

Learning to Remove Wrinkled Transparent Film with Polarized Prior

Jiaqi Tang^{1,2,3} Ruizheng Wu⁴ Xiaogang Xu^{5,6} Sixing Hu⁴ Ying-Cong Chen^{1,2,3*}

¹The Hong Kong University of Science and Technology (Guangzhou)

²The Hong Kong University of Science and Technology ³HKUST(GZ) – SmartMore Joint Lab

⁴SmartMore Corporation ⁵The Chinese University of Hong Kong ⁶Zhejiang University

jtang092@connect.hkust-gz.edu.cn, {ruizheng.wu, david.hu}@smartmore.com

xiaogangxu00@gmail.com, yingcongchen@ust.hk

Abstract

In this paper, we study a new problem, Film Removal (FR), which attempts to remove the interference of wrinkled transparent films and reconstruct the original information under films for industrial recognition systems. We first physically model the imaging of industrial materials covered by the film. Considering the specular highlight from the film can be effectively recorded by the polarized camera, we build a practical dataset with polarization information containing paired data with and without transparent film. We aim to remove interference from the film (specular highlights and other degradations) with an end-to-end framework. To locate the specular highlight, we use an angle estimation network to optimize the polarization angle with the minimized specular highlight. The image with minimized specular highlight is set as a prior for supporting the reconstruction network. Based on the prior and the polarized images, the reconstruction network can decouple all degradations from the film. Extensive experiments show that our framework achieves SOTA performance in both image reconstruction and industrial downstream tasks. Our code will be released at <https://github.com/jqtangust/FilmRemoval>.

1. Introduction

Various deep-learning-based recognition models have been employed in the industrial environment, e.g., defect detection [28], code recognition [25], etc. However, the model failures would sometimes happen due to the insufficient robustness [40] towards different perturbations [8, 10, 24, 32, 39]. The wrinkled transparent film is one of them, which is usually covered or packaged on industrial materials or products for protection. Its interference could cause the failure of varying downstream tasks, e.g., text OCR and

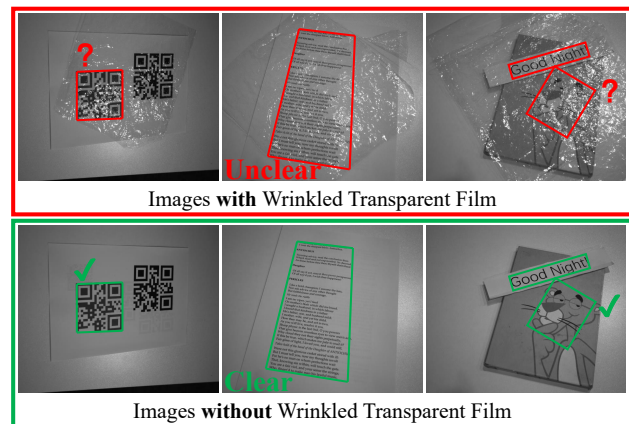


Figure 1. The **Red** box presents a challenge in industrial recognition systems, where the product information is often hidden beneath the wrinkled transparent film. The **Green** box is the image we expect to generate, with the film layer removed. Removing the wrinkled film makes the information on industrial material clearer. QR code recognition, in Fig. 1. Regarding the wide usage of such films in industrial scenarios, it is worth opening the research direction to remove these films from images.

For the first time, we address a novel problem of wrinkled transparent Film Removal (FR), which aims to remove the transparent film and reveal the hidden information, benefiting the robustness of the industrial downstream models.

Although some solutions have attempted to remove surface highlight [6, 31, 34], they have not accurately modeled the imaging of the wrinkled transparent film. Except for the highlight, they cannot remove the effects of other various degradations from the film, e.g., light transmittance and material texture. Therefore, they cannot remove the transparent film thoroughly.

In this paper, we explicitly model the imaging of wrinkled transparent film into two parts: the specular highlight I_h from the film, and the diffuse reflection I_{md} from the materials under the film, as shown in Fig. 2. The diffuse reflection can be influenced by the properties of the film. Thus, it is divided into the original component (I_m) and various

*Corresponding author.

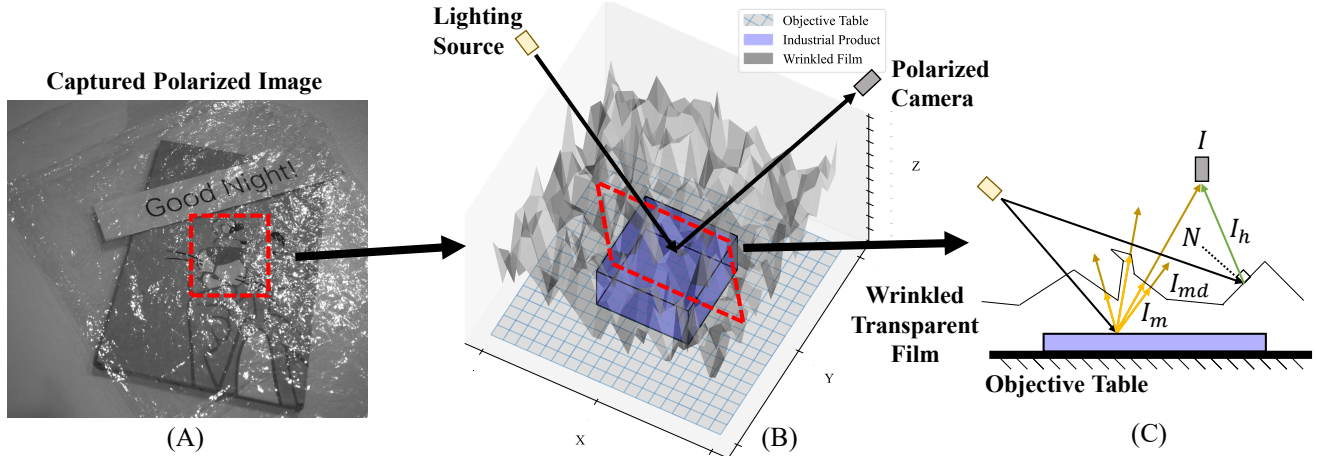


Figure 2. Wrinkled Transparent Film Model. (A) The polarized image. (B) The 3D physics model of the local region. The light is reflected through the transparent wrinkled film and captured by the polarization camera. (C) The light path diagram. Polarized cameras capture two components: **Specular Reflection** (I_h), and **Diffuse Reflection** (I_{md}). The **Original Diffuse Reflection** (I_m) would be interfered by various degradations ($I_{md} - I_m$) from the film.

degradations ($I_{md} - I_m$). Therefore, our objective is to decouple the specular highlight and other degradations caused by the film and reconstruct the original diffuse reflection.

