Guided Lensless Polarization Imaging

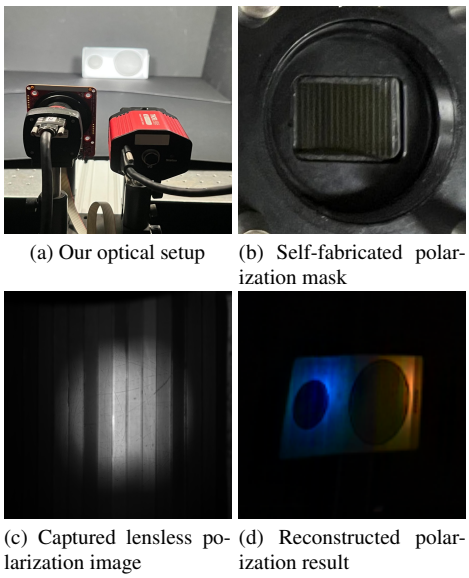
Noa Kraicer Erez Yosef Raja Giryes
Tel Aviv University
School of Electrical Engineering, Faculty of Engineering
noakraicer0@gmail.com erez.yo@gmail.com raja@tauex.tau.ac.il

Abstract

Polarization imaging captures the polarization state of light, revealing information invisible to the human eye yet valuable in domains such as biomedical diagnostics, autonomous driving, and remote sensing. However, conventional polarization cameras are often expensive, bulky, or both, limiting their practical use. Lensless imaging offers a compact, low-cost alternative by replacing the lens with a simple optical element like a diffuser and performing computational reconstruction, but existing lensless polarization systems suffer from limited reconstruction quality. To overcome these limitations, we introduce an RGB-guided lensless polarization imaging system that combines a compact polarization-RGB sensor with an auxiliary, widely available conventional RGB camera providing structural guidance. We reconstruct multi-angle polarization images for each RGB color channel through a two-stage pipeline: a physics-based inversion recovers an initial polarization image, followed by a Transformer-based fusion network that refines this reconstruction using the RGB guidance image from the conventional RGB camera. Our two-stage method significantly improves reconstruction quality and fidelity over lensless-only baselines, generalizes across datasets and imaging conditions, and achieves high-quality real-world results on our physical prototype lensless camera without any fine-tuning.

1. Introduction

Polarization is a fundamental property of light that describes the orientation and phase relationship between its orthogonal electric-field components. Because reflections and material anisotropy affect the polarization state, it encodes information about surface geometry, reflectance, and composition, details often inaccessible to standard intensity or RGB imaging [21, 46]. Recent approaches to polarization imaging include division-of-focal-plane (DoFP) sensors, which assign different polarization filters to adjacent pixels; time-sequential polarimetry, which uses rotating or tunable elements to capture multiple polarization states over time [46], and lensless polarization imaging, which replaces conventional lenses with spatially coded optical elements and reconstructs the image computationally. Despite its utility, polarization imaging has yet to achieve widespread adoption, largely due to the cost, size, and complexity of conventional polarization cameras.



(a) Our optical setup (b) Self-fabricated polarization mask

(c) Captured lensless polarization image (d) Reconstructed polarization result

Figure 1. **RGB-guided lensless polarization imaging system:** (a) optical setup; (b) custom polarization mask; (c) captured lensless image under front illumination with two orthogonally polarized projectors; and (d) reconstructed grayscale polarization result, visualized by mapping the $0°$, $45°$, and $90°$ outputs to the R, G, and B channels.

Lensless imaging offers a compelling alternative by replacing lenses with simple optical elements, such as diffusers or coded masks, and shifting hardware complexity to computation. Such schemes enable compact, low-cost, and scalable imaging systems [11]. Recent works have extended lensless imaging to polarization by combining polarization-sensitive components with various optical coding schemes [5, 16, 25, 47]. Yet, reconstruction quality remains limited

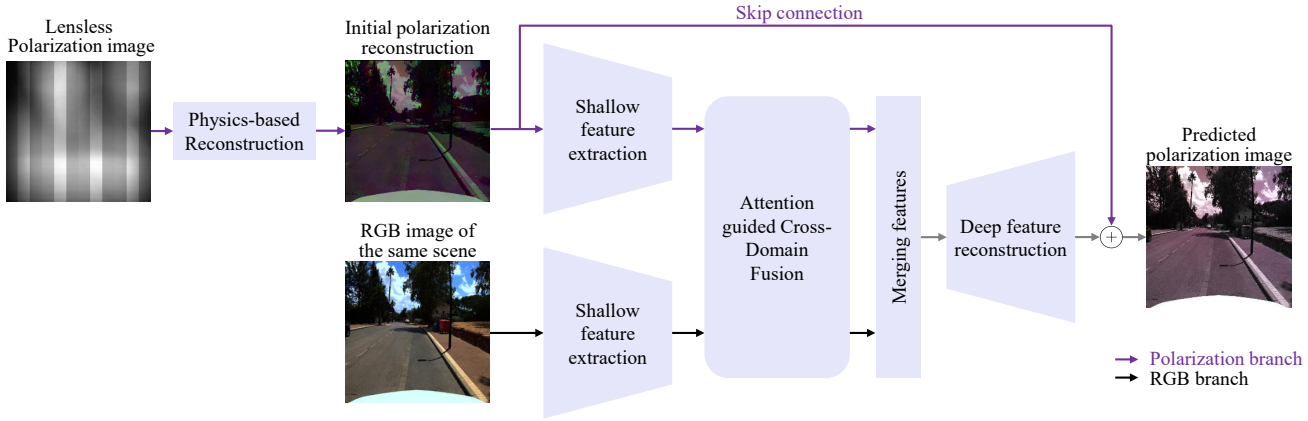


Figure 2. Overview of the proposed RGB-guided reconstruction pipeline. The process consists of two stages: (1) polarization intensity images (RGB or grayscale) are reconstructed from lensless measurements using a physics-based algorithm (FISTA/ADMM); and (2) the initial reconstruction and an RGB image of the same scene are separately encoded and fused through cross-domain attention to produce a refined polarization reconstruction. For visualization, the grayscale reconstructions at three polarization angles ($0°$, $45°$, $90°$) are mapped to the R, G, and B channels. The pipeline is compatible with more general input configurations.

due to the highly compressed measurements that jointly encode structural and polarization information, compounded by the inherently ill-posed nature of the inverse problem.

Classical model-based solvers, such as the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [8] and the Alternating Direction Method of Multipliers (ADMM) algorithm [13], can recover coarse polarization estimates but often fail to recover high-frequency details and are sensitive to noise and deviations from the assumed imaging model. While deep learning models have significantly improved reconstruction quality in general lensless imaging tasks [11], to the best of our knowledge, they have not yet been explored for lensless polarization imaging.

