

PolarAnything: Diffusion-based Polarimetric Image Synthesis

Kailong Zhang^{1†} Youwei Lyu^{1†} Heng Guo^{1,2*} Si Li¹ Zhanyu Ma¹ Boxin Shi^{3,4}

¹Beijing University of Posts and Telecommunications, ²Xiong'an Aerospace Information Research Institute

³State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University

⁴National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

{zhangkailong, youweilv, guoheng, lisi, mazhanyu}@bupt.edu.cn shiboxin@pku.edu.cn

Abstract

Polarization images facilitate image enhancement and 3D reconstruction tasks, but the limited accessibility of polarization cameras hinders their broader application. This gap drives the need for synthesizing photorealistic polarization images. The existing polarization simulator Mitsuba relies on a parametric polarization image formation model and requires extensive 3D assets covering shape and PBR materials, preventing it from generating large-scale photorealistic images. To address this problem, we propose PolarAnything, capable of synthesizing polarization images from a single RGB input with both photorealism and physical accuracy, eliminating the dependency on 3D asset collections. Drawing inspiration from the zero-shot performance of pretrained diffusion models, we introduce a diffusion-based generative framework with an effective representation strategy that preserves the fidelity of polarization properties. Experiments show that our model generates high-quality polarization images and supports downstream tasks like shape from polarization. The project could be found in https://flzt11.github.io/PA_project/.

1. Introduction

Polarization-based computer vision tasks, including shape-from-polarization [1, 2, 39], image dehazing [62], and reflection removal [29, 31, 38, 57], leverage additional polarization cues to enhance RGB-based methods. Different from RGB cameras commonly equipped on mobile phones, polarization cameras are expensive and inaccessible to most users, limiting the application range of polarization-based vision systems. On the other hand, diverse and large-scale polarization image-based datasets with ground truth (GT) labels such as depth, surface normals, and semantic segmentation masks rarely exist, but these datasets are important for improving the performance of the correspond-

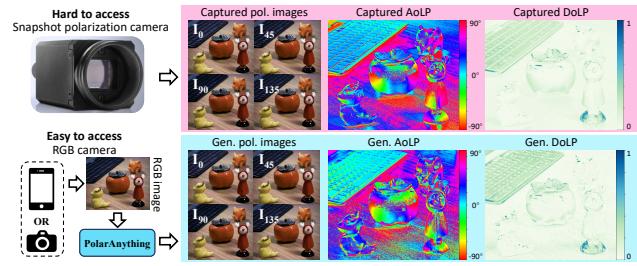


Figure 1. PolarAnything generates polarization images covering diverse shapes and materials from a single image captured by an accessible RGB camera while preserving the polarization properties on both AoLP and DoLP. The top and bottom rows show the polarization images captured by a snapshot polarization camera and the ones generated by PolarAnything, respectively.

ing learning-based computer vision tasks. In contrast, such datasets based on RGB images are widely available. Therefore, it is desired to develop a polarization image simulator taking as input RGB images to easily create polarization datasets and make the polarization-based vision algorithms accessible to users even without specialized polarization cameras.

Existing polarization image simulators are either based on collected measured polarized Bidirectional Reflectance Distribution Function (pBRDF) [3] or a parametric polarized reflectance model [4, 19, 28]. Due to the scarcity of polarization datasets, data-based simulators using measured pBRDFs are inherently limited in representing diverse real-world scenes. On the other hand, a simulator based on a parametric pBRDF model, such as Mitsuba [43], still has a gap in synthesizing polarization properties of real-world scenes due to its reflectance assumption, as demonstrated by fitting the measured pBRDF in [3]. Moreover, Mitsuba requires 3D assets including both surface geometry and paired physically-based material maps such as roughness and albedo to render polarization images. Compared to RGB images, 3D assets are hard to collect, and their amount is much smaller than that of RGB images. Besides, even with plenty of 3D assets, selecting and seamlessly inte-

* corresponding author. † equally contributed authors.

grating them into a scene for scene-level polarization image synthesis is even more challenging. To summarize, current polarization simulators still face considerable difficulties in generating large-scale photorealistic polarization images.

To handle these problems, we propose PolarAnything, a polarization image simulator taking a single RGB image as input, capable of handling diverse shape and materials, as shown in Fig. 1. RGB images are easier to acquire against 3D assets and there are plenty of RGB-based datasets with labeled ground truth (*e.g.*, depth, surface normal, segmentation) for a broad range of vision tasks. Furthermore, objects in daily RGB images are naturally distributed, making simulated polarization images photorealistic. To eliminate assumptions inherent in parametric pBRDF models, we propose a diffusion-based model for polarization image generation. Specifically, this model conditions an RGB image to perform diffusion on the encoded Angle of Linear Polarization (AoLP) and Degree of Linear Polarization (DoLP). Compared to simply conducting diffusion on the polarization images under different polarization angles, we show such a design can better preserve polarization information.