We build an end-to-end framework for decoupling two different degradations in the Wrinkled Transparent Film Model (Fig. 2), which consists of the prior estimation and reconstruction network. We notice that specular highlight is significantly related to polarization angles, while other parts are not. Based on this observation, we use an Angle Estimation Network to learn the polar angle corresponding to the minimized specular highlight, which is driven by a Polarization-based Location Model. Images with minimized specular highlights are set as the priors for the later reconstruction network to remove all degradations.

There is currently no suitable dataset for the FR problem, since most of the existing datasets [5, 18, 34] are targeted towards specular reflection removal only. Also, most datasets do not consider the real industrial environment. Therefore, we build a new dataset that consists of paired images covered by the film and the uncovered ground truth in the industrial optical photography system. Moreover, it's proved that the specular highlight from the film can be effectively captured by the polarization camera [21, 34, 37], which is now cheap to be installed in industrial systems [22]. Thus, we follow the collection pipeline of existing polarization datasets [4, 9], and capture each object with four polarized images under four polar directions in one shot. It's empirically proved by our experiments that with the input of these polarization clues, networks can better locate the specular highlight and recover the hidden information under the film.

Extensive experiments prove that our designed network achieves SOTA performance in the FR problem. Our contributions are summarized as follows:

- To the best of our knowledge, we are the first to address the new problem of Film Removal (FR), which

aims to remove the whole wrinkled transparent film in industrial scenarios.

- To solve FR, we model the wrinkled film physically and propose an end-to-end reconstruction network for FR with a learnable polarization-based prior, which helps the network locate the specular highlight reflection in the film.
- We also build a new practical dataset in the real industrial optical photography system, which contains various polarized image pairs with and without the film.
- Extensive experiments are conducted to prove the effectiveness of our dataset and method, which achieves SOTA performance in both image reconstruction and downstream industrial tasks.

2. Related Work

2.1. Polarization Model and Application

Polarization refers to the property of the transverse wave oscillating in different directions. Since light is a kind of wave, this phenomenon describes the distribution of light waves in all directions [1]. Conventional cameras or human eyes are insensitive to polarization. Thus, polarization is often used as a way to supplement additional visual information. It is often used in a wide range of fields such as optics [29], materials science [27], and physics [7, 36], etc.

In the field of computer vision, polarization provides different angles of view, hence allowing a more efficient interpretation of complex scenes. In recent years, polarization has been widely used for complex tasks such as Integral Imaging [38], Rendering [2], 3D Shape [44], Segmentation [14], and Reflection Removal [12, 15, 34], etc.

In the real world, since natural light is mixed with multiple wavelengths, its refraction time is different when it is injected into an optically active material, thus, a phase shift

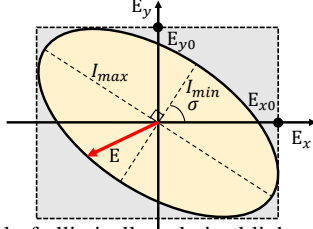


Figure 3. Model of elliptically polarized light. E represents polarized light at any angle, which can be calculated by this model. I_{max} and I_{min} are two components indicating the maximum and minimum intensity of elliptically polarized light.

occurs. A physical description of this phenomenon is elliptically polarized light (*i.e.*, partially polarized light), as shown in Eq. (1).

$$\begin{cases} E_x = E_{x0} \cos(\omega t) \\ E_y = E_{y0} \cos(\omega t - \sigma) \end{cases}, \quad (1)$$

where (x, y) is the Cartesian basis in the space of Jones vectors, E_x and E_y is the component of the light on the basis, (x, y) . ω is the frequency and σ is the phase difference of E_x and E_y . E_{x0} and E_{y0} are the field strength of a pair of orthogonal waves. Eq. (1) describes the vibration of the polarized light in different angles, t . Based on this model, our solution only considers a simplified situation where $\sigma = 90^\circ$. Eq. (1) can be simplified to Eq. (2),

$$\frac{E_x^2}{E_{x0}^2} + \frac{E_y^2}{E_{y0}^2} = 1. \quad (2)$$

2.2. Specular Highlight Removal via Polarization

Since the film includes unpredictable specular reflection, our task includes the procedure to remove these degradations. There have been several solutions to remove specular reflection using polarization information. Nayar et al. [21] first used polarization to determine the color of the specular component to separate the interfaces. Then, Umeyama et al. [30] adopted independent component analysis to separate the diffuse and specular reflection components of surface reflection. Zhang et al. [42] considered the effect of polarization angle and attempted to get the appropriate global angle using Newton’s method, but this method did not make full use of the local information. Wen et al. [34] separated specular reflection regions by using image chromaticity.

Although conventional methods are available for removing specular highlight reflections, they are not able to accurately model the imaging of wrinkled transparent films, and thus do not adequately address the problem of eliminating all degradations from wrinkled transparent films.

3. Dataset

While some datasets exist for the removal of specular reflection [11, 13, 34], there is currently no dataset for film removal in the industrial environment. As shown in Fig. 2, the characteristics of the specular highlight information from the film can be effectively modeled by the polarized image.

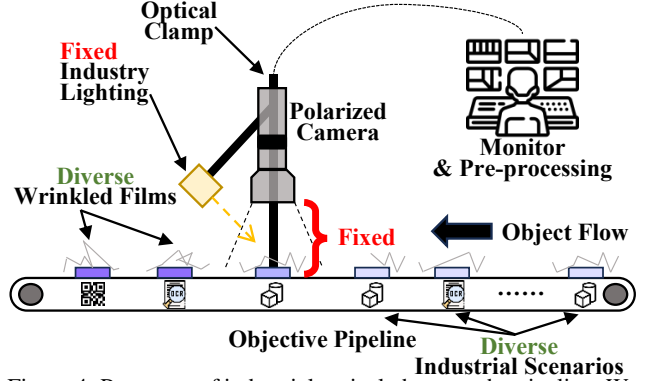


Figure 4. Prototype of industrial optical photography pipeline. We have built an optical pipeline for capturing the dataset in the industrial environment. As objects traverse the objective pipeline, the polarizing camera captures images continuously. Subsequently, the acquired data is sent to the monitor for pre-processing.