RGB images provide complementary geometric and edge information consistent across polarization states, offering a natural way to regularize the ill-posed inversion and recover fine details. While adding an RGB camera increases system complexity compared to a fully lensless design, RGB cameras are compact, low-cost, and widely available, keeping the system simpler and more affordable than dedicated polarization cameras. This is particularly important for size- and cost-constrained applications such as compact microscopy and endoscopy [38, 39, 42, 52].

Leveraging the complementary geometric information provided by RGB images, we introduce the **first RGB-guided lensless polarization imaging system** that integrates physics-based reconstruction with data-driven refinement (Figure 2), enabling accurate recovery of fine structures and details. The first stage performs a model-based inversion (e.g., FISTA or ADMM) to obtain a physically consistent initialization from raw lensless measurements. The second stage refines this estimate using a Swin Transformer-based fusion network, based on SwinFuSR [3], which fuses polarization features from the initial reconstruction with RGB features from a lensed camera via alternating self- and cross-attention. Our approach generalizes across diverse scenes and imaging conditions, and achieves strong performance on real-world measurements without additional fine-tuning.

Our main contributions may be summarized as follows: (i) We propose the first **RGB-guided** lensless polarization imaging system, combining simple and low-cost hardware with a reconstruction algorithm achieving state-of-the-art results for lensless polarization imaging; (ii) We design a two-stage reconstruction approach that integrates a physics-based solver (e.g., FISTA/ADMM) with an adapted version of Swin Transformer utilized for cross-modal fusion, enabling RGB-guided reconstruction of polarization intensity images through self- and cross-attention; (iii) We conduct extensive experiments on multiple simulated datasets, demonstrating consistent improvements over lensless-only baselines in PSNR, SSIM, and LPIPS, with strong generalization across datasets and unseen point-spread-functions (PSFs); (iv) We demonstrate promising real-world results on a prototype lensless polarization camera, validating the method's practical feasibility without additional fine-tuning. **Project page:** https://noa-kraicer.github.io/Guided-Lensless-Polarization-Imaging

2. Related Work

Polarization Imaging. Most existing polarization cameras are based on sequential filtering or division-of-focal-plane (DoFP) architectures [46]. DoFP sensors enable single-shot capture but suffer spatial resolution loss due to pixel subdivision. Sequential filtering preserves full resolution via multiple exposures with rotating or switching polarizers but introduces motion artifacts and mechanical complexity, lim-

iting real-time use. Recent designs address these trade-offs using stacked polarizer architectures [41] and flat-optics or metasurface-based polarization elements [20, 27, 65], signaling a shift toward miniaturized polarization cameras and motivating exploration of lensless variants.

Lensless Imaging. Lensless imaging replaces traditional lenses with coded optics and computational reconstruction, enabling compact imaging systems. Several designs have been proposed, including FlatCam [4], which uses a static amplitude mask; DiffuserCam [1], based on a diffuser and compressive sensing; phase masks [10]; and programmable optics [18, 19, 35, 62, 64]. A comprehensive review of lensless imaging systems can be found in Boominathan et al. [12]. Subsequent research has extended lensless imaging to additional modalities, including hyperspectral [36], depth [6], and temporal imaging [2], highlighting the versatility of the approach. However, lensless polarization imaging has received limited attention and introduces a more complex forward model.

Lensless Polarization Imaging Combining lensless imaging with polarization sensing enables compact, multi-modal imaging systems. Prior works have explored various hardware configurations for single-shot capture, including the use of diffusers with polarization mask [16, 25], phase masks and polarization-encoded apertures [5], and coded masks paired with DoFP polarization sensor [47].

Despite these hardware advances, accurate image reconstruction remains a major challenge. The highly compressed and multiplexed nature of the measurements leads to severely ill-posed inverse problems that are sensitive to noise, artifacts, and other real-world imperfections.

Lensless Imaging Reconstruction. Classical lensless reconstruction methods formulate image recovery as a variational inverse problem solved by iterative optimization algorithms such as FISTA, ADMM, and their variants [1, 10, 36]. For lensless polarization imaging, Elmalem and Giryes [16] use TV-regularized FISTA, while Baek et al. [5] and Wang et al. [47] adopt ADMM-based solvers. These physics-driven methods are interpretable but remain sensitive to noise and PSF mismatch, and cannot fully recover fine details. Deep learning has advanced lensless imaging reconstruction from classical physics-based optimization to learned data-driven models. Early approaches learn direct mappings from sensor measurements to images using convolutional neural networks [43], while later methods integrate model-based priors through unrolled or hybrid architectures such as ISTA-Net [59], ADMM-Net [45]. Hybrid systems like FlatNet [23, 24], DifuzCam [55], and GANESH [34] combine physical inversion with learned refinement for improved fidelity and generalization. Recent modular frameworks further enhance robustness under domain shifts [9] or employing efficient adaptations for new distributions [54].

However, existing approaches for lensless *polarization* imaging remain purely physics-based, relying on explicit forward models, iterative optimization, and handcrafted priors. Although learning-based approaches have shown promise for polarization demosaicing [37, 58], denoising [28], deblurring [63], and low-light enhancement [17, 51], these efforts focus on lens-based systems and have not been demonstrated for lensless polarization reconstruction.

Cross-Modal Guidance. Cross-modal fusion has proven to be effective in enhancing degraded modalities using complementary ones. SwinFuSR [3] uses high-resolution RGB images to guide thermal super-resolution via a Swin Transformer, while Yosef and Giryes [56] presented an RGB image denoising using scene textual description with a diffusion model. In polarization imaging, Liu et al. [29] uses RGB guidance to demosaic simulated sparse DoFP data and recover Stokes parameters. PolarFree [53] shows that polarization can act as a powerful auxiliary signal for reflection removal in RGB images. PolarAnything [60] generates polarization images directly from RGB inputs using diffusion models; however, it relies solely on learned priors and lacks physical fidelity. Yet, RGB guided reconstruction for *lensless polarization* imaging has not been explored. We address this gap with a two-stage framework that leverages RGB guidance to enhance reconstruction fidelity and robustness.

3. Method

Our framework reconstructs high-quality polarization intensity images from a single-shot lensless measurement, leveraging guidance from a registered RGB image. We first formulate the lensless polarization imaging setup (3.1), and describe the synthetic data generation process used for the training (3.2). We then describe our two-stage reconstruction pipeline: a physics-based optimization stage (3.3) and an RGB-guided transformer-based refinement stage (3.4). Figure 2 provides an overview of the full pipeline.