Based on PolarAnything, we show existing RGB-based datasets such as Stanford-ORB [30] can be transferred to their polarimetric version. Specifically, we create Polar-Stanford-ORB with 300 paired polarimetric images and their corresponding GT surface normal maps, benefiting the training of learning-based polarimetric vision tasks such as DeepSfP [2] requiring labeled GT.

To summarize, the contributions are threefold:

- We propose PolarAnything, the first polarization image simulator based on RGB image as input, reducing the effort required to obtain polarization images;
- We show that finetuning Stable Diffusion on the AoLP and DoLP instead of the polarization images can better preserve the polarization information;
- We create PolarStanford-ORB based on our photorealistic polarization image synthesis and demonstrate the effectiveness of PolarAnything on downstream polarization-based computer vision tasks.

2. Related Works

2.1. Polarization Applications and Acquisition

Polarization applications By providing additional measurements, polarization images can enhance a wide range of downstream vision tasks in various ways. Polarization encodes detailed geometry and reflectance information of object surfaces and is well studied for surface normal recovery [1, 2, 6, 18, 32, 39, 41], depth estimation [21, 25, 51, 55], 3D reconstruction [7, 14, 16, 33, 61], and reflectance decompositions [3, 4, 12], further imposing geometric constraints on 6D pose predictions [10]. As an important light property, polarization status reflects the light propagation process,

such as scattering in air and refraction in the semi-reflector, and provides unique cues for solving dehazing [62], reflection removal [29, 31, 38, 57], and transparent object segmentation [26, 40] problems. Light polarization characterizes oscillation magnitudes along two orthogonal orientations on the 2D frame, which inspires researchers to utilize polarized light for designing the structure light system [20] and alpha matte extraction setup [8].

Polarization image capture and synthesis Polarization image capture requires a specialized polarization camera or manually rotated polarizer filters, and the laboursome acquisition process restricts broader applications of polarization images. Mitsuba 2 [43] is a retargetable forward and inverse renderer aiming to formulate the light transport process, which enables polarization image rendering and data synthesis. However, it is necessary to carefully configure light conditions and use a large amount of tailored 3D models and Physically Based Rendering (PBR) materials in rendering, making the large-scale production of high-quality polarization data costly. PolarAnything aims to take as input the most common RGB images and produce polarization images based on a diffusion model at a low cost, further facilitating polarization research for the community.

2.2. Image Synthesis with Diffusion Models

With the advent of deep learning, data-driven approaches have been the widely adopted tool for data synthesis. Given large-scale datasets for supervision, UNet-based convolutional neural networks could generate 4D RGBD light-field images [54], and dual-pixel images [34], and simulate motion blur [5]. Adversarial generative network (GAN) [13] is shown to be a powerful generative method, while diffusion models [17, 52, 53] demonstrate superior image synthesis capabilities. Denoising Diffusion Probabilistic Models (DDPMs) [17] have become an emerging research spot of generative models, which present impressive high-quality image synthesis results. DDPMs learn to invert a diffusion process that incrementally corrupts images with Gaussian noise, enabling them to generate samples from the data distribution by applying the reverse process to random noise. Conditional diffusion models extend DDPMs by incorporating additional information, such as text [48] and images [47]. In the field of text-based image generation, Stable Diffusion [46] is based on a latent diffusion model trained on extensively large-scale datasets [49] and achieves unprecedented image synthesis quality. With distilled knowledge from vast data, the strong diffusion priors could help to produce surprising zero-shot performance. Several studies have investigated methods to leverage the powerful pre-trained model. ControlNet [60] proposes a general fine-tune framework to integrate additional conditions with pre-trained diffusion models such as Stable Diffusion [46] to