Leveraging this polarization information can significantly enhance image reconstruction. Therefore, we construct a paired dataset based on polarized images. Polarized images can capture light amplitude information from wrinkled films at different angles, and making full use of this information will significantly facilitate the decoupling of degradation information occurring from the film.

3.1. Industrial Optical Photography Pipeline

Fig. 4 illustrates the prototype of our pipeline. Within this industrial pipeline, we maintain a consistent posture and angle for both the camera and lighting resources. The objective pipeline sequentially passes different detected objects under the camera.

To effectively capture optical information in multiple polarization directions within a single image, we employ the HIKROBOT MV-CH050-10UP camera¹, which integrates the Sony IMX250MZR CMOS sensor². This sensor contains four different angles of polarization (0° , 45° , 90° , and 135°) as Bayer pattern, allowing us to capture four polarization angles in high definition with a single shot.

Finally, there is a monitor that controls the camera shutter and adjusts camera parameters. We utilize auto-exposure and auto-focus strategies to set the appropriate focal length, exposure time, and ISO for each specific scenario automatically. This is essential since the thickness and surface of the products on the industrial line may vary, requiring slight adjustments in camera parameters to ensure image quality.

3.2. Capturing Polarized Images

When capturing polarized images, the imaging system initially captures an image of the ground truth uncovered by the film, I_{gt}^{raw} . Subsequently, with the complex transpar-

¹<https://www.hikrobotics.com/cn/machinevision/productdetail?id=3886>

²<https://www.sony-semicon.com/en/products/is/industry/polarization.html>

ent film placed over the ground truth, we capture the image I_{input}^{raw} that needs to be recovered.

3.3. Data Diversity and Robustness

To build a diverse and robust dataset, we follow the rules in industrial manufacturing. As depicted in Fig. 4, the current industrial pipeline includes 315 dynamic industrial scenarios, which can be categorized into three types: QR codes, text, and products. To enhance the diversity, we have different films with diverse material properties, coverage areas, film thicknesses, and levels of wrinkling. The film exhibits significant variability across each scenario.

On the other hand, to ensure the stability of the industrial imaging pipeline, we maintained a consistent intensity level for the industrial light source and fixed the distance between the camera and the object flow. This helps to minimize the influence of errors external to the industrial system.

3.4. Preprocessing

Each pixel of the image captured by the polarization sensor is represented by four-pixel dots, each corresponding to the intensity of light at four distinct polarization angles. We first need to decompose it into 4 subgraphs with different angles, and then we restore it to its original resolution using edge-aware residual interpolation (EARL) demosaicking [20]. This process is described in Eq. (3).

$$\{I_{gt}^0, I_{gt}^{45}, I_{gt}^{90}, I_{gt}^{135}\} = M(F_d(I_{gt}^{raw})), \quad (3)$$

$\{I_{input}^0, I_{input}^{45}, I_{input}^{90}, I_{input}^{135}\} = M(F_d(I_{input}^{raw}))$, where $F_d(\cdot)$ is the decomposing operator, $M(\cdot)$ is EARL demosaicking. Taking I_{gt}^{raw} and I_{input}^{raw} as inputs, we obtain full-resolution polarized images with four different angles.

To generate the ground truth image, we follow the standard procedure from the polarized image processing library, polanalyser³. Firstly, we introduce stokes parameters [3]. According to this physical model, the first stoke parameter S_0 describes the total intensity of the optical beam, and it can be calculated by Eq. (4).

$$S_0 = E_{x0}^2 + E_{y0}^2 = I^x + I^y, \quad (4)$$

where $I^x \perp I^y$. E_{x0}^2 and E_{y0}^2 are the field strength of a pair of orthogonal waves in Fig. 3, which can be calculated by a pair of orthogonal polarization components, I^x and I^y . Subsequently, we can calculate one ground truth I_{gt} as

$$I_{gt} = G\left(\frac{S_0}{2}\right) = \left(\frac{I_{gt}^0 + I_{gt}^{45} + I_{gt}^{90} + I_{gt}^{135}}{4}\right)^{\frac{1}{\gamma}}, \quad (5)$$

where $G(\cdot)$ is a gamma correction function, and we empirically set the gamma value, γ , to 2.2.

3.5. Training and Testing

During training, our network mixed all the scenes for training, and the final network is applicable to data from all scenarios. Besides, to ensure the robustness and generalization of our dataset, we adopt 10-fold cross-validation [23]

³<https://github.com/elerac/polanalyser/wiki>

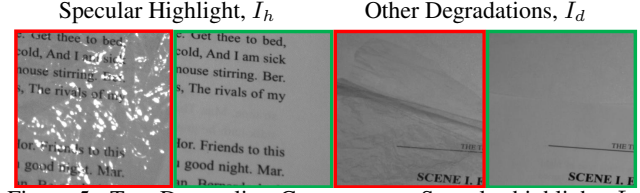


Figure 5. Two Decoupling Components. Specular highlight, I_h and other degradations, I_d . The Red box shows the degradations, the Green box is the Ground Truth.

to evaluate the results. The dataset is divided into ten parts, and nine of them are used as training data and one as test data in turn. Each test will yield a corresponding accuracy rate, which is then averaged as the final accuracy.

4. Method

Although some conventional methods have used polarization information to remove surface specular highlight reflection, they assume the light intensity as a *binary* composition [34], i.e., transmission and reflection. However, the light intensity from the wrinkled film is more complex, which is not only influenced by film surface highlight but also mixed with various degradations, e.g., light transmittance and material texture.

In this section, we first model the wrinkled transparent film physically in Fig. 2, which is our method’s motivation. Then, we use an end-to-end network for reconstructing the original information in Fig. 6. Besides, since the highlight region is more tricky to recover, we build a polarization-based prior into the end-to-end framework to assist in locating highlight regions.

4.1. Modelling the Wrinkled Transparent Film

Based on the Industrial Optical Photography System in Sec. 3.1, the light intensity for materials covered with the wrinkled transparent film will be captured by the polarized camera in Fig. 2(A)(B). The captured light intensity consists of two parts, i.e., the specular highlight reflection of the film’s polarized regions [19, 26], the diffuse reflection from the materials under the film, which could be influenced by light transmittance, the texture of the film in Fig. 2(C). Such a composition is written in Eq. (6).

$$I = I_{md} + I_h = I_m + I_d + I_h, \quad (6)$$

where $+$ denotes the linear superposition of different light components, I is the light intensity captured by the camera, I_{md} is the diffuse reflection component of the material through various film degradations, I_h is the specular reflection part from the film’s highlighted regions. Then, I_{md} can be decoupled to two parts, I_m and I_d . I_m is the original diffuse reflection component of the material and I_d is caused by other various degradations through film. In Fig. 2(C), the optical path diagram illustrates this process.

