

Pol-CACTI: A System and Dataset for High-Speed Polarized Video Compressive Imaging

Yunfeng Song^{1,2*}, Yidong Luo^{1,2*}, Ping Wang^{2†}, Xingjian Jiang^{1,2}, and
Xin Yuan^{2†}

¹ Zhejiang University, Hangzhou, China

² School of Engineering, Westlake University, Hangzhou, China

{songyunfeng, luoyidong, wangping, jiangxingjian, xyuan}@westlake.edu.cn

Abstract. Capturing high-speed polarization dynamics is crucial for revealing physical cues like stress variations and surface structures that are invisible in conventional intensity imaging. While snapshot compressive imaging (SCI) enables high-speed video reconstruction from a single measurement, extending it to division-of-focal-plane (DoFP) polarization sensing introduces a fundamentally more ill-posed inverse problem due to the coupling between temporal multiplexing and spatial polarization mosaicing. To address this challenge, we propose Polarization Coded Aperture Compressive Temporal Imaging (Pol-CACTI), a hardware–software co-designed framework for polarized video SCI. First, we build an optical system that encodes DoFP polarization dynamics into a single snapshot through spatiotemporal modulation and integration. Second, we construct a large-scale high-quality polarized video dataset with carefully denoised Stokes-domain supervision, enabling supervised learning of polarization video SCI reconstruction models. Third, we develop a physics-driven reconstruction network that jointly performs temporal recovery and polarization demosaicing, where the Polarization-Aware Learnable Back-Projection (PA-LBP) module is proposed to enforce measurement consistency with the raw DoFP mosaic and mitigates error accumulation caused by decoupled reconstruction pipelines. Extensive experiments demonstrate that our method significantly improves both intensity fidelity and polarization estimation accuracy. The dataset and code are released to <https://github.com/kaeqxhailesxya/Pol-CACTI>.

Keywords: Polarization Imaging System · Polarized Video Dataset · Snapshot Compressive Imaging

1 Introduction

High-speed video acquisition is typically achieved using high-frame-rate sensors, which directly increase temporal sampling at the cost of significant hardware complexity and data bandwidth. Video snapshot compressive imaging (SCI)

* Equal contribution. † Corresponding authors.

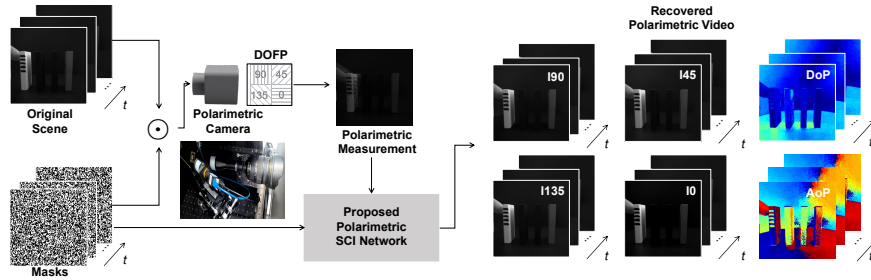


Fig. 1: Overview of the polarized video SCI system and reconstruction framework.

provides an elegant alternative that trades computation for bandwidth. As illustrated in Figure 1, a typical SCI pipeline consists of a *hardware encoder* that temporally multiplexes multiple frames into a single measurement, and a *software decoder* that subsequently recovers the latent frames from the compressed measurement [8, 9, 16, 30, 33, 38, 49, 51]. This sensing-to-computation paradigm shifts the burden of high-speed acquisition from hardware to algorithmic reconstruction, enabling imaging beyond conventional frame-rate limitations.

While SCI enables high-speed intensity acquisition, it remains insufficient for many physical analysis tasks where intensity alone cannot capture material, structural, or geometric variations. Polarization imaging provides complementary physical cues beyond radiance, including Angle of Polarization (AoP) and Degree of Polarization (DoP), which are essential for surface characterization [1, 2, 10, 24, 32, 45], stress analysis [34], and material discrimination under low-contrast or textureless conditions [35, 42]. As illustrated in Figure 1, standard intensity imaging fails to distinguish the black rotating base from the black blocks due to their similarly low reflectance, whereas polarization cues clearly reveal material boundaries and rotational dynamics [2]. This demonstrates the value of polarization beyond intensity, especially for dynamic scenes with low radiometric contrast. However, capturing high-speed polarization video remains fundamentally challenging. Division-of-focal-plane (DoFP) polarization cameras provide single-shot measurements but suffer from limited frame rates, leading to motion blur and temporal information loss in dynamic scenes [23].

Existing attempts to address high-speed polarization imaging face a fundamental sensing bottleneck. While DoFP sensors are constrained by spatial multiplexing and insufficient temporal resolution, alternative approaches such as event-based Mueller-matrix sensing and single-shot Mueller polarimetry remain limited by object motion, sensor bandwidth, and signal stability. Rapid motion may distort polarization cues, where dense event bursts interfere with polarimetric signals and introduce motion artifacts, while high-frequency events may exceed readout and transmission bandwidth, leading to data dropouts. Moreover, practical optical implementations, such as rotating waveplates, suffer from additional temporal noise caused by mechanical jitter, surface contamination, and specular reflections [20, 21]. These limitations indicate that existing hardware-centric solutions are fundamentally inadequate for reliable high-speed polarization imaging.

Directly combining SCI with DoFP polarization sensing introduces a tightly coupled inverse problem, where temporal multiplexing is entangled with spatial polarization mosaicing. In this setting, a naive two-stage pipeline—first applying SCI reconstruction and then performing polarization demosaicing—is fundamentally suboptimal. SCI reconstruction artifacts are introduced during temporal inversion and subsequently propagated and amplified during demosaicing, where spatial interpolation redistributes errors across polarization channels. Moreover, polarization quantities such as AoP and DoP are nonlinear functions of the recovered Stokes parameters, leading to further amplification of reconstruction errors and physically inconsistent estimates [25]. Existing demosaicing methods, designed for clean mosaic inputs, are not robust to SCI-induced degradations, making them unsuitable in this setting. These challenges highlight the need for a unified framework that jointly models temporal reconstruction and polarization demosaicing under a physically consistent formulation.

In this paper, we propose Pol-CACTI, a hardware–software co-designed framework for high-speed polarized video snapshot compressive imaging. Our main contributions are summarized as follows:

- **System.** We present the first high-speed polarized video snapshot compressive imaging prototype, by integrating a DoFP polarization sensor with a Coded Aperture Compressive Temporal Imaging (CACTI) system, enabling single-snapshot capture of high-speed polarization dynamics.
- **Dataset.** We construct a large-scale polarized video dataset containing 8,889 frames at 2048×2048 resolution. The dataset provides Stokes-domain supervision computed from denoised polarization measurements, enabling reliable supervised learning for polarization video reconstruction.
- **Algorithm.** We design a physics-driven joint reconstruction network that simultaneously performs SCI reconstruction and polarization demosaicing. A key component is the Polarization-Aware Learnable Back-Projection (PALBP) module, which enforces measurement consistency with the raw DoFP mosaic and mitigates error accumulation in decoupled pipelines.

Extensive experiments on both simulated and real-world captured scenes demonstrate consistent improvements in intensity fidelity and polarization estimation accuracy across compression ratios, producing more stable AoP and DoP reconstructions and better preserving physical consistency in dynamic scenes.

