($2048 \times 2048 \times 31$, batch size 8) and simultaneous training of multiple networks, we employ heterogeneous data and model parallelism for distributed training, as shown in Fig. 5. Adam optimizer with momentum $\beta_1 = 0.9, \beta_2 = 0.999$ is adopted in training. Our system is trained with learning rates of $2e-4$ for weights and $4e-4$ for biases. In the first stage, training runs for 10,000 iterations until the minimum LSR is reached, followed by a second stage of 20,000 iterations. Learning rates are halved after the first 2,000 iterations and further reduced by 70% every 1,000 iterations. The biases and weights are respectively zero and orthogonally initialized.

6. Results and Discussion

Fig. 6 compares the laser suppression ratio, phase mask, and PSF of our NeuSee system with the half-ring mask [28, 30], which was optimized using a genetic algorithm. Our NeuSee achieves a suppression over a broader laser spectral, which on average provides $5\times$ stronger suppression of visible laser

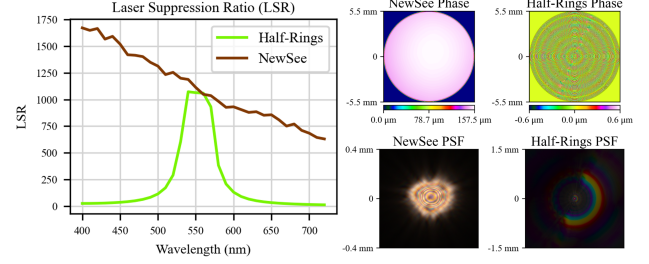


Figure 6. Comparison of laser suppression ratio, DOE (or phase mask), and PSF between our jointly learned NeuSee system and the heuristically learned half-ring system [28, 30]. Our NeuSee achieves, on average, 100 times stronger suppression of visible laser light than the half-ring.

light than the half-ring mask.

Qualitative evaluation of our NeuSee imaging is shown in Figure 7, where the proposed NeuSee results are compared with a traditional half-ring DOE that is learned using heuristic methods [30]. Background illumination strength $\alpha_b = 0.7$, coefficients of photon noise $c_1 = 20\%$ and $c_2 = 1.0$, dark current noise $\mu_c = 0.002e^-$, as well as the read noise $\mu_r = 390e^-$ and $\sigma_r = 10.5e^-$ remain the same in simulation. We showcase the presence of five lasers at violet, blue, green, yellow, red bands. Each has a different strength, ranging from mild $\alpha_l = 1e4$ to potentially damaging laser dazzle ($\alpha_l = 1e6$) strengths. A laser free case is also demonstrated. In both the laser-free and laser-dazzle cases, our NeuSee outperforms the heuristically learned half-ring phase mask and produces the consistently highest-fidelity reconstructions for the anti-dazzle imaging. Quantitative evaluations are performed on a set of 7K test images simulated from a thousand groundtruth scenes and seven laser strengths $\alpha_l = \{0, 10^k | k = 1, 2, \dots, 6\}$. Other parameters follow the sampling scheme of the training set. Compared to the half-ring, our NeuSee improves the restored image quality by 10.1% (L1 metric on average).

7. Conclusion

We present NeuSee, a computational imaging framework that jointly learns a neural diffractive optical element (DOE) and a frequency-space Vision Mamba-GAN network for image restoration. Trained end-to-end on 100K images, it protects sensors from laser flare and damage using a linear, broadband, and instantaneous optical mechanism. The DOE provides immediate protection, with system latency determined only by post-processing. In simulation, NeuSee achieves an irradiance dynamic range of 10^6 over sensor saturation and remains robust under varying laser intensities, angles, ambient lighting, and noise, enabling potential applications in autonomous vehicles, security cameras, HDR imaging, and laser-safe headsets.