3.1. Lensless Polarization Imaging Setup

We employ a lensless imaging system that captures a spatio-polarimetrically multiplexed measurement in a single shot. Our optical design, adapted from [16, 25], combines a diffuser PSF with a striped polarization mask composed of linear polarizers oriented at $0°$, $45°$, $90°$, and $135°$ on an RGB sensor. Placed directly on the sensor, this mask spatially encodes the polarization state of incident light by transmitting different polarization components through distinct stripe orientations according to Malus' law [14], as illustrated in Figure 3. The diffuser mixes this information in the captured image, but due to its frequency response, it loses the high-frequency information. A detailed description of the optical setup is provided in the Supplementary Material. The resulting measurement can be modeled as

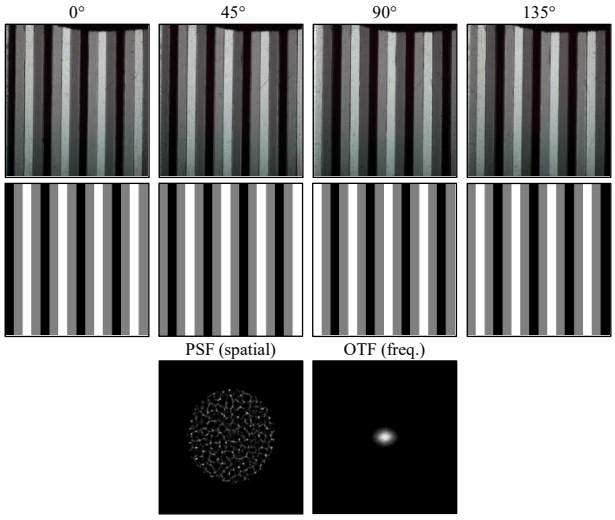


Figure 3. Real and simulated polarization mask responses for four polarization angles in grayscale. (Top) Real masks captured with our prototype; (center) simulated masks; and (bottom) measured grayscale PSF of our diffuser in the spatial domain and its corresponding Optical Transfer Function (OTF) in the frequency domain. The PSF is low-pass, leading to loss of high-frequency information.

a linear inverse problem: $\mathbf{y} = A\mathbf{x} + \mathbf{e}$, where $\mathbf{y}$ denotes the captured lensless measurement, $\mathbf{x}$ is the multi-angle polarization intensity image to be reconstructed, and A is the forward imaging operator combining the effects of the diffuser and polarization mask. The additive term $\mathbf{e}$ accounts for measurement noise. Due to strong spatial multiplexing, high-frequency loss, and partial polarization sampling, this inverse problem is highly ill-posed. To address this, we use an additional conventional RGB camera capturing the same scene, and leverage the RGB image as guidance to provide structural cues and recover high-frequency details.

Forward Model. Our lensless polarization camera follows the same compressive imaging principles as the Spectral DiffuserCam [36]. The polarization-independent diffuser convolves the scene with its PSF, where each narrow feature acts like a micro-lens, mapping a point source to a point on the sensor and spatially multiplexing light from all scene points (Figure 3). This multiplexing enables reconstruction from a subset of sensor pixels, allowing the polarization mask to perform partial sampling across polarization angles. The mask transmits light according to its orientation, applying a multiplicative modulation to the incident intensity. While the diffuser's PSF enables recovery with the mask, it loses some high-frequency information. Let $\mathbf{x} \in \mathbb{R}^{H\times W\times C\times P}$ denote the intensity of the scene per angle of polarization, where H and W are the spatial dimensions, $C \in \{1, 3\}$ is the number of color channels (grayscale or RGB), and $P \in \{3, 4\}$ corresponds to angles $0°, 45°, 90°$ (and optionally $135°$). Each $\mathbf{x}_{:,:,c,p}$ is the intensity that would be observed after an ideal polarizer at angle p. Accordingly, the polarization mask is modeled as a binary spatial selector $\mathbf{S}_p$ that assigns sensor regions to each orientation, since the angular dependence is already encoded in $\mathbf{x}_{:,:,c,p}$. For each color channel c, the measurement is:

$$\mathbf{y}_{:,:,c} = \sum_{p=1}^{P} \mathbf{S}_p \odot \left(\mathbf{x}_{:,:,c,p} * \mathbf{k}_c\right) \tag{1}$$

where $\mathbf{k}_c$ is the diffuser PSF for channel c, $\odot$ denotes element-wise multiplication, and $*$ denotes 2D convolution over spatial dimensions.

3.2. Synthetic Data Generation

Training deep neural networks requires a large number of labeled samples, which is impractical to collect from a real lensless polarization camera at scale. Moreover, to the best of our knowledge, there are no publicly available datasets specifically designed for lensless polarization reconstruction. To address this, we simulate lensless measurements from existing datasets: Polarimetric Imaging for Perception (PIP) [7], UPLight [30], and ZJU-RGB-P [50]. Each dataset provides pixel-level aligned RGB-polarimetric data with four polarization orientations per scene, from which we generate synthetic lensless measurements using the forward model in Equation (1).

An equivalent unpolarized RGB image is computed as:

$$\mathbf{f}_{\text{RGB}} = \frac{1}{2}(\mathbf{I}_{0°} + \mathbf{I}_{45°} + \mathbf{I}_{90°} + \mathbf{I}_{135°}), \tag{2}$$

representing the total unpolarized intensity [29]. For simulation, we adopt two configurations:

1. a three-angle grayscale setup capturing dominant polarization behavior [49].
2. a four-angle RGB configuration matching our hardware and improving robustness and noise reduction [26].

The simulated polarization mask reproduces the prototype's periodic structure while omitting fabrication artifacts (e.g., dust, edge roughness), enabling consistent data generation independent of a specific physical mask's imperfections. Each mask pattern consists of four repeated vertical stripe sequences (manufacturing-quality tradeoff). It is applied identically across color channels to maintain consistency with the hardware design. For realism, we use the PSF measured from our physical device for the simulation process. Figure 3 illustrates the four-angle configuration, shown in grayscale for visualization, along with the corresponding measured PSF and hardware mask for direct comparison.