Based on Eq. (6), the FR task can be implemented by retaining the information of I_m and decoupling I_h and I_d (both of them are caused by the film layer). This is expressed in Eq. (7), as

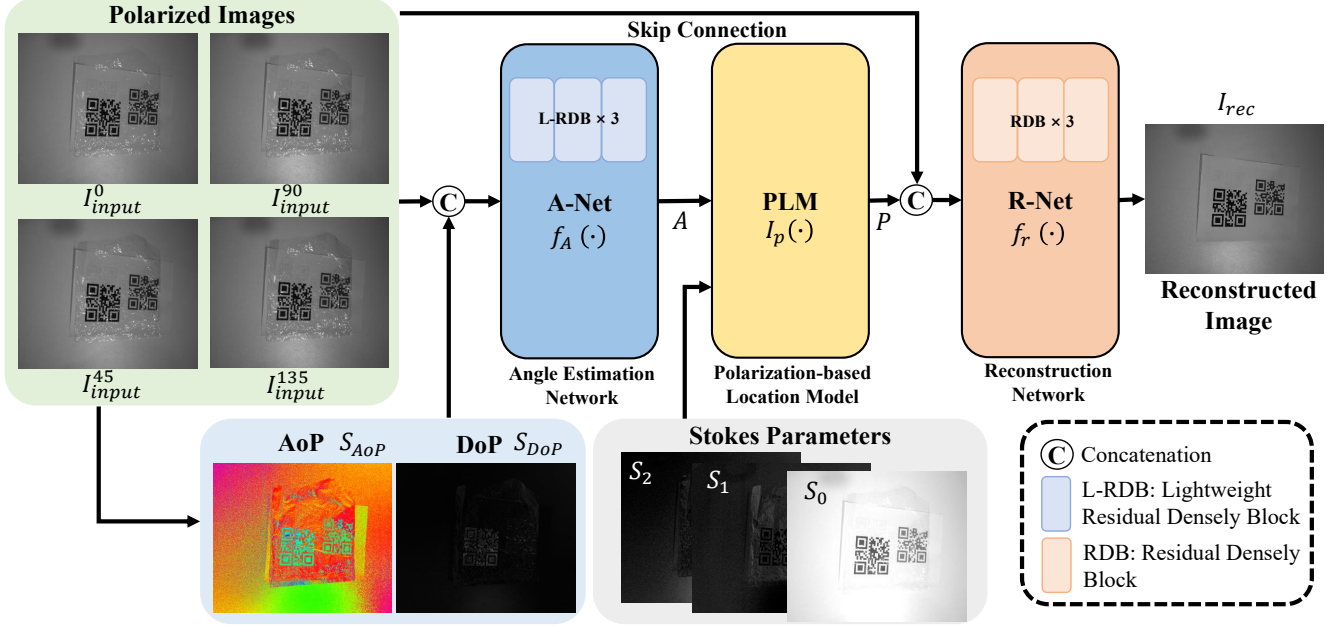


Figure 6. Overall framework. The polarized images, AoP and DoP are fed into Angle Estimation Net (A-Net), denoted as f_A , for estimating the angle, A . Subsequently, the Polarization-based Location Model (PLM), represented as I_p , takes A as input to estimate the image prior, P . This prior provides important highlight location information for the reconstruction network. Finally, the reconstruction network (R-Net) is trained to restore the original diffuse reflection in industrial materials.

$$I_m = I - I_h - I_d, \quad (7)$$

where $-$ denotes the decoupling operator. Fig. 5 visualizes these two components for decoupling in our FR task.

Based on this model, our whole framework is based on an end-to-end reconstruction network for decoupling these two parts in Sec. 4.3. Before that, to support the network for decoupling I_h , we estimate a polarized prior for locating the highlight regions in Sec. 4.2.

4.2. Estimating a Polarized Prior for Locating I_h

To decouple the specular reflection components, it is better to locate the highlight regions on the surface of the wrinkled film, which provides a prior to facilitate the decoupling network. However, these regions are hard to predict since images are captured in variable scenarios. To this end, we introduce the polarization information for this problem.

Based on Fresnel's theory [35], the specular reflection component of optically active materials is an elliptically polarized light, which changes under different angles of polarization orientation, while the rest of the components remain almost constant. It can be employed in estimating I_h , which is a specular reflection component. Thus, we propose a polarized prior P , which is represented as the optimized I with the minimized I_h , as shown in Eq. (8). The difference between P and the input I indicates the regions of I_h .

$$P = I_m + I_d + \min I_h. \quad (8)$$

Eq. (8), *i.e.*, the polarized version of Eq. (6), can be acquired with Malus's Law [16] and the elliptical polarization model in Sec. 2.1. We use a pair of orthogonal maximum and minimum components and the angle variable

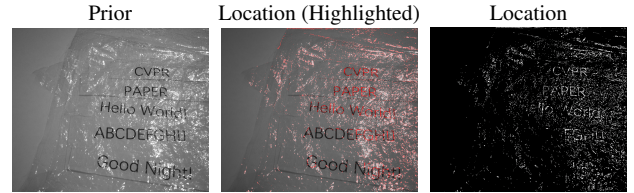


Figure 7. Location of specular highlight. The P is the polarized prior. We can calculate the location of highlight by $I_h - \min I_h$.

$\theta \in [0, 2\pi)$ to rewrite Eq. (8) as

$$I_h = I_p(\theta) = I_{max} \cos^2 \theta + I_{min} \sin^2 \theta, \quad (9)$$

where $I_p(\theta)$ is the Polarization-based Location Model to describe the polarized light in the angle of θ , I_{max} and I_{min} is a pair of orthogonal maximum and minimum components in Fig. 3. Given the input data $\{I_{input}^0, I_{input}^{45}, I_{input}^{90}, I_{input}^{135}\}$, I_{max} and I_{min} can be computed by the Stokes parameters [3], from Eq. (10) and Eq. (11).

$$I_{max} = S_0 + \frac{\sqrt{S_1^2 + S_2^2}}{2}, I_{min} = S_0 - \frac{\sqrt{S_1^2 + S_2^2}}{2}, \quad (10)$$

$$S_0 = E_{x0}^2 + E_{y0}^2 = I_{input}^0 + I_{input}^{90},$$

$$S_1 = E_{x0}^2 - E_{y0}^2 = I_{input}^0 - I_{input}^{90}, \quad (11)$$

$$S_2 = E_{a0}^2 - E_{b0}^2 = I_{input}^{45} - I_{input}^{135},$$

where (a, b) is the Cartesian basis rotated by 45° in the space of (x, y) , and E_{a0}^2 as well as E_{b0}^2 are the field strengths under this basis.