2 Related Work

2.1 Video Snapshot Compressive Imaging

Video Snapshot Compressive Imaging (SCI) [49, 51] captures high-speed video from a single compressed measurement by exploiting temporal multiplexing and computational reconstruction. As a representative implementation, CACTI employs spatial light modulators, such as coded aperture translation [16], Liquid Crystal on Silicon (LCoS) [8, 30], or Digital Micromirror Devices (DMD) [38],

to impose temporally varying masks during exposure. Given this encoding process, SCI reconstruction is formulated as an ill-posed inverse problem of recovering latent video frames from a single measurement. Early approaches relied on model-based optimization methods that impose handcrafted priors on natural video structure, including sparsity [8, 30], total variation [48], and nonlocal low-rank [15]. However, these approaches suffer from slow speed and limited quality. Recent advances have shifted toward data-driven approaches that learn reconstruction priors directly from large-scale datasets. Early deep learning methods have extensively exploited on straight [3, 36, 37, 40, 43], U-shaped [26, 39], recurrent [4, 5], unrolling [18, 22, 46, 52], and plug-and-play [47, 50] architectures to model spatiotemporal dependencies, achieving better fidelity and faster inference than optimization-based methods.

Despite these advances, most SCI methods assume intensity-only imaging under monochrome or RGB sensing configurations. Extending SCI to DoFP polarization sensors introduces additional mosaic sampling due to micro-polarizer arrays, which couples spatial subsampling with temporal multiplexing. This leads to a more ill-posed inverse problem and motivates polarization video SCI for high-speed polarized scene capture.

2.2 Video Polarization Demosaicing and Datasets

Data availability remains a key factor for learning-based polarization imaging. Existing datasets, including MMS-DATA [14] and that of Rahman *et al.* [28], are mainly composed of static scenes, limiting their applicability to high-speed dynamic polarization recovery. In dynamic settings, reliable temporal supervision of AoP and DoP is further hindered by noise in derived polarimetric quantities, making stable ground-truth construction challenging.

In parallel, polarization demosaicing has been extensively studied as an image restoration problem. Early methods rely on interpolation or super-resolution techniques to reduce instantaneous field-of-view artifacts [6, 7, 29], while later approaches incorporate model-based inversion and deep neural networks for improved reconstruction quality [27, 31]. Recent methods, such as Polanalyser [19], PIDSR [53], SwinCPD [41], and PUGDiff [13], further improve performance using joint optimization, Transformer architectures, or diffusion priors. However, these methods are fundamentally designed for single-image inputs and assume clean polarization mosaics.

When extended to video SCI scenarios, polarization demosaicing cannot be treated independently from the reconstruction process. In particular, SCI reconstruction and polarization demosaicing become tightly coupled when applied sequentially: errors introduced during SCI inversion propagate into polarization channels and are further amplified by spatial interpolation, while nonlinear transformations from Stokes parameters to AoP and DoP exacerbate instability. This indicates that treating SCI reconstruction and polarization demosaicing as separate stages is suboptimal, motivating the joint modeling of temporal reconstruction and polarization demosaicing within a unified framework.

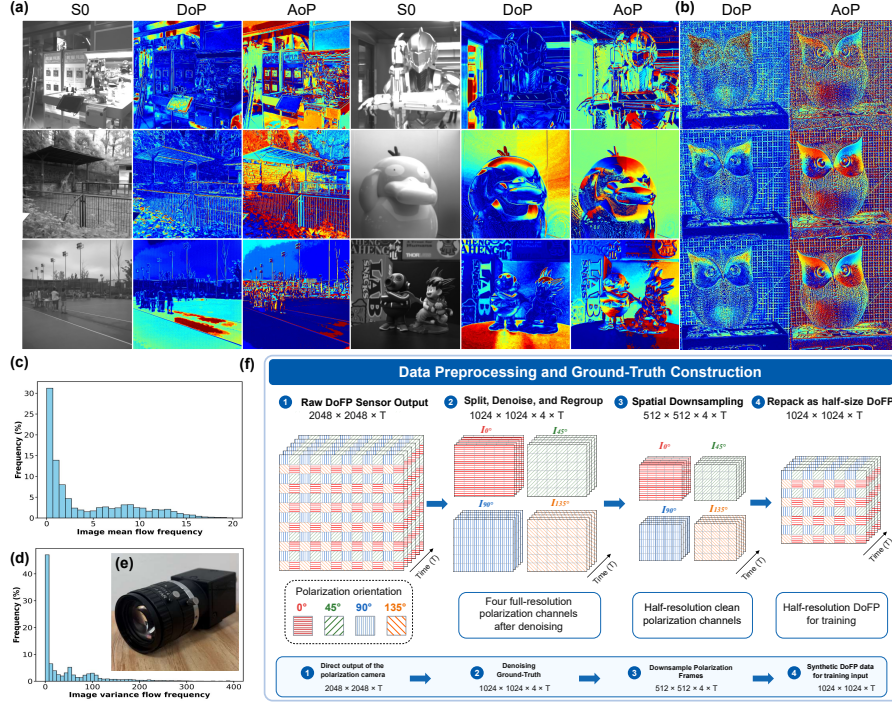


Fig. 2: (a) Example video clips (S0, DoP, and AoP); (b) Denoising comparison: noisy inputs (top), denoised outputs (middle), and averaged GT (bottom); (c) Histogram of mean optical flow magnitude; (d) Histogram of optical flow variance; (e) The polarization camera used for data acquisition; (f) Data preprocessing and GT construction pipeline. More video frames are in the supplement.

Fig. 3: Specifications and statistics of the proposed polarized video dataset. The dataset provides high-resolution, global shutter sequences across diverse scenarios and material types to ensure physical reliability.

Acquisition		Dataset Diversity	
Lens	LBTEK LBMVL-MV1236M-MP	Total Sequences	317 (307 Train / 10 Test)
Sensor	Sony IMX250MZR (DoFP)	Total Frames	8,889 (8,409 Train / 480 Test)
Shutter	Global Shutter	Materials	Metal, Plastic, Glass, Ceramic, Fabric
Resolution	2048 × 2048	Polarization	Specular, Diffuse, Mixed Cues
Frame Rate	30 fps	Lighting	Daylight, Night, Artificial Light
Pixel Size	3.45 μ m × 3.45 μ m	Scenarios	Indoor, Outdoor, Streets, Animals, Object Rotation & Translation

3 Mathematical Model

Video SCI. Figure 1 illustrates the video compressive imaging (SCI) encoding pipeline. In a monochrome video SCI system, an input sequence of T grayscale frames $\{X_t\}_{t=1}^T \in \mathbb{R}^{n_x \times n_y}$ undergoes spatial modulation via predetermined binary masks $\{M_t\}_{t=1}^T \in \mathbb{R}^{n_x \times n_y}$. The modulated frames are subsequently integrated across the temporal dimension, producing a single compressive measurement $Y \in \mathbb{R}^{n_x \times n_y}$ at the native sensor resolution captured by the standard

image sensor. The encoding model is expressed as: where $\odot$ denotes element-wise multiplication, and $\mathbf{Z}$ represents measurement noise.

$$\mathbf{Y} = \sum_{t=1}^T \mathbf{X}_t \odot \mathbf{M}_t + \mathbf{Z}, \quad (1)$$

Color Video SCI. In a color video SCI system, we employ a commonly used Bayer-patterned filter array sensor, where each pixel records only one of the three primary color channels, red (R), green (G), or blue (B), arranged spatially over the sensor plane in patterns such as ‘RGGB’. Given that neighboring pixels correspond to distinct color channels, we decompose the original compressive measurement $\mathbf{Y}$ into four sub-measurements $\{\mathbf{Y}^r, \mathbf{Y}^{g1}, \mathbf{Y}^{g2}, \mathbf{Y}^b\} \in \mathbb{R}^{(n_x/2) \times (n_y/2)}$, following the Bayer filter’s spatial configuration.