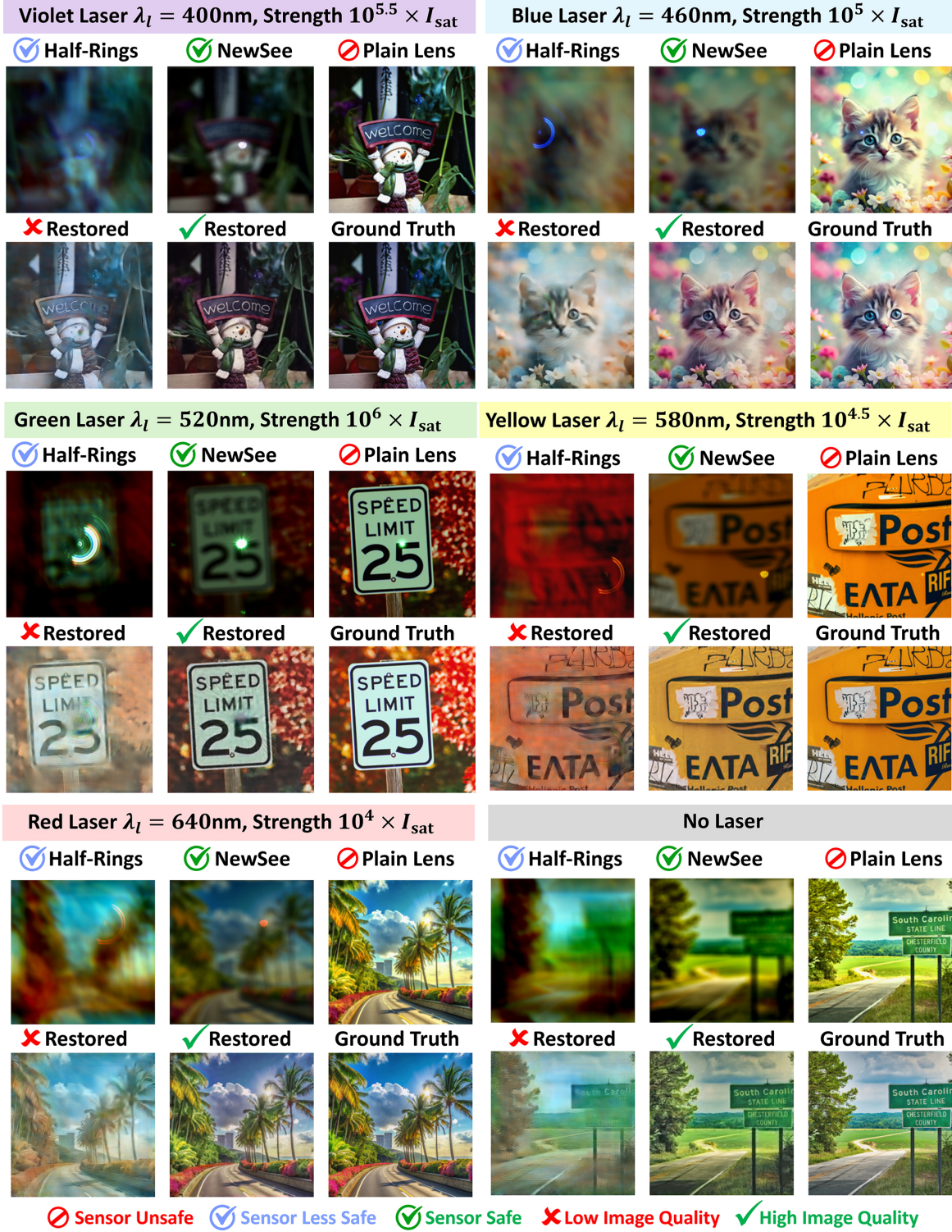


Figure 7. Row 1 compares sensor image of NeuSee (learned in stage-1), the heuristically learned half-ring DOE [30], and plain lens (lens without mask) systems. Row 2 compares image quality between NeuSee (two-stage learning) and half-ring (Mamba-GAN restored) against the corresponding ground truth images. Six lasers (10 nm full width at half maximum, 0.1% floor noise) are presented. Our NeuSee outperforms the half-ring DOE in not only better laser suppression and sensor protection but also higher-fidelity image.