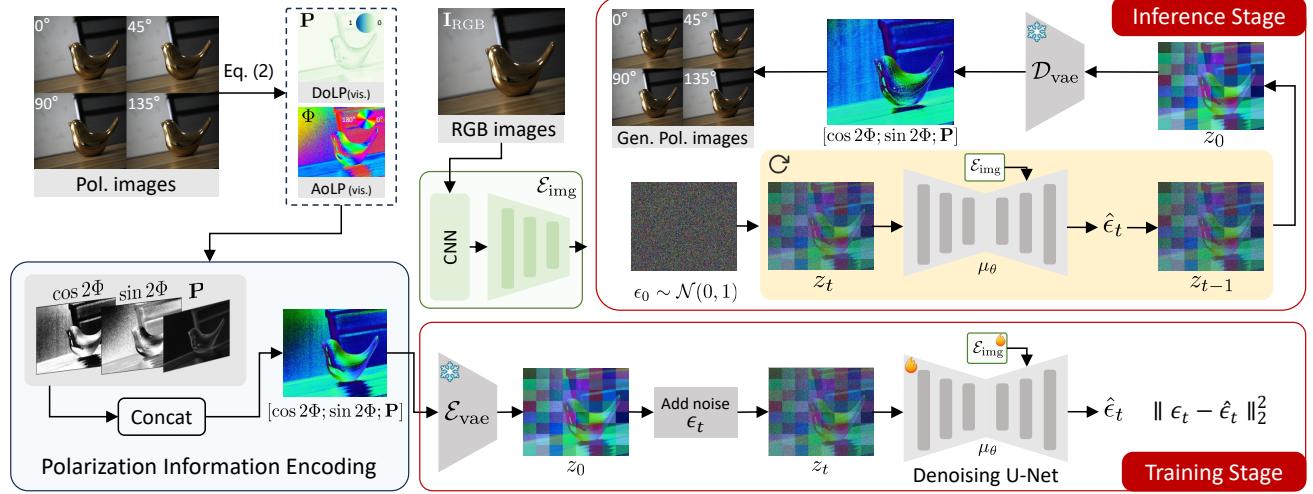


Figure 2. Framework of PolarAnything. In the training stage, the image encoder takes as input the RGB image and extracts the conditional feature for prompting U-Net denoising. During inference, PolarAnything is fed with an RGB image and generates high-quality AoLP and DoLP via steps of denoising.

accomplish more customized tasks. Recent works, StableNormal [58], GeoWizard [9], and Marigold [27], propose to use diffusion models as geometric information predictors. Gao *et al.* [11] design a position-aware warping scheme providing an initial light field pattern, which conditions the diffusion model to generate light field images. StereoDiffusion [56] introduces the disparity map to guide the stereo pixel shift operation into the diffusion process, which predicts stereo images without finetuning the original model. Motivated by diffusion models generating remarkable results in discriminative tasks, we propose PolarAnything to restore polarization properties based on powerful pretrained priors, achieving physically reliable polarization prediction.

3. Method

PolarAnything is a generalizable latent diffusion model for synthesizing polarimetric observations under arbitrary polarization angles from an RGB image. We achieve this goal by creating a high-quality polarization image dataset and finetuning a pretrained image-conditioned generative model on the dataset. Figure 2 displays the overall architecture of PolarAnything.

3.1. Preliminaries on Polarization

Polarization images are acquired with polarizer filtering at different polarizer angles. The intensity of the light passing through a polarizer with an angle of Θ is given by:

$$\mathbf{I}_\Theta = \frac{\mathbf{I}_{\text{RGB}}}{2} (1 + \mathbf{P} \cos(2\Theta - 2\Phi)), \quad (1)$$

in which \mathbf{I}_{RGB} denotes incoming light intensities before the polarizer. DoLP \mathbf{P} denotes the strength ratio of linearly po-

larized light to total incident light. Φ denotes AoLP, *i.e.*, the oscillation orientation of the polarized component. With a snapshot polarization camera, four polarization images, *i.e.*, $\mathbf{I}_{0^\circ}, \mathbf{I}_{45^\circ}, \mathbf{I}_{90^\circ}, \mathbf{I}_{135^\circ}$, can be obtained with a single shot, and the polarization properties AoLP and DoLP can be computed as follows:

$$\begin{aligned} \Phi &= \frac{1}{2} \operatorname{atan2} \frac{\mathbf{I}_{45^\circ} - \mathbf{I}_{135^\circ}}{\mathbf{I}_{0^\circ} - \mathbf{I}_{90^\circ}}, \\ \mathbf{P} &= \frac{\sqrt{(\mathbf{I}_{0^\circ} - \mathbf{I}_{90^\circ})^2 + (\mathbf{I}_{45^\circ} - \mathbf{I}_{135^\circ})^2}}{(\mathbf{I}_{0^\circ} + \mathbf{I}_{45^\circ} + \mathbf{I}_{90^\circ} + \mathbf{I}_{135^\circ})/2}. \end{aligned} \quad (2)$$