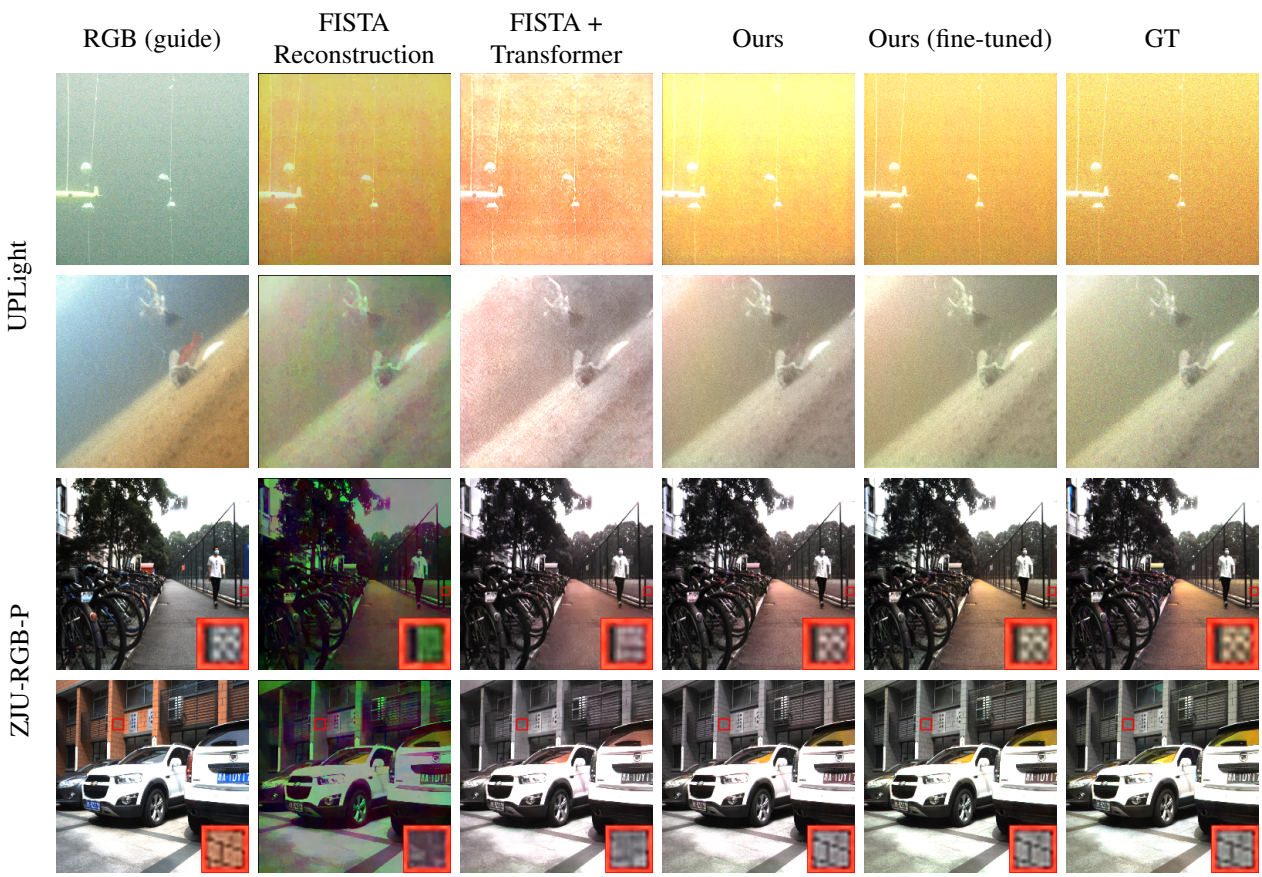


Figure 4. Qualitative reconstruction results on **UPLight** and **ZJU-RGB-P**. Columns: RGB guidance, FISTA reconstruction, FISTA + Transformer w/o RGB, our full RGB-guided model, its fine-tuned version, and the ground-truth polarization image. Each polarization grayscale triplet (0°, 45°, 90°) is visualized as an RGB composite. Note how the RGB guidance improves the high-frequency recovery.

Table 1. Quantitative results (PSNR ↑ / SSIM ↑ / LPIPS ↓) for both **four-angle RGB** and **three-angle grayscale** configurations.

Modality	**Model**	**PIP**			**UPLight**			**ZJU-RGB-P**		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Color	FISTA	14.76	0.47	0.44	12.20	0.18	0.60	15.27	0.48	0.45
	ADMM	12.92	0.30	0.57	10.93	0.14	0.65	14.69	0.34	0.53
	FISTA + Transf.	28.31	0.86	0.12	15.98	0.37	0.36	25.65	0.86	0.15
	ADMM + Transf.	25.41	0.81	0.17	14.63	0.35	0.41	22.38	0.79	0.22
	Ours (ADMM input)	32.48	**0.95**	**0.04**	19.63	**0.51**	0.29	29.11	**0.96**	**0.04**
	Ours (FISTA input)	**33.05**	**0.95**	**0.04**	**20.06**	**0.51**	**0.28**	**30.38**	**0.96**	**0.04**
Grayscale	FISTA	13.87	0.45	0.45	16.72	0.26	0.53	14.50	0.46	0.44
	ADMM	13.06	0.31	0.63	11.98	0.18	0.71	14.90	0.36	0.59
	FISTA + Transf.	28.85	0.88	0.12	17.93	0.44	0.53	27.20	0.89	0.19
	ADMM + Transf.	24.87	0.81	0.20	15.37	0.40	0.76	23.32	0.80	0.29
	Ours (ADMM input)	34.40	**0.97**	**0.03**	18.45	**0.53**	0.46	29.44	**0.97**	0.09
	Ours (FISTA input)	**35.13**	**0.97**	**0.03**	**20.49**	0.52	**0.32**	**31.19**	**0.97**	**0.07**

3.3. Stage I: Physics-based reconstruction

Recovering the polarization intensity image $\hat{\mathbf{x}}$ is achieved by solving the optimization problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2\sigma_e^2} \|\mathbf{y} - A\mathbf{x}\|_2^2 + s(\mathbf{x}), \tag{3}$$

where A denotes the forward operator (Equation (1)), and σ_e is the standard deviation of measurement noise. The first term enforces fidelity to the sensor measurement, and $s(\mathbf{x})$ is a regularization term promoting desirable image priors. The unknown polarization image is represented as $\mathbf{x} \in \mathbb{R}^{H\times W\times C\times P}$, while the observed measurement

$\mathbf{y} \in \mathbb{R}^{H \times W \times C}$ corresponds to either a real sensor acquisition or a simulated observation generated as described in Section 3.2. We solve this inverse problem using FISTA [8] with a fixed number of iterations, employing a weighted 3D Total Variation (3DTV) prior [22] to enforce smoothness across spatial and polarization dimensions, along with a non-negativity constraint. For simulated data, we use the same PSF and polarization mask as in the simulation, while for real measurements, we use the device-measured PSF and mask shown in Figure 3. We also implement an alternative solver based on ADMM [13] using the same prior, demonstrating the generality of our formulation. Additional implementation details, including iteration settings, are provided in the supplementary.

3.4. Stage II: RGB-Guided Deep Refinement

While stage I provides a coarse polarization estimate $\hat{\mathbf{x}}$, it lacks fine spatial details due to the loss of high frequencies (see Figure 3). To recover the high-frequency details and improve reconstruction quality, we employ an RGB-guided refinement network based on SwinFuSR [3], a dual-branch Transformer originally designed for RGB-guided thermal super-resolution.

We adapt SwinFuSR to our task by (i) modifying input/output channels for our polarization data; (ii) training on full-resolution images instead of patches to correct global, spatially correlated artifacts from FISTA/ADMM reconstruction; and (iii) incorporating an LPIPS perceptual loss [61] to enhance perceptual quality, consistently with prior SR and reconstruction studies [15, 32, 57]. To improve robustness to small registration errors in real-world measurements, we apply random translation augmentation during training, synthetically shifting both the RGB image and the ground-truth polarization image jointly by up to ± 4 pixels in both horizontal and vertical directions.