References

- [1] Ove Steinvall. The potential role of laser in combating uav: part 2; laser as a countermeasure and weapon. In *Technologies for Optical Countermeasures XVIII and High-Power Lasers: Technology and Systems, Platforms, Effects V*, volume 11867. SPIE, 2021. 1
- [2] Ove Steinvall. Laser dazzling: an overview. In *Technologies for Optical Countermeasures XIX*, volume 12738, pages 17–31. SPIE, 2023.
- [3] Gareth D Lewis, Alexander Borghraef, and Marijke Vandewal. The disruptive impact of dynamic laser dazzling on template matching algorithms applied to thermal infrared imagery. In *Technologies for Optical Countermeasures XIX*, volume 12738, page 1273803. SPIE, 2023. 1
- [4] Ranjie Duan, Xiaofeng Mao, A Kai Qin, Yuefeng Chen, Shaokai Ye, Yuan He, and Yun Yang. Adversarial laser beam: Effective physical-world attack to dnns in a blink. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16062–16071, 2021. 1
- [5] Kyulim Kim, Jeongsoo Kim, Seungri Song, Jun-Ho Choi, Chulmin Joo, and Jong-Seok Lee. Engineering pupil function for optical adversarial attacks. *Optics Express*, 30(5): 6500–6518, 2022.
- [6] Yitong Sun, Yao Huang, and Xingxing Wei. Embodied laser attack: Leveraging scene priors to achieve agent-based robust non-contact attacks. In *ACM Multimedia 2024*. 1
- [7] Franco Quercioli. Beyond laser safety glasses: augmented reality in optics laboratories. *Applied optics*, 56(4):1148–1150, 2017. 1
- [8] Grzegorz Owczarek, Mieszko Wodzyński, Joanna Szkudlarek, and Marcin Jachowicz. Virtual reality (vr) for laser safety training. In *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)*, pages 1–3. IEEE, 2021.
- [9] JM Deniel and S Thommet. Occupational eye protection using augmented reality: a proof of concept. *Radioprotection*, 57(2):165–173, 2022.
- [10] Ke Li, Aradhana Choudhuri, Susanne Schmidt, Tino Lang, Reinhard Bacher, Ingmar Hartl, Wim Leemans, and Frank Steinicke. Mixed reality for laser safety at advanced optics laboratories. In *International Laser Safety Conference*, number PUBDB-2023-07345. Control-System, 2023. 1
- [11] Laser incidents. www.faa.gov/about/initiatives/lasers/laws. 1
- [12] Laser hazard. www.osha.gov/laser-hazards. 1
- [13] Deepthi Malayanur, Venkataram Nagaraj Mysore, et al. Laser safety eyewear. *CosmoDerma*, 2, 2022. 1
- [14] www.ilda.com/camera-sensor-damage.htm. 1
- [15] Gunnar Ritt. Laser safety calculations for imaging sensors. *Sensors*, 19(17), 2019. 1
- [16] Francis Théberge, Michel Auclair, Jean-François Daigle, and Dominik Pudo. Damage thresholds of silicon-based cameras for in-band and out-of-band laser expositions. *Applied Optics*, 61(10):2473–2482, 2022. 1
- [17] Gunnar Ritt, Bastian Schwarz, and Bernd Eberle. Preventing image information loss of imaging sensors in case of laser dazzle. *Optical Engineering*, 58(1):013109, 2019. 1