3.2. PolarAnything

Polarized information encoding For polarization image generation, one option is to denoise polarization images conditioned on specific polarizer angles. However, we observe that the polarization properties computed from the image are damaged, which may be attributed to the absence of physical polarization constraint during the denoising steps. To generate physically plausible polarization properties, we propose to estimate corresponding AoLP and DoLP maps and then simulate polarization images under arbitrary polarizer angles. Considering the π -periodicity of AoLP, we encode AoLP in the sinusoidal form as $(\cos 2\Phi, \sin 2\Phi)$, which is also a continuous representation facilitating network learning [63]. Consistent with the value range of encoded AoLP and the VAE, we normalize DoLP within $[-1, 1]$, and then concatenate the encoded AoLP with DoLP as diffusion output: $[\cos 2\Phi; \sin 2\Phi; \mathbf{P}]$. The image of encoded AoLP and DoLP is visualized in Fig. 2. Given an RGB image \mathbf{I}_{RGB} and its AoLP and DoLP, polarization images of arbitrary polarizer angles can be obtained by Eq. (1).

Diffusion-based polarization generator Pre-trained diffusion models have shown impressive zero-shot performance in text-based image generation and discriminative tasks. A Denoising Diffusion Probabilistic Model (DDPM) depicts data distributions by progressively transforming noisy samples via reverse diffusion processes $z_{t-1} = \beta_t z_t - \mu_\theta(z_t, t) + \epsilon_t$, where β_t is a variance schedule, ϵ_t denotes the noise added at timestep t , and the model learns to iteratively denoise z_t via U-Net μ_θ parameterized by θ .

We propose PolarAnything to generate polarization properties from a sole RGB image \mathbf{I}_{RGB} by a latent denoising diffusion model, which models the conditional distribution $p(\cos 2\Phi; \sin 2\Phi; \mathbf{P} | \mathbf{I}_{\text{RGB}})$. The overview of PolarAnything is shown in Fig. 2. PolarAnything is based on the pretrained large diffusion model Stable Diffusion v1.5 [46], which learns representative and generalizable image priors from LAION-5B [49]. The model consists of an image encoder \mathcal{E}_{img} , a pretrained VAE encoder \mathcal{E}_{vae} , a denoising U-Net μ_θ , and a pretrained VAE decoder \mathcal{D}_{vae} . We obtain the conditional signal by encoding the input RGB image with the condition encoder \mathcal{E}_{img} . The condition encoder comprises two cascaded modules: the first is an RGB image feature extractor, which consists of several convolutional layers with SiLU activations; the latter feature encoder shares the same structure as that of the denoising U-Net encoder. The condition encoder \mathcal{E}_{img} generates hierarchical guidance features for feature fusion in the decoder of denoising U-Net like ControlNet [60]. The VAE encoder \mathcal{E}_{vae} and decoder \mathcal{D}_{vae} conduct the conversion between the encoded AoLP and DoLP map and its latent code z_t . The weights of the VAE encoder and decoder are fixed in the finetuning stage. The U-Net μ_θ is conditioned on the hierarchical guidance features $\mathcal{E}_{\text{img}}(\mathbf{I}_{\text{RGB}})$ to denoise polarization properties in the latent space. We finetune PolarAnything with the following objective:

$$\min_{\theta} \mathbb{E}_{x \sim \mathcal{E}_{\text{vae}}, t, c \sim \mathcal{N}(0, 1)} \|\epsilon_t - \mu_\theta(z_t, t, c, \mathcal{E}_{\text{img}}(\mathbf{I}_{\text{RGB}}))\|_2^2, \quad (3)$$

in which c is the text prompt embedding generated by CLIP [45].

At the inference stage, the latent code of polarization property maps is initialized as standard Gaussian noise, which is iteratively denoised with the same schedule as finetuning.

3.3. Real-world Polarization Dataset

To generate more realistic and physically plausible polarization images, we train PolarAnything solely with real captured images. Collecting polarization images requires manual rotation of a lens-mounted polarizer or a specialized polarization camera, which makes large-scale dataset creation time-consuming. Current real-world polarization image datasets are insufficient to support PolarAnything training. 1) Small amount of high-quality data: Morimatsu *et al.* [42] and Qiu *et al.* [44] provide only 40 and 38 sets of polarization images, respectively; 2) large amount of data captured in limited scenes: HAMMER [23], House-Cat6D [24], and PPP [10] contain 930, 2358, 5000 polarization images, respectively, but are captured in similar indoor scenes mainly for 6D pose estimation.

al. [42] and Qiu *et al.* [44] provide only 40 and 38 sets of polarization images, respectively; 2) large amount of data captured in limited scenes: HAMMER [23], House-Cat6D [24], and PPP [10] contain 930, 2358, 5000 polarization images, respectively, but are captured in similar indoor scenes mainly for 6D pose estimation.