Polarized Video SCI. In a polarized video SCI system, we utilize a Division of Focal Plane (DoFP) sensor to capture polarization information. Let $\mathbf{X}_t^{(\alpha)} \in \mathbb{R}^{n_x \times n_y}$ denote the full-resolution intensity at polarization angle $\alpha \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ and time index t , and let $\mathbf{M}_t$ be the encoding mask. Similar to the Bayer case, the DoFP sensor induces a spatial mosaic via a micro-polarizer array $\mathbf{D}^{(\alpha)}$, which imposes an angle-specific downsample pattern and only allows light of corresponding polarization orientations to pass at each pixel location. The captured raw measurement $\mathbf{Y}$ is thus formulated as:

$$\mathbf{Y} = \sum_{t=1}^T \mathbf{M}_t \odot \left(\sum_{\alpha} \mathbf{D}^{(\alpha)} \odot \mathbf{X}_t^{(\alpha)} \right) + \mathbf{Z}. \quad (2)$$

Analogous to the Bayer filter array in color imaging (which uses an ‘RGGB’ pattern), the DoFP sensor employs a micro-polarizer array arranged in a periodic 2×2 super-pixel configuration. Given that neighboring pixels correspond to distinct polarization orientations, we decompose the raw polarized measurement $\mathbf{Y}$ into four sub-measurements $\{\mathbf{Y}^{(0)}, \mathbf{Y}^{(45)}, \mathbf{Y}^{(90)}, \mathbf{Y}^{(135)}\} \in \mathbb{R}^{\frac{n_x}{2} \times \frac{n_y}{2}}$, following the micro-polarizer mosaic’s spatial configuration. Consequently, the reconstruction involves a joint problem of reconstruction and demosaicing: we aim to recover the full-resolution polarization volume $\{\mathbf{X}^{(\alpha)}\} \in \mathbb{R}^{n_x \times n_y \times 4}$ from these spatially downsampled sub-measurements.

After vectorization, define $\mathbf{D}_t = \text{diag}(\text{vec}(\mathbf{M}_t))$ and $\mathbf{H} = [\mathbf{D}_1, \dots, \mathbf{D}_T]$. For each polarization-angle channel,

$$\mathbf{y}^{(\alpha)} = \mathbf{H}^{(\alpha)} \mathbf{x}^{(\alpha)} + \mathbf{z}^{(\alpha)}, \quad \mathbf{H}^{(\alpha)} = \mathbf{S}^{(\alpha)} \mathbf{H}, \quad (3)$$

where $\mathbf{S}^{(\alpha)}$ is the sampling matrix corresponding to the angle-specific down-sampling, and $\mathbf{x}^{(\alpha)} = [\text{vec}(\mathbf{X}_1^{(\alpha)})^\top, \dots, \text{vec}(\mathbf{X}_T^{(\alpha)})^\top]^\top$. Stacking four polarization angles yields the block-diagonal forward model

$$\mathbf{y}_p = \begin{bmatrix} \mathbf{H}^{(0)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^{(45)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}^{(90)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}^{(135)} \end{bmatrix} \mathbf{x}_p + \mathbf{z}_p, \quad (4)$$

with $\mathbf{y}_p = [\mathbf{y}^{(0)}; \mathbf{y}^{(45)}; \mathbf{y}^{(90)}; \mathbf{y}^{(135)}]$ and $\mathbf{x}_p = [\mathbf{x}^{(0)}; \mathbf{x}^{(45)}; \mathbf{x}^{(90)}; \mathbf{x}^{(135)}]$. Here, $\mathbf{z}^{(\alpha)}$ and $\mathbf{z}_p$ are the vectorized noise terms corresponding to $\mathbf{Z}$.

4 Proposed Pol-CACTI System

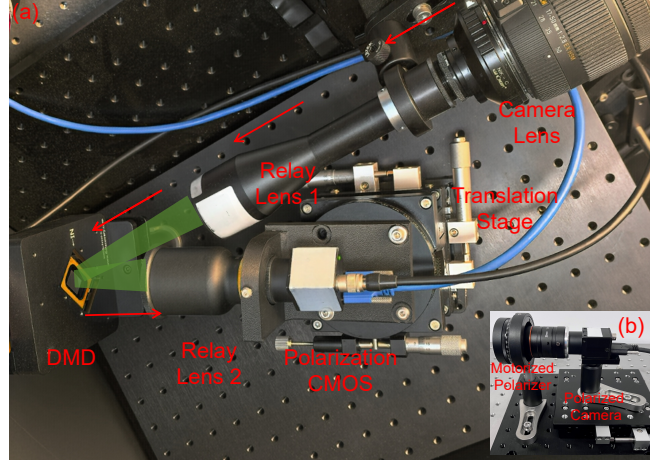


Fig. 4: (a) The proposed Pol-CACTI system featuring a DMD and a polarization camera. (b) The optical calibration setup utilized to measure the sensor’s polarimetric calibration matrix.

Pol-CACTI Hardware. The optical configuration of our practical polarization video SCI prototype is shown in Fig. 4. Light from the moving target is imaged by a camera lens (Nikon Macro 35 mm) and a primary relay lens (Coolens WWK10-110-111) onto a digital micromirror device (DMD) (Fldiscovery F4320, 1920×1080 , $10.8 \mu\text{m}$ pitch). The DMD displays preloaded 1-bit pseudo-random masks to encode the high-speed scene, and the coded snapshot is relayed by a secondary lens (Daheng Optics GCO-232206) onto a monochrome polarization camera (FLIR BFS-U3-51S5P-C, USB 3.1, $3.45 \mu\text{m}$ pixel size), which records a single compressive measurement per exposure. The camera operates at 10 fps in our real prototype acquisition setting; with compression ratio C_r , the reconstructed temporal sampling rate is $10 \times C_r$ fps.

Hardware Alignment. We aggregate each 4×4 DMD micromirror block as one encoding unit and project it to align with a corresponding 4×4 block of sensor pixels, which contains four complete 2×2 DoFP macro-pixels. This DMD-to-sensor registration is calibrated using a high-precision multi-axis linear translation stage, learnably adjusting translation and focus to minimize misregistration between mask edges and the pixel grid. The resulting physical alignment ensures that the same mask pattern modulates all polarization angles consistently, without requiring software-based geometric warping. To ensure accuracy, we characterized the sensor’s response using a uniform integrating sphere field and a precision rotating polarizer before the experiment (see Supplementary

Material for calibration procedures). In our setup, we set $C_r = 8$ and synchronize the DMD refresh rate to 80 Hz while the camera captures at 10 fps, thereby compressing eight latent frames into a single polarized snapshot.

5 Polarized Video Dataset

To bridge the critical gap between theoretical algorithms and real-world deployment, we introduce a novel and high-fidelity polarimetric video dataset.

5.1 Data Acquisition and Preprocessing

Data Acquisition. As shown in Figure 2, the data acquisition system is built around a FLIR BFS-U3-51S5P-C polarization camera equipped with the Sony IMX250MZR DoFP sensor and an LBTEK LBMVL-MV1236M-MP lens. The DoFP sensor integrates a micro-polarizer array (MPA) on the focal plane and simultaneously captures four linear polarization orientations ($0^\circ, 45^\circ, 90^\circ, 135^\circ$). The camera operates in global-shutter mode, which avoids rolling-shutter distortions and enables reliable recording of dynamic scenes. We configure the capture frame rate to 30 fps under typical indoor illumination conditions to balance signal-to-noise ratio (SNR) and motion sharpness. Each frame is acquired at a native resolution of 2048×2048 with a pixel size of $3.45 \mu\text{m} \times 3.45 \mu\text{m}$.