- [18] Gunnar Ritt and Bernd Eberle. Use of complementary wavelength bands for laser dazzle protection. *Optical Engineering*, 59(1):015106, 2020. 1
- [19] Garreth J Ruane, Abbie T Watnik, and Grover A Swartzlander. Reducing the risk of laser damage in a focal plane array using linear pupil-plane phase elements. *Applied optics*, 54(2):210–218, 2015. 1
- [20] Haoqi Luo, Yangliang Li, Junyu Zhang, Hao Zhang, Yunlong Wu, and Qing Ye. Performance analysis of spiral axicon wavefront coding imaging system for laser protection. *Current Optics and Photonics*, 8(4):355–365, 2024. 1
- [21] Jade Caillieaudeaux, Olivier Muller, Morgane Guerchoux, Célia Bruder, Lionel Merlat, Anne-Sophie Schuller, and Christelle Delaite. Thermoset polymers as host for optical limiting. *Journal of Applied Polymer Science*, 141(3): e54810, 2024. 1
- [22] Ling Wang. Self-activating liquid crystal devices for smart laser protection. *Liquid Crystals*, 43(13-15):2062–2078, 2016. 1
- [23] Ruicong Zhang, Zhibo Zhang, Jiecai Han, Lei Yang, Jiajun Li, Zicheng Song, Tianyu Wang, and Jiaqi Zhu. Advanced liquid crystal-based switchable optical devices for light protection applications: principles and strategies. *Light: Science & Applications*, 12(1):11, 2023. 1
- [24] Austin Howes, Zhihua Zhu, David Curie, Jason R Avila, Virginia D Wheeler, Richard F Haglund, and Jason G Valentine. Optical limiting based on huygens’ metasurfaces. *Nano Letters*, 20(6):4638–4644, 2020. 1
- [25] Nicolas Bonod, Pierre Brianceau, Jérôme Daurios, Sylvain Grosjean, Nadja Roquin, Jean-Francois Gleyze, Laurent Lemaître, and Jérôme Neauport. Linear-to-circular polarization conversion with full-silica meta-optics to reduce nonlinear effects in high-energy lasers. *Nature Communications*, 14(1):5383, 2023. 1
- [26] GD Lewis, CN Santos, and M Vandewal. Mitigation of laser dazzle effects on a mid-wave infrared thermal imager by reducing the integration time of the focal plane array. In *Technologies for Optical Countermeasures XVI*, volume 11161, page 1116108. International Society for Optics and Photonics, 2019. 1
- [27] Ric HMA Schleijpen, Sven Binsbergen, Amir Vosteen, Karin de Groot-Trouw, Denise Meuken, and Alexander MJ Van Eijk. Smoke as protection against high energy laser effects. In *Technologies for Optical Countermeasures XVIII and High-Power Lasers: Technology and Systems, Platforms, Effects V*, volume 11867, page 1186704. SPIE, 2021. 1
- [28] Jacob H Wirth, Abbie T Watnik, and Grover A Swartzlander. Half-ring point spread functions. *Optics letters*, 45(8):2179–2182, 2020. 1, 7
- [29] Helen Peng. *Computational Imaging and Its Applications*. Rochester Institute of Technology, 2022. 2
- [30] Xiaopeng Peng, Erin F Fleet, Abbie T Watnik, and Grover A Swartzlander. Learning to see through dazzle. *arXiv preprint arXiv:2402.15919*, 2024. 7, 8