The network processes $\hat{\mathbf{x}}$ and the approximately aligned (with residual shifts of up to ± 4 pixels) RGB image $\mathbf{f}_{\mathrm{RGB}}$ through separate branches of shallow convolutional layers and Swin Transformer layers (STL). Feature fusion is later performed using Attention-guided Cross-domain Fusion (ACF) blocks, which alternate between self-attention and cross-attention to integrate information from both modalities. The outputs of the two branches are then merged via concatenation followed by convolution. The fused features are subsequently refined through additional STLs and convolutional layers, where a skip connection adds the initial reconstruction to the output to preserve consistency with our physical forward model. Further implementation details appear in the supplementary.

4. Experimental Results

We evaluate our proposed method on both synthetic and real-world data. For quantitative evaluation, we used PSNR, SSIM [48] and LPIPS [61]. Our RGB-guided method outperforms existing state-of-the-art methods and an equivalent non-guided approach across all tested datasets and metrics. We demonstrate the advantages of our method on both synthetic datasets and real-world data using our prototype camera. We also present ablation studies to analyze the impact of PSF variations, initialization steps, fusion strategy, translation augmentation, and RGB-only input.

Implementation Details. We train our model on the PIP dataset [7], which is the largest RGB-polarization dataset ($\sim$12.6K samples). We split the data into training, validation, and test sets with no scene overlap. Training is conducted under the two configurations as earlier described in the simulation process (Section 3.2) with RGB guidance images. We use AdamW [31] with a OneCycle schedule [44] (peak 1.5×10^{-4}) for 30 epochs. The loss combines an ℓ_1 term and LPIPS [61], with weights of 1.0 and 0.1, respectively, with early stopping on validation loss. Inputs for training and evaluation are resized to 250×250 pixels and run on an NVIDIA RTX 2080 Ti. Translation augmentation is used only during training of the RGB-guided models. All reported results are obtained without augmentation.

4.1. Synthetic Data Results

We compare our approach against two physics-based reconstruction baselines: 3D FISTA with total variation (TV) regularization [36], and 3D ADMM variant with similar regularization. Both are widely used in lensless polarization reconstruction. In addition, we evaluate a learning-based baseline derived from our architecture, denoted *FISTA/ADMM + Transformer*, in which the refinement network operates solely on the initial reconstruction, without RGB guidance. This is implemented by feeding the same input to both branches and removing cross-modal fusion, thereby disentangling the contribution of RGB guidance. We also compare against two further learning-based baselines: FlatNet [24], which employs a learnable inversion followed by a perceptual refinement U-Net, and PolarAnything [60] that synthesizes polarization images from RGB inputs based on Stable Diffusion v1.5 [40]. We adapt PolarAnything to predict polarization *intensity* images instead of the original angle and degree of linear polarization (AoLP/DoLP), testing two conditioning modes—an RGB image and the FISTA reconstruction. Both FlatNet and PolarAnything are evaluated under the **three-angle grayscale** configuration, since their architectures do not natively support multi-channel RGB-polarization inputs, and are retrained on our dataset for fairness. Table 2 shows that neither of them performs well on the PIP test set. FlatNet's learned deconvolution struggles under partial polarization sampling, lacking the iterative regularization and measurement consistency of physics-based solvers. PolarAnything, despite its strong generative prior, performs poorly when

Table 2. Comparison of FlatNet, PolarAnything, FISTA + Transformer, and our method under the three-angle grayscale configuration on the PIP dataset.

Model	PSNR↑	SSIM↑	LPIPS↓
FlatNet	21.57	0.68	0.45
PolarAnything (FISTA input)	21.51	0.64	0.31
PolarAnything (RGB input)	22.02	0.66	0.29
FISTA + Transf.	28.85	0.88	0.12
Ours (FISTA input)	**35.13**	**0.97**	**0.03**

adapted to polarization intensity prediction and conditioned on either RGB or FISTA inputs, failing to match the accuracy of our method with or without RGB guidance. They both fail to generalize to the unseen datasets (see Tab. 1, supplementary).

The remaining methods (excluding FlatNet and PolarAnything) are evaluated on two additional datasets: ZJU-RGB-P [50] (394 samples) and UPLight [30] (1,991 samples). Both datasets introduce significant domain shift relative to the PIP dataset, particularly the underwater scenes in UPLight. Quantitative results are presented in Table 1, and qualitative examples appear in Figure 4 (grayscale configuration). Our method consistently outperforms both physics-based and learning-based baselines across datasets, recovering richer high-frequency details and improving structural fidelity.

While the base model generalizes well, we further explore whether performance can be improved through fine-tuning on a small number of samples from the target domain. Specifically, for each of the ZJU-RGB-P and UPLight datasets, we fine-tune the PIP-trained model using 10 randomly selected pairs of RGB-polarization images for 5 epochs using the same training setup. The qualitative results (Figure 4) show that an additional improvement in performance can be achieved using a small number of examples from the target domain. This makes domain adaptation more practical, without needing to collect a large dataset. The quantitative results are provided in the supplementary.

4.2. Real-world Results

We validate our method on data captured using the prototype lensless polarization camera, as shown qualitatively in Figure 5. We tested two polarized lighting setups: (i) front-illuminated scenes with two orthogonally polarized projectors (rows 1-3) to assess source separation, and (ii) a back-illuminated scene with a polarized screen and transparent plastic bag (row 4), highlighting polarization-dependent variations such as internal stress. For the grayscale configuration, the polarization images are first reconstructed using FISTA with three angles provided as input to the refinement network. Despite the domain gap and differences in the polarization mask used for training, our RGB-guided method recovers sharper structures and finer details compared to FISTA. The FISTA+Transformer baseline fails to generalize in this setting, further highlighting the robustness and effectiveness of our proposed approach under real-world conditions. The prototype introduces additional challenges, including microfabrication artifacts in the polarization mask, the distance between the sensor and the polarization mask (due to the sensor's cover glass), and various fabrication and assembly inaccuracies that are not fully calibrated and compensated [25]. These issues can lead to deviations from the assumed forward model, potentially degrading the performance of physics-based methods.