Since I_h is the only polarized component that is determined by θ , P in Eq. (8) can also be formulated as Eq. (12).

$$\begin{aligned} P &= I_m + I_d + \min I_h \\ &= I_m + I_d + \min_{\theta} I_p(\theta) \\ &= I_m + I_d + \min_{\theta} (I_{max} \cos^2 \theta + I_{min} \sin^2 \theta). \end{aligned} \quad (12)$$

Different pixels in one image correspond to varying optimized values for θ . Thus, we estimate pixel-wise θ with a learning-based network, obtaining the angle map A . The input includes images with polarization information, i.e., I_0 , I_{45} , I_{90} , and I_{135} . In addition, our input also includes two essential physics statistics, angle of polarization (AoP) and degree of polarization (DoP) [17]. The AoP provides information about the object’s surface normal, which helps to analyze the difference in the surface structure between the object and film. The DoP offers information on the intensity of polarized light, so it can facilitate the network to utilize the polarized light accurately. Both of them can promote the model to learn the appropriate angle better. Eq. (13) describes this procedure.

$$A = f_A(I_{input}^0 \oplus I_{input}^{45} \oplus I_{input}^{90} \oplus I_{input}^{135} \oplus S_{AoP} \oplus S_{DoP}), \quad (13)$$

where, $A \in \mathbb{R}^{h \times w \times 1}$ is the pixel-wise angle map, $S_{AoP} \in \mathbb{R}^{h \times w \times 1}$ and $S_{DoP} \in \mathbb{R}^{h \times w \times 1}$ indicate the AoP and DoP map respectively, \oplus is the concatenation operator, and $f_A(\cdot)$ is angle estimation network. The network structure employs a lightweight Residual Dense Network [43] with a large perception field to capture more global information.

After obtaining the angle map A , we can get the prior with minimized I_h by Eq. (14).

$$P = I_m + I_d + I_{max} \cos^2 A + I_{min} \sin^2 A, \quad (14)$$

where $P \in \mathbb{R}^{h \times w \times 1}$. Fig. 7 visualize the Prior, P , and the location of I_h (highlighted). We can accurately estimate the location of the specular highlight.

4.3. Reconstructing I_m with a Prior

In this section, we set a reconstruction network f_r to decouple both I_d and I_h , with the input of I (i.e., $I_{input}^0, I_{input}^{45}, I_{input}^{90}$, and I_{input}^{135}) and P . The obtained prior P from Sec. 4.2 can already provide the estimation of I_h by comparing with I . Thus, f_r can be focused on the assessment of I_d . The reconstruction network is implemented as a common Residual Dense Network [43]. The reconstruction process can be expressed as Eq. (15).

$$I_{rec} = f_r(I_{input}^0 \oplus I_{input}^{45} \oplus I_{input}^{90} \oplus I_{input}^{135} \oplus P), \quad (15)$$

where, $I_{rec} \in \mathbb{R}^{h \times w \times 1}$ is the reconstructed image.

4.4. Details in Implementation

Our framework is implemented by PyTorch, and we use Polanalyser library⁴ to process the polarized images. It is trained with $L1$ loss between I_{rec} and I_{gt} in an end-to-end manner. Besides, the learning rate is 5×10^{-5} , and it decays to half of the original for every 2×10^4 iterations.

5. Experiment

In our experiment, we evaluate the performance of the proposed framework in three aspects. First, as an image reconstruction task, we conduct qualitative and quantitative evaluations with reconstruction-related metrics (e.g.,

⁴<https://github.com/elerac/polanalyser/wiki>

PSNR, SSIM). Second, as an upstream task in the industry, two downstream scenarios (e.g., QR code reading, and Text OCR) are selected to evaluate the reliability of our proposed algorithm. Finally, we conduct ablation studies to analyze the roles of different components.

5.1. Baselines

There is no existing baseline for this new problem. Thus, we choose two general SOTA methods in image reconstruction, Uformer [33] and Restormer [41] to evaluate the performance. Considering the connection between the film removal and highlight removal, we also compare with two SOTA polarization and non-polarization highlight removal baselines, Polar-HR [34] and SHIQ [6], respectively.

To make a fair comparison, we follow the same training setting in our framework and input four polarized images to reconstruct one image without wrinkled transparent film in Uformer [33], Restormer [41]. In Polar-HR [34], it is a traditional training-free model and, therefore, shares the same input and output data as our approach. Besides, in SHIQ [6], since it cannot support polarized images, we convert four polarized images to one unpolarized image.

5.2. Evaluation on Reconstruction Task

Quantitative Evaluation To measure the quantitative performance of the algorithm, we refer to commonly used metrics for Image Reconstruction: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). Also, to compare the overall performance of our dataset, we use 10-fold cross-validation during training and testing. This strategy ensures the evaluation’s reliability and further proves the proposed algorithm’s robustness.

The quantitative results are shown in Table 1. In the last column of Table 1, μ is the mean value, and σ is the variance of the 10-fold performance. These experimental results show the average PSNR is 36.48 and the average SSIM is 0.9824, and the results demonstrate the high quality of our reconstructed images. In addition, the variance of PSNR and SSIM are 0.57 and 1.23×10^{-5} , respectively, and this proves the proposed algorithm is stable and robust.

Qualitative Evaluation To evaluate the reconstruction performance of the algorithm in a qualitative way, we show some visual results. Our results are shown in Fig. 8. Although Polar-HR [34] and SHIQ [6] can remove some of the highlights, their models cannot model the film correctly and, therefore, cannot remove the film itself. Besides, it can be observed that our algorithm can reconstruct regions where text or QR codes are corrupted by highlights through polarization information, while Uformer [33] and Restormer [41] produce fake artifacts in these regions.