The resulting dataset contains 317 polarized video sequences, including 307 sequences for training and 10 sequences for testing, totaling 8,889 high-resolution frames. Among them, 8,409 frames are used for training and 480 frames are reserved for testing, as summarized in Table 3. To ensure diverse polarimetric observations and robust generalization, the dataset covers multiple material categories, including metal, plastic, glass, ceramic, and fabric, and contains specular, diffuse, and mixed polarization cues. The scenes are captured under daylight, nighttime, and artificial lighting conditions, spanning indoor and outdoor environments, streets, animals, and object rotation and translation. Example S0, DoP, and AoP video clips are shown in Figure 2(a), while the optical-flow statistics in Figure 2(c) and Figure 2(d) further characterize the motion diversity of the collected dynamic sequences.

Ground-Truth Generation. The dataset and the Pol-CACTI prototype are related but not strictly coupled. The dataset provides supervised dynamic polarimetric sequences, while the forward model of Pol-CACTI defines how training pairs are generated under a specific coding and sampling configuration. Specifically, the native 2048×2048 at 30 fps recordings are converted from DoFP RAW measurements into demosaiced polarization images and denoised Stokes-domain labels for supervision. During model training, data augmentation operations, including cropping, resizing, and temporal sampling, change the effective spatial resolution, frame rate, and temporal range of the input-output pairs. Accordingly, the collected sequences serve as high-quality dynamic polarimetric video sources with reliable temporal consistency, from which training samples are synthesized according to the Pol-CACTI imaging model.

Data Preprocessing. Raw DoFP measurements contain noise that is strongly amplified when computing AoP and DoP. To produce reliable ground truth, we adopt a physics-aware denoising pipeline based on NLCSR [17]. We use patch size 8×8 and search window 30×30 , and extend non-local regularization from spatial to spatio-temporal matching so that similar patches are grouped across neighboring frames. The grouped patches are optimized with adaptive dictionary learning and centralized sparse coding, which suppresses random sensor noise while preserving polarization edges and textures used by Stokes estimation. Visual examples are shown in Figure 2(b); full denoising settings are provided in the supplementary material.

6 Proposed Reconstruction Method

6.1 Video SCI Reconstruction

To recover high-fidelity polarization video, we propose a physics-driven joint reconstruction network following a coarse-to-fine reconstruction paradigm. First, a Video Reconstruction Backbone reconstructs the raw DoFP mosaic sequence from the snapshot measurement. We employ EfficientSCI [36] for this stage explicitly due to its demonstrated robustness and computational efficiency in handling high-dimensional compressive data.

6.2 Polarization Demosaicing

To recover a full-resolution polarization video from DoFP mosaic sub-measurements, we introduce a refinement module equipped with a Polarization-Aware Learnable Back-Projection (PA-LBP) mechanism. The module refines predictions in feature space while explicitly enforcing consistency with the raw mosaic input. Since the DoFP measurement preserves polarization cues before any learned transformation, the back-projection step serves as a physics anchor that suppresses spurious details and steers the reconstruction toward a solution compatible with the sensor observation. The refinement stage consists of two parts: Deep Feature Refinement and PA-LBP Correction.

Deep Feature Refinement. Let $\mathbf{F}_{sub}$ denote the feature map from the preceding reconstruction stage. We first lift $\mathbf{F}_{sub}$ with a convolutional head and then apply $K = 5$ residual blocks to enhance high-frequency polarization structures. Each block follows a standard residual form, $\text{Res}(\mathbf{x}) = \mathbf{x} + \text{Conv}(\sigma(\text{Conv}(\mathbf{x})))$. We then upsample and further filter the features to mitigate grid artifacts, producing a high-resolution residual map. This process is written as:

$$\begin{aligned}
 \mathbf{F}_{head} &= \text{Conv}(\mathbf{F}_{sub}), \\
 \mathbf{F}_k &= \text{Res}(\mathbf{F}_{k-1}), \quad k = 1, \dots, 5, \\
 \mathbf{F}_{body} &= \text{Conv}(\mathbf{F}_5) + \mathbf{F}_{head}, \\
 \mathbf{F}_{up} &= \sigma(\text{Conv}(\sigma(\text{Conv}(\text{Up}_{bi}(\mathbf{F}_{body}))))), \\
 \mathbf{I}_{res} &= \text{Conv}(\mathbf{F}_{up}),
 \end{aligned} \tag{5}$$

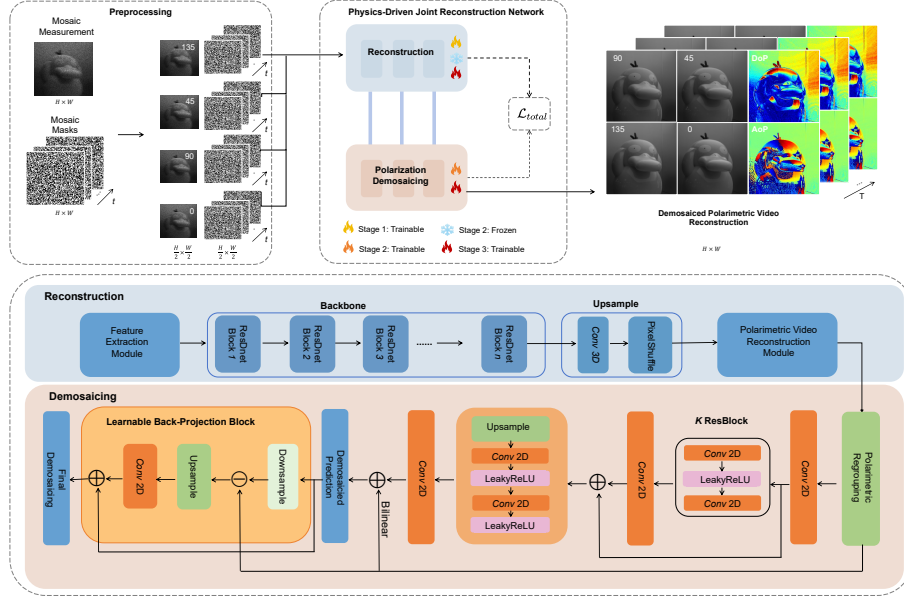


Fig. 5: Overview of the proposed joint reconstruction network.

where Conv denotes a 3×3 convolution, σ is LeakyReLU, and Up_{bi} is bilinear upsampling. $\mathbf{I}_{res}$ represents the learned high-frequency residual.

PA-LBP Correction. We first combine the learned residual with the bilinearly upsampled polarization component tensor to obtain the model-predicted demosaiced estimate $\mathbf{I}_{pred}$. The PA-LBP step then re-projects $\mathbf{I}_{pred}$ to the polarization component space using a learnable strided convolution $\text{Conv}_\downarrow$, computes the discrepancy with the split polarization components, and back-projects this error to correct the full-resolution output:

$$\begin{aligned}
 \mathbf{I}_{pred} &= \text{Up}_{bi}(\mathbf{I}_{pc}) + \mathbf{I}_{res}, \\
 \mathbf{I}_{proj} &= \text{Conv}_\downarrow(\mathbf{I}_{pred}), \\
 \mathbf{E}_{diff} &= \mathbf{I}_{pc} - \mathbf{I}_{proj}, \\
 \mathbf{I}_{corr} &= \text{Conv}(\text{Up}_{bi}(\mathbf{E}_{diff})), \\
 \mathbf{I}_{final} &= \mathbf{I}_{pred} + \mathbf{I}_{corr}.
 \end{aligned} \tag{6}$$

Here, $\mathbf{I}_{pc}$ denotes the regrouping polarization component tensor $(I_0, I_{45}, I_{90}, I_{135})$, and $\mathbf{I}_{res}$ is the learned high-resolution residual. $\mathbf{I}_{pred}$ is the model-predicted demosaiced estimate before PA-LBP correction, and $\mathbf{I}_{proj}$ is obtained by $\text{Conv}_\downarrow$. $\mathbf{E}_{diff}$, $\mathbf{I}_{corr}$, and $\mathbf{I}_{final}$ denote the discrepancy, back-projected correction, and final corrected high-resolution polarization output, respectively.