- [31] Junyu Zhang, Qing Ye, Yunlong Wu, Yangliang Li, Yihua Hu, and Haoqi Luo. Learned phase mask to protect camera under laser irradiation. *Optics Express*, 32(24):42674–42691, 2024.
- [32] Haoqi Luo, Junyu Zhang, Ye Liu, Weibing Sun, Yunlong Wu, Qing Ye, and Yihua Hu. Opsecurecam: optically enhanced secure camera via an engineering point spread function. *Optics Express*, 33(10):20880–20893, 2025.
- [33] Johannes Meyer, Michael Henrichsen, Christian Eisele, Bastian Schwarz, Jürgen Limbach, Gunnar Ritt, Stefanie Dengler, Lukas Dippon, and Christian Kludt. Laser protection via jointly learned defocus and image reconstruction. *Authorea Preprints*, 2025. 1
- [34] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *Acm Siggraph 2006 Papers*, pages 795–804. 2006. 2
- [35] Yifan Peng, Qiang Fu, Felix Heide, and Wolfgang Heidrich. The diffractive achromat full spectrum computational imaging with diffractive optics. *ACM Transactions on Graphics (TOG)*, 35(4):1–11, 2016. 2
- [36] Xiong Dun, Hayato Ikoma, Gordon Wetzstein, Zhanshan Wang, Xinbin Cheng, and Yifan Peng. Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging. *Optica*, 7(8):913–922, 2020. 2
- [37] Qilin Sun, Ethan Tseng, Qiang Fu, Wolfgang Heidrich, and Felix Heide. Learning rank-1 diffractive optics for single-shot high dynamic range imaging. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1386–1396, 2020. 2
- [38] Christopher A Metzler, Hayato Ikoma, Yifan Peng, and Gordon Wetzstein. Deep optics for single-shot high-dynamic-range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1375–1385, 2020. 2
- [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 papers*, pages 1–10. 2008. 2
- [40] Mushfiqur Rouf, Rafał Mantiuk, Wolfgang Heidrich, Matthew Trentacoste, and Cheryl Lau. Glare encoding of high dynamic range images. In *CVPR 2011*, pages 289–296. IEEE, 2011. 2
- [41] Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)*, 37(4):1–13, 2018. 2
- [42] Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. Phasecam3d—learning phase masks for passive single view depth estimation. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2019.
- [43] Shiyu Tan, Yicheng Wu, Shou-I Yu, and Ashok Veeraraghavan. Codedstereo: Learned phase masks for large depth-of-field stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7170–7179, 2021. 2
- [44] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Transactions on Graphics (TOG)*, 26(3):69, 2007. 2
- [45] Xiaopeng Peng and Grover A Swartzlander. Shape reconstruction and orientation estimation of transparent microscopic object using light field microscopy. In *Frontiers in Optics*, pages FTh5C–3. Optica Publishing Group, 2016. 2
- [46] Vivek Boominathan, Jacob T Robinson, Laura Waller, and Ashok Veeraraghavan. Recent advances in lensless imaging. *Optica*, 9(1):1–16, 2022. 2
- [47] Carlos Hinojosa, Juan Carlos Niebles, and Henry Arguello. Learning privacy-preserving optics for human pose estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2573–2582, 2021. 2
- [48] Zaid Tasneem, Giovanni Milione, Yi-Hsuan Tsai, Xiang Yu, Ashok Veeraraghavan, Manmohan Chandraker, and Francesco Pittaluga. Learning phase mask for privacy-preserving passive depth estimation. In *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pages 504–521. Springer, 2022. 2
- [49] Xiaopeng Peng, Garreth J Ruane, and Grover A Swartzlander Jr. Randomized aperture imaging. *arXiv preprint arXiv:1601.00033*, 2016. 2
- [50] Xiaopeng Peng and Grover A Swartzlander. Mirror swarm space telescope. In *2014 IEEE Western New York Image and Signal Processing Workshop (WNYISPW)*, pages 37–41. IEEE, 2014. 2
- [51] Daniel S Jeon, Seung-Hwan Baek, Shinyoung Yi, Qiang Fu, Xiong Dun, Wolfgang Heidrich, and Min H Kim. Compact snapshot hyperspectral imaging with diffracted rotation. *ACM Transactions on Graphics (TOG)*, 38(4):1–13, 2019. 2
- [52] Seung-Hwan Baek, Hayato Ikoma, Daniel S Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H Kim. Single-shot hyperspectral-depth imaging with learned diffractive optics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2651–2660, 2021. 2
- [53] Xi Mou, Xiaopeng Peng, and Jianping Wang. P-48: Evaluation of field of view in optical see-through near eye displays. In *SID Symposium Digest of Technical Papers*, volume 55, pages 1548–1550. Wiley Online Library, 2024. 2
- [54] Xi Mou, Xiaopeng Peng, and Tongsheng Mou. 38-2: Invited paper: Evaluating optical performance and image quality in augmented reality eyewear: Standardization, challenges, and measurement methods. In *SID Symposium Digest of Technical Papers*, volume 55, pages 324–326. Wiley Online Library, 2024.
- [55] Baizhou Yang, Ling Chen, Xiaopeng Peng, Jiashen Chen, Yani Tang, Wei Wang, Dingyi Fang, and Chao Feng. Rf-sauron: Enabling contact-free interaction on eyeglass using conformal rfid tag. *IEEE Internet of Things Journal*, 2025. 2