5.3. Evaluation on Downstream Applications

In industrial environments, it is common to use transparent films to cover products, which may negatively im-

		K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	$\mu \uparrow$	$\sigma \downarrow$
SHIQ [6]	PSNR	23.47	22.11	21.95	21.69	21.77	21.03	20.86	20.46	21.10	21.31	21.58	0.64
	SSIM	0.7899	0.7640	0.7416	0.7439	0.7459	0.7465	0.7499	0.7412	0.7465	0.7300	0.7499	2.41×10^{-4}
Polar-HR [34]	PSNR	23.31	22.80	22.13	21.58	21.94	22.00	22.03	21.99	22.18	21.95	22.19	0.22
	SSIM	0.7642	0.7421	0.7220	0.7099	0.7064	0.7098	0.7128	0.7017	0.7102	0.6968	0.7176	3.80×10^{-4}
Uformer [33]	PSNR	31.85	31.95	31.39	31.19	31.81	32.04	31.68	31.98	31.85	31.01	31.68	0.11
	SSIM	0.9519	0.9456	0.9371	0.9364	0.9434	0.9421	0.9438	0.9435	0.9457	0.9363	0.9426	2.17×10^{-5}
Restormer [41]	PSNR	34.35	35.02	34.44	33.71	34.88	35.13	34.31	34.33	34.51	32.49	34.32	0.52
	SSIM	0.9771	0.9770	0.9721	0.9678	0.9757	0.9746	0.9742	0.9741	0.9759	0.9633	0.9731	1.75×10^{-5}
Ours	PSNR	36.76	37.29	36.62	35.12	36.93	37.21	36.24	36.67	36.94	35.02	36.48	0.57
	SSIM	0.9852	0.9859	0.9822	0.9767	0.9845	0.9833	0.9836	0.9830	0.9850	0.9749	0.9824	1.23×10^{-5}

Table 1. Quantitative evaluation in 10-fold cross-validation. The K-I indicates the I_{th} fold.

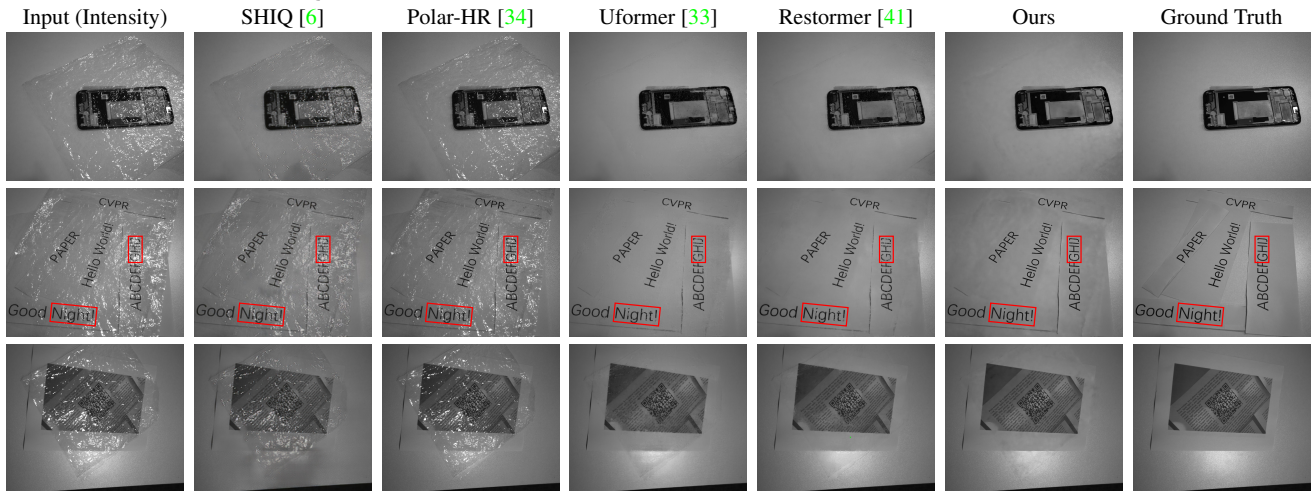


Figure 8. Qualitative Evaluation. Compared with other baselines, our model can reconstruct more realistic details in highlight regions instead of fake artifacts. **Please zoom in for more details.**

	SHIQ [6]	Polar-HR [34]	Uformer [33]
Read Number	7	17	18
Read Rate	8.05%	19.54%	20.69%
	Restormer [41]	Ours	Ground Truth
Read Number	24	63	87
Read Rate	27.58%	72.41%	100%

Table 2. QR code reading rate. Compared with other baselines, our approach can improve the performance of QR codes scanner in the industrial environment.

impact the robustness of downstream algorithms. To evaluate the effectiveness of our solution in such settings, we conduct two tests on two downstream tasks: QR Code Reading and Text Optical Character Recognition (Text OCR). These tasks are particularly relevant as they require access to raw information, which provides a rigorous test of the effectiveness of our algorithms in industrial scenarios.

QR Code Reading The read rate is a critical performance metric for manufacturing pipelines in industrial systems. In this experiment, we compare our method with other baselines in Table 2. Our upstream algorithm leads to a significant improvement in the performance of the QR code reading. The qualitative results in Fig. 9 demonstrate that our algorithm not only removes the film but also significantly

recovers the QR code information under the film.

Text OCR Text OCR is also an important downstream industrial task. We compared our algorithm with other baselines, as shown in Fig. 10. Our method can restore more hidden text information accurately.

5.4. Ablation Study

We conduct ablation studies, which mainly validate the effectiveness of our proposed polarization dataset and one crucial component in our framework. To ensure that the experimental settings are consistent, we perform 10-fold cross-validation for the ablation experiments as well and draw two boxplots for further analysis.

Effectiveness of the Polarized Information To prove the effectiveness of polarized information, we eliminate all components that need polarization information, which include four polarization images. Since the Polarization-based Location Model is driven by the I_{max} and I_{min} , which need to be calculated with polarization images as input, this structure is also removed. Other settings remain constant. Without the support of polarized information, the quantitative performance has a considerable decrease, as shown in Fig. 11 (w/o Polarized information). Also, the performance of the network is limited by the lack of guidance

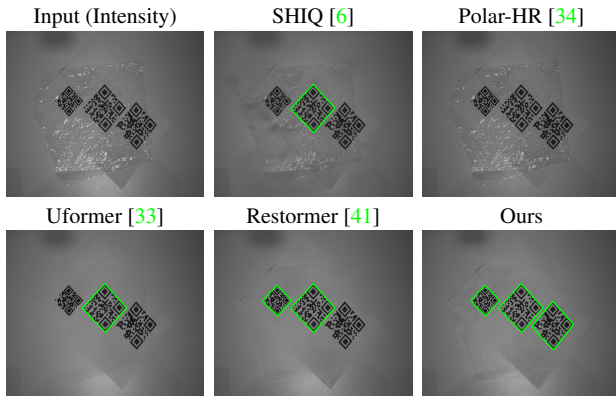


Figure 9. Performance of QR code reading in the industry. In the original image, the QR code scanner fails to detect QR code information. Polar-HR [34] and SHIQ [6] can only eliminate the specular reflection on the QR code, and Uformer [33] and Restormer [41] can only generate false artifacts in highlight regions. Our method can help industrial QR code scanners achieve a higher performance.