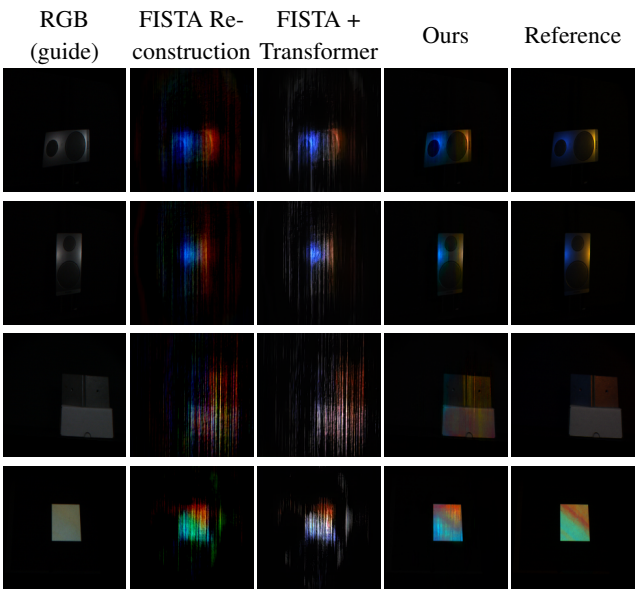


Figure 5. Qualitative results on real lensless polarization data (3-angle grayscale). Each reconstructed polarization triplet ($0°$, $45°$, $90°$) is visualized as an RGB composite. Note the significant improvement in the structural details achieved by RGB guidance.

For real-world lensless measurements, pixel-wise aligned ground-truth polarization images are not available due to hardware differences between the reference and lensless cameras. The reference images in Figure 5, captured using a conventional RGB camera with a rotating polarizer, serve as qualitative benchmarks. Unlike synthetic data, our real measurements require alignment. We perform a one-time homography-based alignment using manually selected correspondences, justified by the short inter-camera baseline relative to scene depth (Fig. 1, supplementary), assuming synchronized acquisition. However, residual misalignment may persist. To handle this, we employ translation augmentation during training (see Section 3.4). Additional qualitative results for the 4-angle RGB configuration are provided in the Supplementary Material (Fig. 3).

Table 3. Ablation study results (PSNR ↑ / SSIM ↑ / LPIPS ↓). Additional details appear in the supplementary.

Dataset \ Model	**FISTA**			**FISTA + Transf.**			**Ours (FISTA input)**		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
PIP (simple fusion)	13.87	0.45	0.45	-	-	-	34.69	0.97	0.03
UPLight (simple fusion)	16.72	0.26	0.53	-	-	-	21.86	0.54	0.30
ZJU-RGB-P (simple fusion)	14.50	0.46	0.44	-	-	-	30.80	0.97	0.08
UPLight (PSF #1)	15.01	0.27	0.62	17.96	0.34	0.63	20.78	0.53	0.32
UPLight (PSF #2)	16.03	0.26	0.54	19.09	0.44	0.53	20.68	0.52	0.32
ZJU-RGB-P (PSF #1)	14.45	0.45	0.47	24.28	0.83	0.27	31.01	0.97	0.08
ZJU-RGB-P (PSF #2)	14.59	0.47	0.42	27.17	0.89	0.19	31.18	0.97	0.07
PIP (5k iters eval, 10k model)	13.09	0.43	0.48	23.61	0.83	0.17	34.83	0.97	0.03
PIP (5k iters eval, 5k model)	13.09	0.43	0.48	27.53	0.87	0.14	34.93	0.97	0.03
UPLight (5k iters eval, 10k model)	14.98	0.25	0.60	18.71	0.42	0.58	20.93	0.53	0.33
UPLight (5k iters eval, 5k model)	14.98	0.25	0.60	16.32	0.39	0.56	21.17	0.53	0.32
UPLight (1k iters)	11.88	0.20	0.72	14.99	0.36	0.71	20.91	0.53	0.43
ZJU-RGB-P (5k iters eval, 10k model)	14.45	0.46	0.45	25.77	0.87	0.21	31.12	0.97	0.08
ZJU-RGB-P (5k iters eval, 5k model)	14.45	0.46	0.45	26.01	0.87	0.21	30.92	0.97	0.08
ZJU-RGB-P (1k iters)	14.93	0.37	0.58	19.73	0.68	0.39	30.19	0.97	0.09

4.3. Ablation Studies

We analyze the impact of RGB fusion strategy, PSF mismatch, and FISTA depth under the three-angle grayscale setup on PIP (Table 3). Additional ablations (RGB-only and translation augmentation) are shown in the supplementary.

To assess the benefit of cross-attention for fusion versus a simpler alternative while retaining the advantages of the attention-based design, we evaluate a **simple fusion** variant. This variant replaces the cross-attention fusion with direct channel concatenation of the two modalities' features, followed by a 3×3 convolution and standard Swin Transformer blocks (RSTB). Unlike our regular fusion module, which employs cross-attention (CRSTB) layers that jointly process both modalities through mutual attention before concatenation, this variant performs early feature-level merging without any adaptive interaction between modalities. Although quantitative differences are modest, the variant exhibits noticeable visual intensity artifacts on both simulated and real data (see qualitative results in Fig. 4, supplementary), similar to those reported in [33].

Next, we evaluate robustness to **PSF mismatch** by training on data simulated and reconstructed with our PSF, and testing on data simulated and reconstructed with two alternative PSFs. These PSFs are measured from different diffusers: the first, from Antipa et al. [1], exhibits a large rectangular pattern, while the second, from Monakhova et al. [36], covers a smaller rectangular area in comparison to our circular shape (see supplementary). Table 3 shows that our model consistently outperforms both FISTA and FISTA+Transformer under these mismatched conditions, demonstrating strong generalization to unseen optics.

We analyze sensitivity to the **number of FISTA iterations** used for the initial reconstruction at test time. Classical FISTA degrades rapidly when iterations are reduced (1k or 5k vs. the 10k baseline used for both training and testing), reflecting its dependence on full convergence. In contrast, our method, leveraging RGB guidance, maintains high fidelity even with weak initializations, demonstrating robustness under limited test-time computation. When trained with fewer iterations (5k instead of 10k), the method remains relatively stable under weaker initialization (Table 3). However, performance still drops compared to models trained and evaluated with fully converged (10k) inputs (Table 1). This highlights the importance of high-quality physics-based initialization and the trade-off between reconstruction accuracy and computational efficiency.

Finally, we evaluate the effect of **translation augmentation**. A model trained without it is sensitive to spatial shifts, whereas incorporating translations largely mitigates this effect, which is critical in real-world settings where perfect alignment cannot be guaranteed (Tab. 5, supplementary).

5. Conclusions

We introduced a modular two-stage framework for RGB-guided reconstruction in lensless polarization imaging, demonstrating high reconstruction quality, strong generalization, and robustness across both simulated and real-world setups. Our framework outperforms all compared baselines, confirming the contribution of our RGB-guided reconstruction approach. Beyond lensless setups, our approach can be extended to conventional polarization cameras with limited resolution or high noise, where RGB guidance can enhance the recovery of fine details. This modular design supports integration with alternative inverse solvers and adapts to a wide range of imaging setups. Future work will focus on improving computational efficiency and extending the framework toward dynamic scenes, advancing compact polarization imaging in practical environments.

Acknowledgment

We thank Tomer Pee'r and Michael Baltaxe (General Motors) for providing a suitable version of the PIP dataset, and Shay Elmalem for fruitful discussions. This work was partially supported by the Center for AI and Data Science at Tel Aviv University (TAD) and by ERC Grant No. 10111339.