6.3 Physics-Constrained Loss Functions

To ensure both radiometric fidelity and polarimetric validity, we adopt a multi-stage physics-constrained objective. The total loss is calculated as the weighted

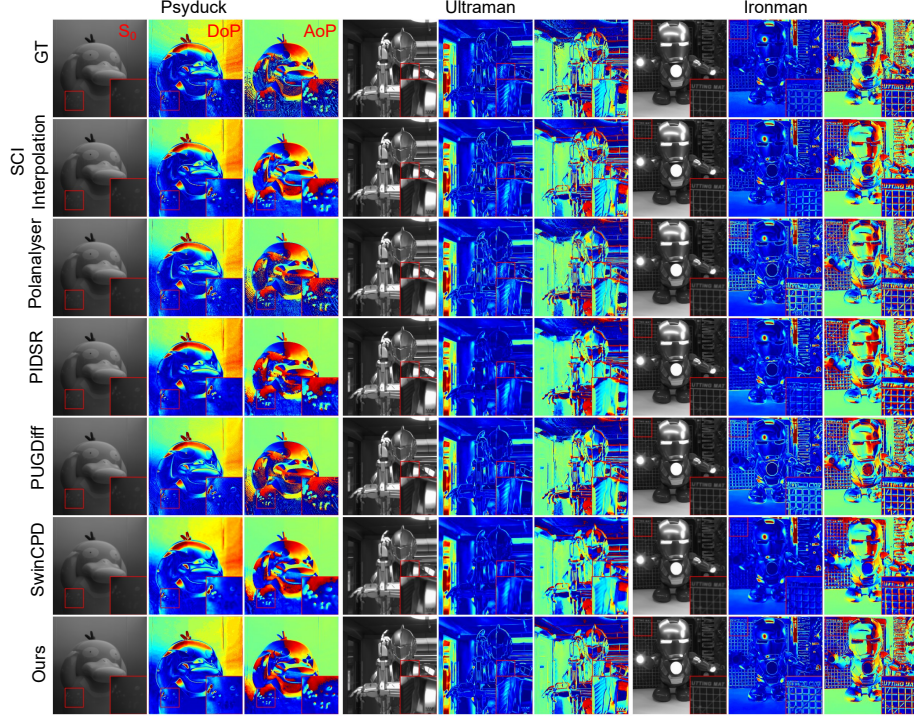


Fig. 6: Reconstruction results on simulated test indoor scenes. All results are visualized with percentile-based normalization for obvious comparison. We provide zoomed-in views of local regions for enhanced detail inspection.

sum of losses applied to both the intermediate mosaic estimation ($\hat{\mathbf{X}}_{\text{mosaic}}$) and the final full-resolution reconstruction ($\hat{\mathbf{X}}_{\text{full}}$):

$$\mathcal{L}_{\text{total}} = \mathcal{L}(\hat{\mathbf{X}}_{\text{mosaic}}, \mathbf{X}_{\text{mosaic}}^{\text{gt}}) + \mathcal{L}(\hat{\mathbf{X}}_{\text{full}}, \mathbf{X}_{\text{full}}^{\text{gt}}). \quad (7)$$

For each stage, the loss is a composite of four terms:

$$\mathcal{L} = \lambda_{\text{raw}} \mathcal{L}_{\text{raw}} + \lambda_{S_0} \mathcal{L}_{S_0} + \lambda_{\text{DoP}} \mathcal{L}_{\text{DoP}} + \lambda_{\text{AoP}} \mathcal{L}_{\text{AoP}}. \quad (8)$$

We adopt the Charbonnier loss [12] as the robust metric for all intensity-based terms. The Charbonnier loss serves as a differentiable variant of the ℓ_1 norm. It offers greater robustness to outliers and heavy-tailed noise distributions, ensuring sharper reconstruction of polarization details. Specifically, we apply $\mathcal{L}_{\text{Char}}$ to supervise the raw analyzer channels ($\mathcal{L}_{\text{raw}}$), the total intensity Stokes component ($\mathcal{L}_{S_0}$), and the Degree of Polarization ($\mathcal{L}_{\text{DoP}}$).

In contrast, for the Angle of Polarization (AoP), standard norm-based losses fail due to phase wrapping. To address the π -periodicity and the singularity of AoP in unpolarized regions, we employ a DoP-weighted cosine difference loss:

$$\mathcal{L}_{\text{AoP}} = \mathbb{E} \left[\text{DoP}_{\text{gt}} \cdot \left(1 - \cos \left(2(\hat{\theta} - \theta_{\text{gt}}) \right) \right) \right]. \quad (9)$$

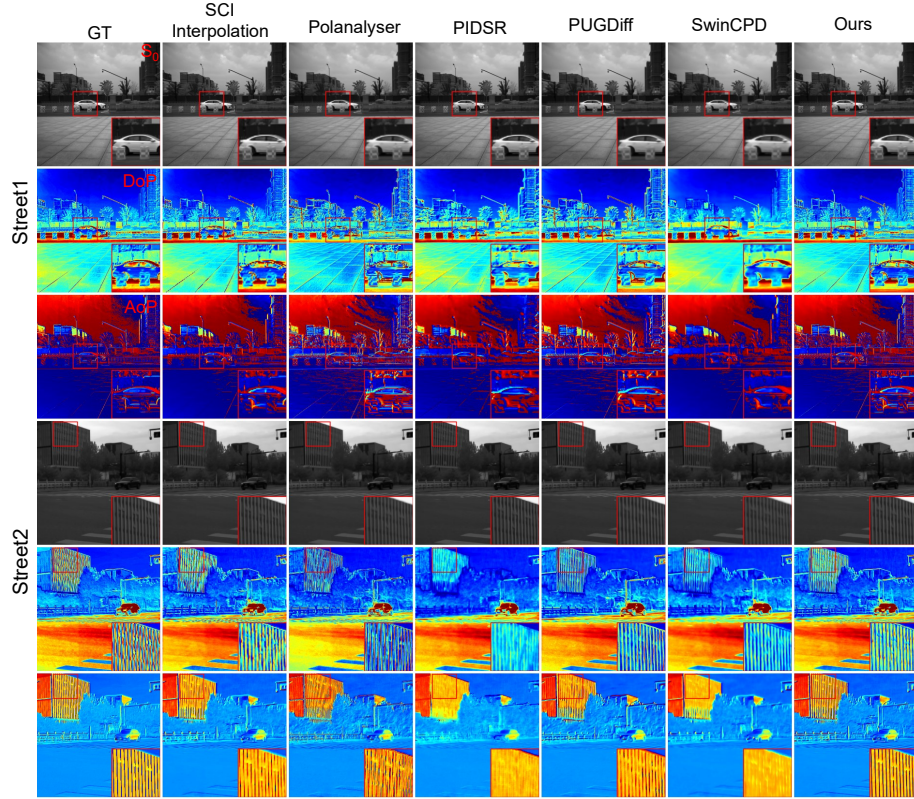


Fig. 7: Qualitative comparison on simulated testing outdoor scenes.