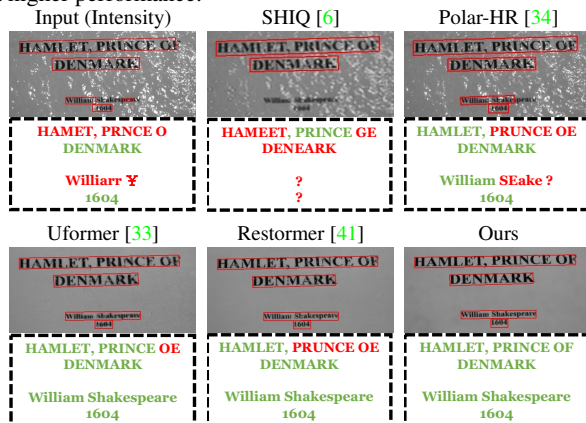


Figure 10. Performance of text OCR in the industry. Compared to other baselines, our method can reconstruct more original text information for recognition.

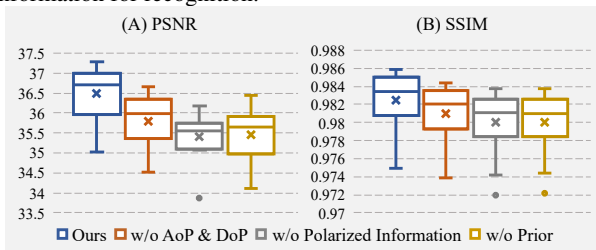


Figure 11. PSNR (A) and SSIM (B) in ablation study. We use two boxplots to describe the performance after t -fold cross-validation. The results show that our dataset, the extracted prior, the AoP, and the DoP contribute to the framework performance improvement. information, so it cannot remove all degradation. Besides, removing the polarization information will induce artifacts in highlight regions, as shown in Fig. 12.

Effectiveness of AoP and DoP To prove the effectiveness of AoP and DoP, we eliminate AoP and DoP in the angle estimation network. The qualitative results presented in Fig. 13 demonstrate that AoP and DoP can help locate the highlight more significantly, and also improve overall

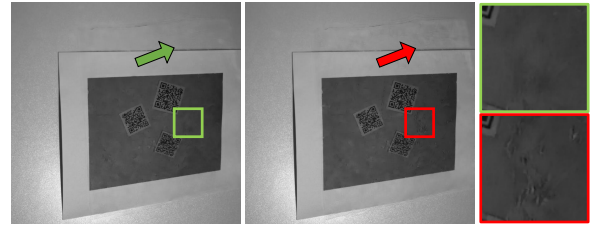


Figure 12. Qualitative evaluation of w/o Polarization Information. Compared with w/o polarization information (Red), introducing polarization information avoids the generation of artifacts and reconstructs the realistic details (Green).

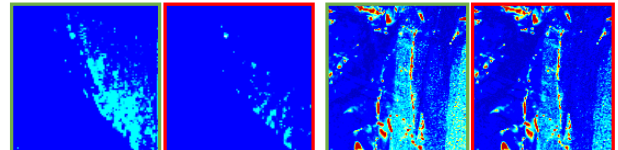


Figure 13. Highlight location heatmap. Compared with “w/o AoP & DoP” (In Red Box), introducing AoP and DoP (In Green Box) can help the A-Net to infer more highlight regions.

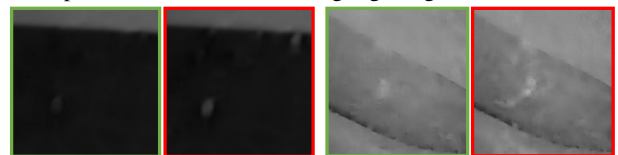


Figure 14. Qualitative Evaluation of w/o Prior. Compared with “w/o Prior” (In Red Box), our solution (In Green Box) is more effective to remove highlight and infer original information. performance in Fig. 11 (w/o AoP & DoP).

Effectiveness of Prior To prove the effectiveness of polarized prior, we eliminate the angle estimation network and the PLM in our network while retaining only the polarized images and the reconstruction network. In this case, the network can only implicitly facilitate the polarized information in the reconstruction network. Since there is no polarized prior that indicates highlight regions, this can lead to a decrease in the performance of the network to recover highlight regions (as shown in Fig. 14), while causing overall performance to decrease in Fig. 11 (w/o Prior).

6. Conclusion

In this study, we pioneer the investigation of the Film Removal (FR) problem, aiming to eliminate the disturbances caused by wrinkled transparent films and to restore the obscured information. We propose an end-to-end framework to effectively remove all degradations caused by the film with a polarized prior to minimizing highlight. Besides, we build a practical polarized dataset containing paired data for this problem. Experiments in the industry have demonstrated the potential application. We believe that the deployment of our algorithms will considerably improve the robustness of downstream industrial recognition systems.

Acknowledgment This work is supported by the National Natural Science Foundation of China (No. 62206068) and the Natural Science Foundation of Zhejiang Province, China under No. LD24F020002.