References

[1] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. Diffusercam: lensless single-exposure 3d imaging. *Optica*, 5(1):1–9, 2018. 3, 8

[2] Nick Antipa, Patrick Oare, Emrah Bostan, Ren Ng, and Laura Waller. Video from stills: Lensless imaging with rolling shutter. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2019. 3

[3] Cyprien Arnold, Philippe Jouvet, and Lama Seoud. Swinfusr: an image fusion-inspired model for RGB-guided thermal image super-resolution. In *CVPR*, pages 3027–3036, 2024. 2, 3, 6

[4] M Salman Asif, Ali Ayremlou, Aswin Sankaranarayanan, Ashok Veeraraghavan, and Richard G Baraniuk. Flatcam: Thin, lensless cameras using coded aperture and computation. *IEEE Transactions on Computational Imaging*, 3(3): 384–397, 2016. 3

[5] Nakkyu Baek, Yujin Lee, Taeyoung Kim, Jaewoo Jung, and Seung Ah Lee. Lensless polarization camera for single-shot full-Stokes imaging. *APL Photonics*, 7(11), 2022. 1, 3

[6] Dhruvjyoti Bagadthey, Sanjana Prabhu, Salman S Khan, D Tony Fredrick, Vivek Boominathan, Ashok Veeraraghavan, and Kaushik Mitra. Flatnet3d: intensity and absolute depth from single-shot lensless capture. *Journal of the Optical Society of America A*, 39(10):1903–1912, 2022. 3

[7] Michael Baltaxe, Tomer Pe'er, and Dan Levi. Polarimetric imaging for perception. In *34th British Machine Vision Conference 2023, BMVC 2023, Aberdeen, UK, November 20-24, 2023*. BMVA, 2023. 4, 6

[8] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009. 2, 6

[9] Eric Bezzam, Yohann Perron, and Martin Vetterli. Towards robust and generalizable lensless imaging with modular learned reconstruction. *IEEE Transactions on Computational Imaging*, 11:213–227, 2025. 3

[10] Vivek Boominathan, Jesse K Adams, Jacob T Robinson, and Ashok Veeraraghavan. PhlatCam: Designed phase-mask based thin lensless camera. *IEEE TPAMI*, 42(7):1618–1629, 2020. 3

[11] Vivek Boominathan, Jacob T Robinson, Laura Waller, and Ashok Veeraraghavan. Recent advances in lensless imaging. *Optica*, 9(1):1–16, 2021. 1, 2

[12] Vivek Boominathan, Jacob T Robinson, Laura Waller, and Ashok Veeraraghavan. Recent advances in lensless imaging. *Optica*, 9(1):1–16, 2022. 3

[13] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3 (1):1–122, 2011. 2, 6

[14] Caslav Brukner and Anton Zeilinger. Malus' law and quantum information. *acta physica slovaca*, 49:647–652, 1999. 3

[15] Xin Cai, Zhiyuan You, Hailong Zhang, Jinwei Gu, Wentao Liu, and Tianfan Xue. Phocolens: Photorealistic and consistent reconstruction in lensless imaging. *Advances in Neural Information Processing Systems*, 37:12219–12242, 2024. 6

[16] Shay Elmalem and Raja Giryes. A lensless polarization camera. In *Computational Optical Sensing and Imaging*, pages CTh7A–1. Optica Publishing Group, 2021. 1, 3

[17] Haofeng Hu, Yang Lin, Xiaobo Li, Pengfei Qi, and Tiegen Liu. Iplnet: a neural network for intensity-polarization imaging in low light. *Optics Letters*, 45(22):6162–6165, 2020. 3

[18] Yi Hua, Shigeki Nakamura, M Salman Asif, and Aswin C Sankaranarayanan. Sweepcam—depth-aware lensless imaging using programmable masks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(7):1606–1617, 2020. 3

[19] Gang Huang, Hong Jiang, Kim Matthews, and Paul Wilford. Lensless imaging by compressive sensing. In *ICIP*, pages 2101–2105, 2013. 3

[20] Zhaorui Huang, Yaqin Zheng, Junhao Li, Yongzhi Cheng, Jian Wang, Zhang-Kai Zhou, and Lin Chen. High-resolution metalens imaging polarimetry. *Nano letters*, 23(23):10991–10997, 2023. 3

[21] Cong Phuoc Huynh, Antonio Robles-Kelly, and Edwin Hancock. Shape and refractive index recovery from single-view polarisation images. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1229–1236. IEEE, 2010. 1

[22] Ulugbek S Kamilov. A parallel proximal algorithm for anisotropic total variation minimization. *IEEE TIP*, 26(2): 539–548, 2016. 6

[23] Salman S Khan, V R Adarsh, Vivek Boominathan, Jasper Tan, Ashok Veeraraghavan, and Kaushik Mitra. Towards photorealistic reconstruction of highly multiplexed lensless images. In *ICCV*, pages 7860–7869, 2019. 3

[24] Salman Siddique Khan, Varun Sundar, Vivek Boominathan, Ashok Veeraraghavan, and Kaushik Mitra. FlatNet: Towards photorealistic scene reconstruction from lensless measurements. *IEEE TPAMI*, 44(4):1934–1948, 2020. 3, 6

[25] Noa Kraicer, Shay Elmalem, Erez Yosef, Hani Barhum, and Raja Giryes. A lensless polarization camera. *arXiv preprint arXiv:2603.17156*, 2026. 1, 3, 7

[26] Nicolas Lefaudeux, Nicolas Lechocinski, Sebastien Breugnot, and Philippe Clemenceau. Compact and robust linear stokes polarization camera. In *Polarization: Measurement, Analysis, and Remote Sensing VIII*, pages 76–87. SPIE, 2008. 4

[27] Lisa W Li, Jaewon Oh, Harris Miller, Federico Capasso, and Noah A Rubin. Flat, wide field-of-view imaging polarimeter. *Optica*, 12(6):799–811, 2025. 3

[28] Xiaobo Li, Haiyu Li, Yang Lin, Jianhua Guo, Jingyu Yang, Huanjing Yue, Kun Li, Chuan Li, Zhenzhou Cheng, Haofeng Hu, et al. Learning-based denoising for polarimetric images. *Optics express*, 28(11):16309–16321, 2020. 3

[29] Ju Liu, Jin Duan, Youfei Hao, Guangqiu Chen, Hao Zhang, and Yue Zheng. Polarization image demosaicing and rgb image enhancement for a color polarization sparse focal plane array. *Optics Express*, 31(14):23475–23490, 2023. 3, 4

[30] Zhuoyan Liu, Bo Wang, Lizhi Wang, Chenyu Mao, and Ye Li. ShareCMP: Polarization-aware RGB-P semantic segmentation. *IEEE TCSVT*, 2025. 4, 7

[31] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6

[32] Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. Structure-preserving super resolution with gradient guidance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7769–7778, 2020. 6