This formulation effectively treats orientation errors modulo π in a physically consistent manner and suppresses gradients in regions where polarization information is physically meaningless (i.e., $\text{DoP} \approx 0$).

7 Experimental Results

Implementation Details: We implement the model in PyTorch and train on 4 NVIDIA RTX 3090 GPUs with random cropping and horizontal flipping. Adam [11] is used in three stages: SCI backbone (300 epochs, LR 1×10^{-3}), demosaicing branch (100 epochs, LR 1×10^{-3}), and end-to-end fine-tuning (20 epochs, LR 1×10^{-5}). We report PSNR and SSIM [44].

7.1 Results on Simulated Testing Datasets

We evaluate all methods on the simulated testing set under different compression ratios in Table 1 and Figures 6 and 7. Our method consistently achieves the best performance across all compression ratios. At $Cr = 8$, our method obtains 42.122 dB in S_0 PSNR, clearly outperforming the strongest baseline

Polanalyser (36.384 dB), and also achieves the best DoP PSNR (31.327 dB), DoP SSIM (0.874), and AoP MAE (7.279°). As the compression ratio increases, the reconstruction problem becomes more challenging, but our method degrades gracefully. Specifically, our S_0 PSNR changes from 42.122 dB at $Cr = 8$ to 37.190 dB, 34.607 dB, and 33.507 dB at $Cr = 16$, 32, and 48, respectively, remaining higher than all competing methods. For polarimetric fidelity, our method also maintains the lowest AoP MAE across all settings, i.e., 7.279° , 8.164° , 9.466° , and 10.213° from $Cr = 8$ to $Cr = 48$. These results demonstrate that the proposed method preserves both intensity structures and polarization cues more robustly under increasingly severe temporal compression.

Table 1: Quantitative results under different compression ratios on the simulated testing dataset. The evaluation volume is reported as $H \times W \times T \times N$. Best scores in each group are highlighted in bold.

Cr	Method	Eval. Vol.	PSNR $_{S_0} \uparrow$	SSIM $_{S_0} \uparrow$	PSNR $_{DoP} \uparrow$	SSIM $_{DoP} \uparrow$	MAE $_{AoP} \downarrow$
		$(H \times W \times T \times N)$					
$Cr = 8$	SCI Interp.	$512 \times 512 \times 48 \times 10$	36.361	0.962	31.103	0.853	7.804
	Polanalyser	$512 \times 512 \times 48 \times 10$	36.384	0.964	25.382	0.684	16.955
	PIDSR	$512 \times 512 \times 48 \times 10$	35.886	0.964	27.460	0.737	14.311
	PUGDiff	$512 \times 512 \times 48 \times 10$	35.976	0.965	26.261	0.709	17.023
	SwinCPD	$512 \times 512 \times 48 \times 10$	35.923	0.961	29.063	0.807	9.537
	Ours	$512 \times 512 \times 48 \times 10$	42.122	0.987	31.327	0.874	7.279
$Cr = 16$	SCI Interp.	$512 \times 512 \times 48 \times 10$	34.228	0.942	28.651	0.800	9.805
	Polanalyser	$512 \times 512 \times 48 \times 10$	34.526	0.945	25.711	0.691	16.554
	PIDSR	$512 \times 512 \times 48 \times 10$	34.160	0.944	27.208	0.733	14.873
	PUGDiff	$512 \times 512 \times 48 \times 10$	34.208	0.946	26.467	0.712	16.424
	SwinCPD	$512 \times 512 \times 48 \times 10$	34.146	0.942	28.396	0.796	10.245
	Ours	$512 \times 512 \times 48 \times 10$	37.190	0.963	29.587	0.840	8.164
$Cr = 32$	SCI Interp.	$512 \times 512 \times 48 \times 10$	32.807	0.925	27.628	0.780	11.033
	Polanalyser	$512 \times 512 \times 48 \times 10$	33.034	0.928	25.692	0.689	16.734
	PIDSR	$512 \times 512 \times 48 \times 10$	32.767	0.927	26.836	0.725	15.178
	PUGDiff	$512 \times 512 \times 48 \times 10$	32.789	0.928	26.310	0.705	16.628
	SwinCPD	$512 \times 512 \times 48 \times 10$	32.711	0.925	27.305	0.778	11.320
	Ours	$512 \times 512 \times 48 \times 10$	34.607	0.942	28.422	0.814	9.466
$Cr = 48$	SCI Interp.	$512 \times 512 \times 48 \times 10$	32.200	0.919	27.242	0.769	11.436
	Polanalyser	$512 \times 512 \times 48 \times 10$	32.354	0.922	25.456	0.685	16.596
	PIDSR	$512 \times 512 \times 48 \times 10$	32.109	0.921	26.593	0.721	15.207
	PUGDiff	$512 \times 512 \times 48 \times 10$	32.131	0.922	26.071	0.700	16.516
	SwinCPD	$512 \times 512 \times 48 \times 10$	32.131	0.920	27.082	0.771	11.647
	Ours	$512 \times 512 \times 48 \times 10$	33.507	0.936	27.634	0.796	10.213

7.2 Results on Real Testing Datasets

Figure 8 presents evaluations on real ‘Panda’ and ‘Ball’ scenes, acquired as 512×512 polarized raw measurements. From these compressed observations, our method recovers polarized video sequences with enhanced clarity, exhibiting sharper DoP/AoP boundaries and reduced artifacts. Fine structural details—including panda paw contours and the black rectangular patches on the ball—are notably better preserved. Our approach also demonstrates improved temporal coherence compared to alternative methods, with visibly diminished DoP flickering across consecutive video frames.

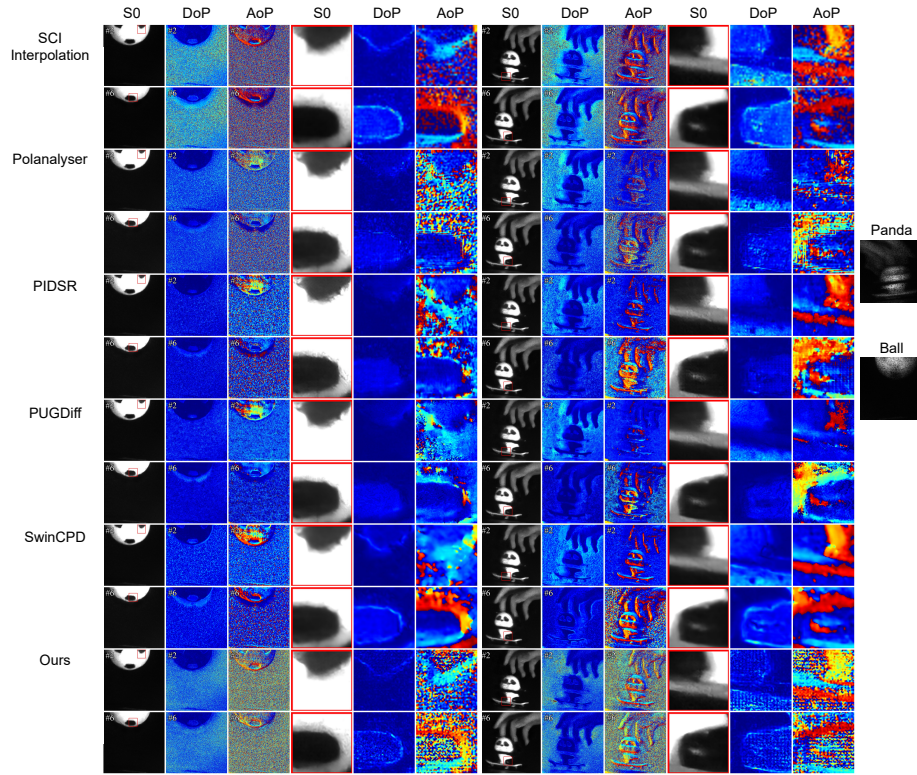


Fig. 8: Qualitative comparison of real captured testing scenes.