References

- [1] Gary A Atkinson. Polarized light in computer vision. In *Computer Vision: A Reference Guide*, pages 1005–1010. Springer, 2021. 2
- [2] Seung-Hwan Baek, Tizian Zeltner, Hyunjin Ku, Inseung Hwang, Xin Tong, Wenzel Jakob, and Min H Kim. Image-based acquisition and modeling of polarimetric reflectance. *ACM TOG*, 2020. 2
- [3] William S Bickel and Wilbur M Bailey. Stokes vectors, mueller matrices, and polarized scattered light. *American Journal of Physics*, 1985. 4, 5
- [4] DJ De Smet. Brewster’s angle and optical anisotropy. *American Journal of Physics*, 1994. 2
- [5] Gang Fu, Qing Zhang, Qifeng Lin, Lei Zhu, and Chunxia Xiao. Learning to detect specular highlights from real-world images. In *ACM Multimedia*, 2020. 2
- [6] Gang Fu, Qing Zhang, Lei Zhu, Ping Li, and Chunxia Xiao. A multi-task network for joint specular highlight detection and removal. In *CVPR*, 2021. 1, 6, 7, 8
- [7] T Gold. Polarization of starlight. *Nature*, 1952. 2
- [8] Yen-Ting Huang, Yan-Tsung Peng, and Wen-Hung Liao. Enhancing object detection in the dark using u-net based restoration module. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2019. 1
- [9] Bhaskar Kanseri et al. Degree and state of polarization control using brewster’s law in a nematic liquid crystal. *Optics & Laser Technology*, 2023. 2
- [10] Chanran Kim, Younkyoung Lee, Jong-Il Park, and Jaeha Lee. Diminishing unwanted objects based on object detection using deep learning and image inpainting. In *2018 International Workshop on Advanced Image Technology (IWAIT)*, 2018. 1
- [11] Chenyang Lei, Xuhua Huang, Chenyang Qi, Yankun Zhao, Wenxiu Sun, Qiong Yan, and Qifeng Chen. A categorized reflection removal dataset with diverse real-world scenes. In *CVPR*, 2022. 3
- [12] Chenyang Lei, Xuhua Huang, Mengdi Zhang, Qiong Yan, Wenxiu Sun, and Qifeng Chen. Polarized reflection removal with perfect alignment in the wild. In *CVPR*, 2020. 2
- [13] Chen Li, Stephen Lin, Kun Zhou, and Katsushi Ikeuchi. Specular highlight removal in facial images. In *CVPR*, 2017. 3
- [14] Yupeng Liang, Ryosuke Wakaki, Shohei Nobuhara, and Ko Nishino. Multimodal material segmentation. In *CVPR*, 2022. 2
- [15] Youwei Lyu, Zhaopeng Cui, Si Li, Marc Pollefeys, and Boxin Shi. Reflection separation using a pair of unpolarized and polarized images. In *NeurIPS*, 2019. 2
- [16] Etienne Louis Malus and French Physician. *Théorie de la double refraction de la lumiere dans les substances cristallisées*. Garnery: Baudouin, 1810. 5
- [17] William H McMaster. Matrix representation of polarization. *Reviews of modern physics*, 1961. 6
- [18] Abhimitra Meka, Maxim Maximov, Michael Zollhoefer, Avishek Chatterjee, Hans-Peter Seidel, Christian Richardt, and Christian Theobalt. Lime: Live intrinsic material estimation. In *CVPR*, 2018. 2
- [19] Mario Monzón, Zaida Ortega, Alba Hernández, Rubén Paz, and Fernando Ortega. Anisotropy of photopolymer parts made by digital light processing. *Materials*, 2017. 4
- [20] Miki Morimatsu, Yusuke Monno, Masayuki Tanaka, and Masatoshi Okutomi. Monochrome and color polarization demosaicking using edge-aware residual interpolation. In *ICIP*, 2020. 4
- [21] Shree K Nayar, Xi-Sheng Fang, and Terrance Boult. Separation of reflection components using color and polarization. *International Journal of Computer Vision*, 1997. 2, 3
- [22] David M. Rebhan, Maik Rosenberger, and Gunther Notni. Principle investigations on polarization image sensors. In *Other Conferences*, 2019. 2
- [23] Payam Refaeilzadeh, Lei Tang, and Huan Liu. Cross-validation. *Encyclopedia of database systems*, 2009. 4
- [24] Mohamed Sayed and Gabriel Brostow. Improved handling of motion blur in online object detection. In *CVPR*, 2021. 1
- [25] Boxiao Shen, Chuan Huang, Wenjun Xu, Tingting Yang, and Shuguang Cui. Blind channel codes recognition via deep learning. *IEEE Journal on Selected Areas in Communications*, 2021. 1
- [26] Yu P Sinichkin, AV Spivak, and DA Yakovlev. Effect of scattering anisotropy and material optical anisotropy of oriented fiber layers on the transmitted light polarization. *Optics and Spectroscopy*, 2010. 4
- [27] V Sundar and RE Newnham. Electrostriction and polarization. *Ferroelectrics*, 1992. 2
- [28] Domen Tabernik, Samo Šela, Jure Skvarč, and Danijel Skočaj. Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, 2020. 1
- [29] R Ulrich and A Simon. Polarization optics of twisted single-mode fibers. *Applied optics*, 1979. 2
- [30] Shinji Umeyama and Guy Godin. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *IEEE TPAMI*, 2004. 3
- [31] Chengjie Wang and Sachiko Kamata. Removal of transparent plastic film specular reflection based on multi-light sources. *2012 Symposium on Photonics and Optoelectronics*, 2012. 1
- [32] Yuan-Kai Wang and Chin-Fa Wang. Face detection with automatic white balance for digital still camera. In *2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2008. 1
- [33] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, 2022. 6, 7, 8
- [34] Sijia Wen, Yinqiang Zheng, and Feng Lu. Polarization guided specular reflection separation. *IEEE TIP*, 2021. 1, 2, 3, 4, 6, 7, 8
- [35] John T. Winthrop and C. R. Worthington. Theory of fresnel images. i. plane periodic objects in monochromatic light. *Journal of the Optical Society of America*, 1965. 5

- [36] L Wolfenstein. Polarization of fast nucleons. *Annual review of nuclear science*, 1956. 2
- [37] Lawrence B Wolff. Using polarization to separate reflection components. In *1989 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1989. 2
- [38] Xiao Xiao, Bahram Javidi, Genaro Saavedra, Michael Eismann, and Manuel Martinez-Corral. Three-dimensional polarimetric computational integral imaging. *Optics express*, 2012. 2
- [39] Tao Yang, Xiaofei Chang, Hang Su, Nathan Crombez, Yasmine Ruichek, Tomas Krajenik, and Zhi Yan. Raindrop removal with light field image using image inpainting. *IEEE Access*, 2020. 1
- [40] Dong Yin, Raphael Gontijo Lopes, Jon Shlens, Ekin Dogus Cubuk, and Justin Gilmer. A fourier perspective on model robustness in computer vision. In *NeurIPS*, 2019. 1
- [41] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022. 6, 7, 8
- [42] Lichi Zhang, Edwin R Hancock, and Gary A Atkinson. Reflection component separation using statistical analysis and polarisation. In *Iberian Conference on Pattern Recognition and Image Analysis*, 2011. 3
- [43] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *IEEE TPAMI*, 2020. 6
- [44] Dizhong Zhu and William AP Smith. Depth from a polarisation+ rgb stereo pair. In *CVPR*, 2019. 2