- [56] Qilin Sun, Jian Zhang, Xiong Dun, Bernard Ghanem, Yifan Peng, and Wolfgang Heidrich. End-to-end learned, optically coded super-resolution spad camera. *ACM Transactions on Graphics (TOG)*, 39(2):1–14, 2020. 2
- [57] Zheng Shi, Yuval Bahat, Seung-Hwan Baek, Qiang Fu, Hadi Amata, Xiao Li, Praneeth Chakravarthula, Wolfgang Heidrich, and Felix Heide. Seeing through obstructions with diffractive cloaking. *ACM Transactions on Graphics (TOG)*, 41(4):1–15, 2022. 3
- [58] Ethan Tseng, Ali Mosleh, Fahim Mannan, Karl St-Arnaud, Avinash Sharma, Yifan Peng, Alexander Braun, Derek Nowrouzezahrai, Jean-Francois Lalonde, and Felix Heide. Differentiable compound optics and processing pipeline optimization for end-to-end camera design. *ACM Transactions on Graphics (TOG)*, 40(2):1–19, 2021. 3
- [59] Yuqi Li, Qiang Fu, and Wolfgang Heidrich. Extended depth-of-field projector using learned diffractive optics. In *2023 IEEE conference virtual reality and 3D user interfaces (VR)*, pages 449–459. IEEE, 2023. 2
- [60] Kaihao Zhang, Wenqi Ren, Wenhan Luo, Wei-Sheng Lai, Björn Stenger, Ming-Hsuan Yang, and Hongdong Li. Deep image deblurring: A survey. *International Journal of Computer Vision*, pages 1–28, 2022. 2
- [61] Xiaopeng Peng, Prateek R Srivastava, and Grover A Swartzlander. Cnn-based real-time image restoration in laser suppression imaging. In *Optical Sensors and Sensing Congress*, pages JTh6A–10. Optica Publishing Group, 2021. 2
- [62] Xiaopeng Peng, Garreth J Ruane, Marco B Quadrelli, and Grover A Swartzlander. Randomized apertures: high resolution imaging in far field. *Optics express*, 25(15):18296–18313, 2017.
- [63] Xiaopeng Peng, Garreth J Ruane, Alexandra B Artusio-Glimpse, and Grover A Swartzlander Jr. Image restoration from a sequence of random masks. In *Computational Imaging XIII*, volume 9401, pages 111–123. SPIE, 2015. 2
- [64] Jingyun Liang, Jie Zhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 2
- [65] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17683–17693, 2022.
- [66] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 2
- [67] Hanxiao Liu, Zihang Dai, David So, and Quoc V Le. Pay attention to mlps. *Advances in Neural Information Processing Systems*, 34:9204–9215, 2021. 2
- [68] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5769–5780, 2022. 2
- [69] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023. 2
- [70] Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. In *Forty-first International Conference on Machine Learning*. 2
- [71] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024.
- [72] Pengfei Zhou, Jie Xia, Xiaopeng Peng, Wangbo Zhao, Zilong Ye, Zekai Li, Suorong Yang, Jiadong Pan, Yuanxiang Chen, Ziqiao Wang, et al. Neural-driven image editing. *arXiv preprint arXiv:2507.05397*, 2025. 2
- [73] Oren Rippel, Jasper Snoek, and Ryan P Adams. Spectral representations for convolutional neural networks. *Advances in neural information processing systems*, 28, 2015. 2
- [74] Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. *Advances in Neural Information Processing Systems*, 33:4479–4488, 2020.
- [75] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547, 2020. 2
- [76] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5978–5986, 2019. 2
- [77] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 2
- [78] Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Chen Change Loy, and Jinwei Gu. Removing diffraction image artifacts in under-display camera via dynamic skip connection network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 662–671, 2021. 2
- [79] Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. *Advances in neural information processing systems*, 31, 2018. 2
- [80] Chieh Hubert Lin, Chia-Che Chang, Yu-Sheng Chen, Da-Cheng Juan, Wei Wei, and Hwann-Tzong Chen. Cogan: Generation by parts via conditional coordinating. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4512–4521, 2019.
- [81] Chieh Hubert Lin, Hsin-Ying Lee, Yen-Chi Cheng, Sergey Tulyakov, and Ming-Hsuan Yang. Infinitygan: Towards