Table 2: Ablation Studies. We evaluate the effectiveness of our proposed framework on the Pol-CACTI dataset. (a) **Training Strategy:** Impact of different training stages. (b) **Component Effectiveness:** Ablation on network architecture and physics-constrained loss functions. Best results are in bold. Detailed runtime and memory usage are available in the supplementary material.

(a) Training Strategies				(b) Component Effectiveness			
Setting	S_0 PSNR	DoP PSNR	AoP MAE	Variant	S_0 PSNR	DoP PSNR	AoP MAE
Serial Cascade	15.78	16.43	32.94°	w/o PA-LBP	30.26	21.55	39.09°
Joint Training	33.16	24.18	30.26°	Heavy ResNet	32.37	29.51	11.81°
Stage 2 (Frozen)	35.91	29.89	14.48°	w/o Phys. Loss	39.47	21.99	14.50°
Stage 3 (Ours)	39.79	30.64	9.33°	Ours (Full)	39.79	30.64	9.33°

7.3 Ablation Studies

To validate the effectiveness of our proposed framework, we conducted comprehensive ablation studies summarized in Table 2.

Impact of Training Strategy. As shown in Table 2(a), the naive “Serial Cascade” approach yields poor results (15.78 dB in S_0 PSNR) due to severe error propagation from the SCI reconstruction. Training the entire network from scratch (“Joint Training”) improves performance to 33.16 dB but suffers from

convergence difficulties. In contrast, our progressive strategy serves as a stable anchor: the “Stage 2” baseline (frozen backbone) reaches 35.91 dB, while the final joint fine-tuning (“Stage 3”) significantly boosts S_0 PSNR to **39.79 dB** and reduces AoP error by over 5° (from 14.48° to **9.33°**), confirming the necessity of end-to-end holistic optimization.

Component Effectiveness. Table 2(b) analyzes the contribution of specific modules. Removing the PA-LBP module (“w/o PA-LBP”) causes a drastic drop in S_0 PSNR to 30.26 dB, proving that the learnable back-projection is critical for detail recovery. Simply increasing the depth of the reconstruction network (“Heavy ResNet”) only achieves 32.37 dB, demonstrating that the performance gain stems from the physics-aware projection design rather than mere parameter scaling. notably, training without physics-constrained losses (“w/o Phys. Loss”) maintains high radiometric quality (39.47 dB S_0 PSNR) but fails to recover accurate polarization signatures, causing DoP PSNR to collapse to 21.99 dB. Our full model achieves the best performance across all metrics, ensuring both radiometric fidelity and polarimetric accuracy.

8 Conclusion

In this paper, we present Pol-CACTI, an end-to-end framework for high-speed polarized video snapshot compressive imaging. By coupling DoFP-based polarimetric sampling with temporal compression in a unified forward model, and solving the resulting inverse problem with a physics-driven joint reconstruction network, Pol-CACTI alleviates the aliasing and error accumulation typically observed in two-step pipelines. Central to our approach is the Polarization-Aware Learnable Back-Projection (PA-LBP) module, which injects measurement consistency during reconstruction, together with physics-constrained losses that promote faithful recovery of polarization cues. Experiments on both our curated dynamic polarization dataset and real captured scenes demonstrate improved radiometric fidelity and more stable polarimetric estimates, yielding notable PSNR gains and sub- 10° AoP accuracy. We will release the Pol-CACTI dataset and related resources to support reproducible research and facilitate broader adoption in dynamic polarization imaging.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant number U2541205, 62271414), National Key R&D Program of China (2024YFF0505603), “Pioneer” and “Leading Goose” R&D Program of Zhejiang (grant number 2024SDXHDX0006, 2024C03182), the 2023 International Sci-tech Cooperation Projects under the purview of the “Innovation Yongjiang 2035” Key R&D Program (grant number 2024Z126), the National Natural Science Foundation of China (grant number 62105372) and Hunan Provincial Research and Development Project (grant number 2025QK3019).

References

1. Atkinson, G.A., Hancock, E.R.: Recovery of surface orientation from diffuse polarization. *IEEE Trans. Image Process.* **15**(6), 1653–1664 (2006)
2. Ba, Y., Gilbert, A.R., Wang, F., Yang, J., Chen, R., Wang, Y., Yan, L., Shi, B., Kadambi, A.: Deep shape from polarization. In: *Eur. Conf. Comput. Vis.* Springer (2020)
3. Cheng, Z., Chen, B., Liu, G., Zhang, H., Lu, R., Wang, Z., Yuan, X.: Memory-efficient network for large-scale video compressive sensing. In: *IEEE Conf. Comput. Vis. Pattern Recog.* pp. 16246–16255 (2021)
4. Cheng, Z., Chen, B., Lu, R., Wang, Z., Zhang, H., Meng, Z., Yuan, X.: Recurrent neural networks for snapshot compressive imaging. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(2), 2264–2281 (2022)
5. Cheng, Z., Lu, R., Wang, Z., Zhang, H., Chen, B., Meng, Z., Yuan, X.: BIRNAT: Bidirectional recurrent neural networks with adversarial training for video snapshot compressive imaging. In: *Eur. Conf. Comput. Vis.* (2020)
6. Gao, S., Gruev, V.: Bilinear and bicubic interpolation methods for division of focal plane polarimeters. *Optics Express* **19**(27), 26161–26173 (2011)
7. Hardie, R.C., Barnard, K.J., Armstrong, E.E.: Super-resolution for imagery from integrated microgrid polarimeters. *Optics Express* **19**(13), 12937–12960 (2011)
8. Hitomi, Y., Gu, J., Gupta, M., Mitsunaga, T., Nayar, S.K.: Video from a single coded exposure photograph using a learned over-complete dictionary. In: *Int. Conf. Comput. Vis.* pp. 287–294. IEEE (2011)
9. Jiang, X., Wang, L., Wang, P., Yuan, X.: Detectsci: Toward object-guided roi reconstruction for high-resolution video snapshot compressive imaging. In: *IEEE Conf. Comput. Vis. Pattern Recog.* pp. 41720–41729 (2026)
10. Kadambi, A., Taamazyan, V., Shi, B., Raskar, R.: Polarized 3D: High-Quality Depth Sensing with Polarization Cues. In: *Int. Conf. Comput. Vis.* pp. 3370–3378 (2015)
11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
12. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(11), 2599–2613 (2019)
13. Li, C., Luo, Y., Zhang, J., Yang, D.: Polarization uncertainty-guided diffusion model for color polarization image demosaicking. *AAAI* **40**(8), 6028–6036 (2026)
14. Lincetto, F., Agresti, G., Rossi, M., Zanuttigh, P.: Multimodalstudio: A heterogeneous sensor dataset and framework for neural rendering across multiple imaging modalities. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2025)
15. Liu, Y., Yuan, X., Suo, J., Brady, D.J., Dai, Q.: Rank minimization for snapshot compressive imaging. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(12), 2990–3006 (2019)
16. Llull, P., Liao, X., Yuan, X., Yang, J., Kittle, D., Carin, L., Sapiro, G., Brady, D.J.: Coded aperture compressive temporal imaging. *Optics Express* **21**(9), 10526–10545 (2013)
17. Luo, Y., Zhang, J., Shao, J., Tian, J., Ma, J.: Learning a non-locally regularized convolutional sparse representation for joint chromatic and polarimetric demosaicking. *IEEE Trans. Image Process.* **33**, 5029–5044 (2024)
18. Ma, J., Liu, X.Y., Shou, Z., Yuan, X.: Deep tensor admm-net for snapshot compressive imaging. In: *Int. Conf. Comput. Vis.* pp. 10223–10232 (2019)