[33] Jiayi Ma, Linfeng Tang, Fan Fan, Jun Huang, Xiaoguang Mei, and Yong Ma. Swinfusion: Cross-domain long-range learning for general image fusion via swin transformer. *IEEE/CAA Journal of Automatica Sinica*, 9(7):1200–1217, 2022. 8

[34] Rakesh Raj Madavan, Akshat Kaimal, Vinayak Gupta, Rohit Choudhary, Chandrakala Shanmuganathan, Kaushik Mitra, et al. Ganesh: Generalizable nerf for lensless imaging. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 9499–9508. IEEE, 2025. 3

[35] Jennifer R Miller, Cheng-Yu Wang, Christine D Keating, and Zhiwen Liu. Particle-based reconfigurable scattering masks for lensless imaging. *ACS nano*, 14(10):13038–13046, 2020. 3

[36] Kristina Monakhova, Kyrollos Yanny, Neerja Aggarwal, and Laura Waller. Spectral diffusercam: lensless snapshot hyperspectral imaging with a spectral filter array. *Optica*, 7(10): 1298–1307, 2020. 3, 4, 6, 8

[37] Mara Pistellato, Filippo Bergamasco, Tehreem Fatima, and Andrea Torsello. Deep demosaicing for polarimetric filter array cameras. In *IEEE TIP*, pages 2017–2026, 2022. 3

[38] Jessica C Ramella-Roman, Ilyas Saytashev, and Mattia Piccini. A review of polarization-based imaging technologies for clinical and preclinical applications. *Journal of Optics*, 22(12):123001, 2020. 2

[39] Carla Rodríguez, Albert Van Eeckhout, Laia Ferrer, Enrique Garcia-Caurel, Emilio González-Arnay, Juan Campos, and Angel Lizana. Polarimetric data-based model for tissue recognition. *Biomed. Opt. Express*, 12(8):4852–4872, 2021. 2

[40] 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. 6

[41] Kiyotaka Sasagawa, Ryoma Okada, Makito Haruta, Hironari Takehara, Hiroyuki Tashiro, and Jun Ohta. Polarization image sensor for highly sensitive polarization modulation imaging based on stacked polarizers. *IEEE Transactions on Electron Devices*, 69(6):2924–2931, 2022. 3

[42] Michael Shribak. Polychromatic polarization microscope: bringing colors to a colorless world. *Scientific reports*, 5: 17340, 2015. 2

[43] Ayan Sinha, Justin Lee, Shuai Li, and George Barbastathis. Lensless computational imaging through deep learning. *Optica*, 4(9):1117–1125, 2017. 3

[44] Leslie N Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications*, pages 369–386. SPIE, 2019. 6

[45] Jian Sun, Huibin Li, Zongben Xu, et al. Deep ADMM-Net for compressive sensing MRI. *NeurIPS*, 29, 2016. 3

[46] J Scott Tyo, Dennis L Goldstein, David B Chenault, and Joseph A Shaw. Review of passive imaging polarimetry for remote sensing applications. *Applied optics*, 45(22):5453–5469, 2006. 1, 2

[47] Wenbo Wang, Wenjun He, and Zhencong Xiong. Lensless polarization imaging system based on a random coded mask. In *Fourth International Computational Imaging Conference (CITA 2024)*, pages 1106–1111. SPIE, 2025. 1, 3

[48] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6

[49] Lawrence B Wolff and Terrance E Boult. Constraining object features using a polarization reflectance model. *IEEE TPAMI*, 13(7):635–657, 2002. 4

[50] Kaite Xiang, Kailun Yang, and Kaiwei Wang. Polarization-driven semantic segmentation via efficient attention-bridged fusion. *Optics Express*, 29(4):4802–4820, 2021. 4, 7

[51] WANMJ XU XY et al. Colorpolarnet: Residual dense network based chromatic intensity polarization imaging in low light environment. *IEEE Transactions on Instrumentation and Measurement*, 71(1), 2022. 3

[52] Kyrollos Yanny, Nick Antipa, William Liberti, Sam Dehaeck, Kristina Monakhova, Fanglin Linda Liu, Konlin Shen, Ren Ng, and Laura Waller. Miniscope3d: optimized single-shot miniature 3d fluorescence microscopy. *Light: Science & Applications*, 9(171), 2020. 2

[53] Mingde Yao, Menglu Wang, King-Man Tam, Lingen Li, Tianfan Xue, and Jinwei Gu. PolarFree: Polarization-based reflection-free imaging. In *CVPR*, pages 10890–10899, 2025. 3

[54] Nebiyou Yismaw, Ulugbek S Kamilov, and M Salman Asif. Domain expansion via network adaptation for solving inverse problems. *IEEE Transactions on Computational Imaging*, 10:549–559, 2024. 3

[55] Erez Yosef and Raja Giryes. Difuzcam replacing camera lens with a mask and a diffusion model for generative ai based flat camera design. *Scientific Reports*, 15(1):43059, 2025. 3

[56] Erez Yosef and Raja Giryes. Tell me what you see: Text-guided real-world image denoising. *IEEE Open Journal of Signal Processing*, 2025. 3

[57] Tianjiao Zeng and Edmund Y Lam. Robust reconstruction with deep learning to handle model mismatch in lensless imaging. *IEEE Transactions on Computational Imaging*, 7: 1080–1092, 2021. 6

[58] Xianglong Zeng, Yuan Luo, Xiaojing Zhao, and Wenbin Ye. An end-to-end fully-convolutional neural network for division of focal plane sensors to reconstruct s0, dolp, and aop. *Optics express*, 27(6):8566–8577, 2019. 3

[59] Jian Zhang and Bernard Ghanem. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *CVPR*, pages 1828–1837, 2018. 3

[60] Kailong Zhang, Youwei Lyu, Heng Guo, Si Li, Zhanyu Ma, and Boxin Shi. Polaranything: Diffusion-based polarimetric image synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 26466–26476, 2025. 3, 6

[61] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 6

[62] Yucheng Zheng, Yi Hua, Aswin C Sankaranarayanan, and M Salman Asif. A simple framework for 3D lensless imaging with programmable masks. In *ICCV*, pages 2603–2612, 2021. 3

[63] Chu Zhou, Minggui Teng, Xinyu Zhou, Chao Xu, Imari Sato, and Boxin Shi. Learning to deblur polarized images. *International Journal of Computer Vision*, pages 1–16, 2025. 3

[64] Assaf Zomet and Shree K Nayar. Lensless imaging with a controllable aperture. *CVPR*, 1:339–346, 2006. 3

[65] Jiawei Zuo, Jing Bai, Shinhyuk Choi, Ali Basiri, Xiahui Chen, Chao Wang, and Yu Yao. Chip-integrated metasurface full-stokes polarimetric imaging sensor. *Light: Science & Applications*, 12(1):218, 2023. 3