- infinite-pixel image synthesis. In *International Conference on Learning Representations*, 2021. 2
- [82] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *stat*, 1050:1, 2014. 2
- [83] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, 31, 2018. 2
- [84] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015. 2
- [85] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2
- [86] Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion gans. In *International Conference on Learning Representations*, 2021. 2
- [87] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 2
- [88] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021. 2
- [89] Hongkai Zheng, Weili Nie, Arash Vahdat, Kamyar Azizzadenesheli, and Anima Anandkumar. Fast sampling of diffusion models via operator learning. In *International Conference on Machine Learning*, pages 42390–42402. PMLR, 2023. 2
- [90] Chenlin Meng, Robin Rombach, Ruiqi Gao, Diederik Kingma, Stefano Ermon, Jonathan Ho, and Tim Salimans. On distillation of guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14297–14306, 2023. 2
- [91] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 745–755, 2022. 3
- [92] <https://github.com/rafael-fuente/diffractsim>. 4
- [93] Matthias Hullin, Elmar Eisemann, Hans-Peter Seidel, and Sungkil Lee. Physically-based real-time lens flare rendering. In *ACM SIGGRAPH 2011 papers*, pages 1–10. 2011. 4
- [94] Yicheng Wu, Qirui He, Tianfan Xue, Rahul Garg, Jiawen Chen, Ashok Veeraraghavan, and Jonathan T Barron. How to train neural networks for flare removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2239–2247, 2021.
- [95] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7k: A phenomenological night-time flare removal dataset. *Advances in Neural Information Processing Systems*, 35:3926–3937, 2022. 4
- [96] Min Hu and Jieqing Tan. Adaptive osculatory rational interpolation for image processing. *Journal of Computational and Applied Mathematics*, 195(1-2):46–53, 2006. 6, 7
- [97] Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in neural information processing systems*, 33:5824–5836, 2020. 6
- [98] Pengfei Zhou, Xiaopeng Peng, Jiajun Song, Chuanhao Li, Zhaopan Xu, Yue Yang, Ziyao Guo, Hao Zhang, Yuqi Lin, Yefei He, et al. Opening: A comprehensive benchmark for judging open-ended interleaved image-text generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 56–66, 2025. 7
- [99] Pengfei Zhou, Xiaopeng Peng, Fanrui Zhang, Zhaopan Xu, Jiaxin Ai, Yansheng Qiu, Chuanhao Li, Zhen Li, Ming Li, Yukang Feng, et al. Mdk12-bench: A comprehensive evaluation of multimodal large language models on multidisciplinary exams. *arXiv preprint arXiv:2508.06851*, 2025. 7
- [100] <https://unsplash.com/license>. 7