19. Maeda, R.: Polanalyser: Polarization image analysis tool (2019), <https://github.com/elerac/polanalyser>
20. Maeda, R., Moon, Y., Baek, S.H.: Event ellipsometer: Event-based mueller-matrix video imaging. In: IEEE Conf. Comput. Vis. Pattern Recog. (2025)
21. McWilliam, A., Al Khafaji, M.A., Svensson, S.J., Pádua, S., Franke-Arnold, S.: Dynamic mueller matrix polarimetry using generalized measurements. *Optics Express* **32**(12), 21909–21924 (2024)
22. Meng, Z., Yuan, X., Jalali, S.: Deep unfolding for snapshot compressive imaging. *Int. J. Comput. Vis.* **131**(11), 2933–2958 (2023)
23. Mihoubi, S., Lapray, P.J., Bigué, L.: Survey of demosaicking methods for polarization filter array images. *Sensors* **18**(11), 3688 (2018)
24. Miyazaki, D., Shigetomi, T., Baba, M., Furukawa, R., Hiura, S., Asada, N.: Surface normal estimation of black specular objects from multiview polarization images. *Optical Engineering* **56**(4), 041303 (2017)
25. Morimatsu, M., Monno, Y., Tanaka, M., Okutomi, M.: Monochrome and color polarization demosaicking based on intensity-guided residual interpolation. *IEEE Sensors Journal* **21**(23), 26985–26996 (2021)
26. Qiao, M., Meng, Z., Ma, J., Yuan, X.: Deep learning for video compressive sensing. *APL Photonics* **5**(3) (2020)
27. Qiu, S., Fu, Q., Wang, C., Heidrich, W.: Linear polarization demosaicking for monochrome and color polarization filter array images. *Computer Graphics Forum* **40**, 77–89 (2021)
28. Rahman, M.D.A.B.A., Monno, Y., Tanaka, M., Okutomi, M.: Polarization denoising and demosaicking: Dataset and baseline method. In: IEEE Int. Conf. Image Process. pp. 2524–2529 (2025)
29. Ratliff, B.M., LaCasse, C.F., Tyo, J.S.: Interpolation strategies for reducing instantaneous-field-of-view artifacts in microgrid polarimeter imagery. In: In Proceedings of the SPIE Defense, Security, and Sensing Conference. vol. 7336, p. 733605 (2009)
30. Reddy, D., Veeraraghavan, A., Chellappa, R.: P2c2: Programmable pixel compressive camera for high speed imaging. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 329–336 (2011)
31. Sargent, I., York, T., Wilson, D.: A conditional generative adversarial network demosaicing strategy for division of focal plane polarimeters. *Optics Express* **28**(25), 38419–38433 (2020)
32. Schechner, Y.Y., Narasimhan, S.G., Nayar, S.K.: Polarization-based vision through haze. *Applied Optics* **42**(3), 511–525 (2003)
33. Song, Y., Liu, M., Yuan, X.: Deep learning based video snapshot compressive imaging of leidenfrost droplets. *iOptics* p. 100020 (2025)
34. Su, F., et al.: A brief review of the photoelasticity based on a pixelated polarization camera: principle, advantages/disadvantages and typical applications. *Journal of Mechanics* **40**, 698–710 (2024)
35. Tyo, J.S., Goldstein, D.L., Chenault, D.B., Shaw, J.A.: Review of passive imaging polarimetry for remote sensing applications. *Applied Optics* **45**(22), 5453–5469 (2006)
36. Wang, L., Cao, M., Yuan, X.: Efficientsci: Densely connected network with space-time factorization for large-scale video snapshot compressive imaging. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 18477–18486 (2023)
37. Wang, L., Cao, M., Zhong, Y., Yuan, X.: Spatial-temporal transformer for video snapshot compressive imaging. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(7), 9072–9089 (2022)

38. Wang, P., Wang, L., Qiao, M., Yuan, X.: Full-resolution and full-dynamic-range coded aperture compressive temporal imaging. *Optics Letters* **48**(18), 4813–4816 (2023)
39. Wang, P., Wang, L., Yuan, X.: Deep optics for video snapshot compressive imaging. In: *Int. Conf. Comput. Vis.* pp. 10646–10656 (2023)
40. Wang, P., Zhang, Y., Wang, L., Yuan, X.: Hierarchical separable video transformer for snapshot compressive imaging. In: *Eur. Conf. Comput. Vis.* pp. 104–122. Springer (2024)
41. Wang, T., Zhang, X., Li, J., Zheng, Y., Ru, J., Xu, H.: SwinCPD: A Swin Transformer-Based Network for Color-Polarization Demosaicing. *IEEE Sensors Journal* **26**(1), 637–651 (2026)
42. Wang, X., Zhang, Z., Gao, J.: Polarization-based camouflaged object detection. *Pattern Recognition Letters* **174**, 106–111 (2023)
43. Wang, Z., Zhang, H., Cheng, Z., Chen, B., Yuan, X.: Metasci: Scalable and adaptive reconstruction for video compressive sensing. In: *IEEE Conf. Comput. Vis. Pattern Recog.* pp. 2083–2092 (2021)
44. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
45. Wolff, L.B.: Polarization-based material classification from specular reflection. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(11), 1059–1071 (1990)
46. Yang, C., Zhang, S., Yuan, X.: Ensemble learning priors driven deep unfolding for scalable video snapshot compressive imaging. In: *Eur. Conf. Comput. Vis.* pp. 600–618 (2022)
47. Yuan, X., Liu, Y., Suo, J., Dai, Q.: Plug-and-play algorithms for large-scale snapshot compressive imaging. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2020)
48. Yuan, X.: Generalized alternating projection based total variation minimization for compressive sensing. In: *IEEE Int. Conf. Image Process.* pp. 2539–2543 (2016)
49. Yuan, X., Brady, D.J., Katsaggelos, A.K.: Snapshot compressive imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine* **38**(2), 65–88 (2021)
50. Yuan, X., Liu, Y., Suo, J., Durand, F., Dai, Q.: Plug-and-play algorithms for video snapshot compressive imaging. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(10), 7093–7111 (2021)
51. Zhang, Z., Zheng, S., Qiu, M., Situ, G., Brady, D.J., Dai, Q., Suo, J., Yuan, X.: A decade review of video compressive sensing: a roadmap to practical applications. *Engineering* **46**, 172–185 (2025)
52. Zheng, S., Yuan, X.: Unfolding framework with prior of convolution-transformer mixture and uncertainty estimation for video snapshot compressive imaging. In: *Int. Conf. Comput. Vis.* pp. 12738–12749 (2023)
53. Zhou, S., Zhou, C., Lyu, Y., Guo, H., Ma, Z., Shi, B., Sato, I.: PIDSr: Complementary polarized image demosaicing and super-resolution. In: *IEEE Conf. Comput. Vis. Pattern Recog.* pp. 16081–16090 (2025)