To facilitate PolarAnything training and testing, we use a polarization camera [22] to create an RGB polarization image dataset featuring various shapes and reflectances under divergent light conditions. Our real-world dataset comprises 1,148 high-quality polarization images, each with a resolution of 1224×1024 . Over 100 different objects were included in the acquisition, spanning a broad spectrum of categories: transparent/opaque, conductive/dielectric, and diffuse/specular surfaces. Given that illumination can influence polarization properties [18], we captured images under 19 varied lighting environments, including 8 outdoor scenes and 11 indoor scenes. An overview of our camera rig and part of the objects is provided in the supplementary material. We combine data from Morimatsu *et al.* [42] with 1,115 images from our real-world dataset to form the PolarAnything training dataset, while the remaining 33 images are reserved for evaluating generation quality.

4. Experiment

4.1. Implementation Details

We finetuned the pretrained Stable Diffusion V1.5 backbone using the AdamW [37] optimizer with parameters $\beta_1=0.9$, $\beta_2=0.999$, and a weight decay of 0.001. We set a fixed learning rate of 4×10^{-5} and a batch size of 16 during finetuning. We finetuned the model for 600 steps on 8 NVIDIA A100 cards, which took about 10h. Different from ControlNet [60], all the weights of denoising U-Net are trainable during finetuning, which is experimentally proved to be more effective. For data augmentation, we randomly crop the original 1224×1024 polarization image into 512×512 patches. Since polarization images captured by [22] consist of 3 channels(R-G-B), we first convert them to grayscale and then compute DoLP and AoLP using Eq. (2) as the ground truth.

4.2. Comparison on Polarization Image Synthesis

Evaluation metric To measure the difference between GT and generated polarization images $\{\mathbf{I}_0, \mathbf{I}_{45}, \mathbf{I}_{90}, \mathbf{I}_{135}\}$, we use PSNR and SSIM as evaluation metrics. To evaluate generated polarization properties, mean absolute error (MAbsE) and mean angular error (MAngE) are adopted to assess generated DoLP and AoLP, respectively.

Due to the π -periodicity of AoLP, we formulate the angular error between GT and synthesized AoLP $(\Phi, \hat{\Phi})$ as

$$\text{AngE}(\Phi, \hat{\Phi}) = \min \left\{ \left| \hat{\Phi} + k\pi - \Phi \right|, k = -1, 0, 1 \right\}. \quad (4)$$

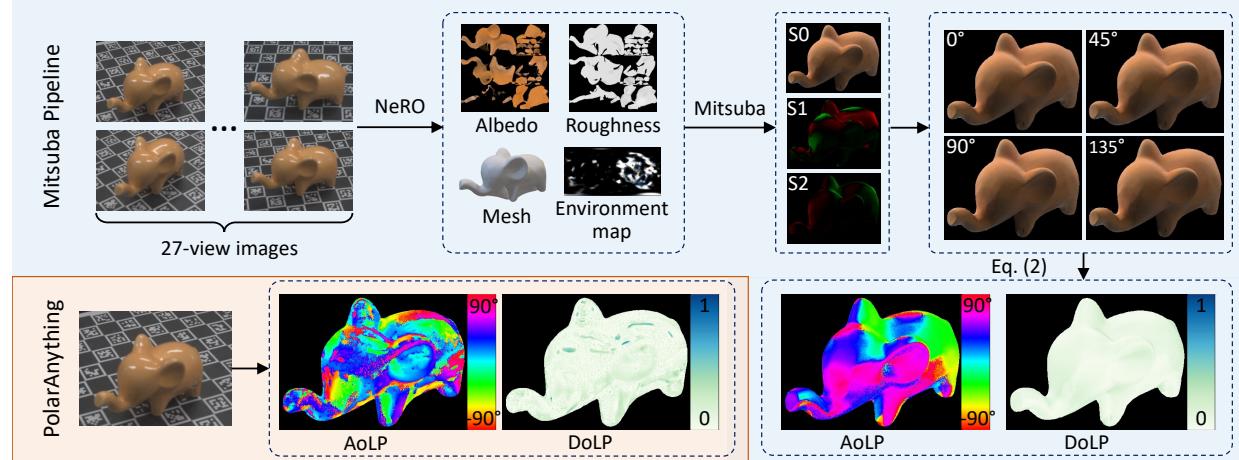


Figure 3. Comparison between Mitsuba and PolarAnything pipelines for generating polarization properties. Multiview images of ELEPHANT are captured to generate the mesh, environment light, and PBR parameter maps for rendering polarization images with Mitsuba [43], while PolarAnything only needs one RGB image for polarization properties generation.

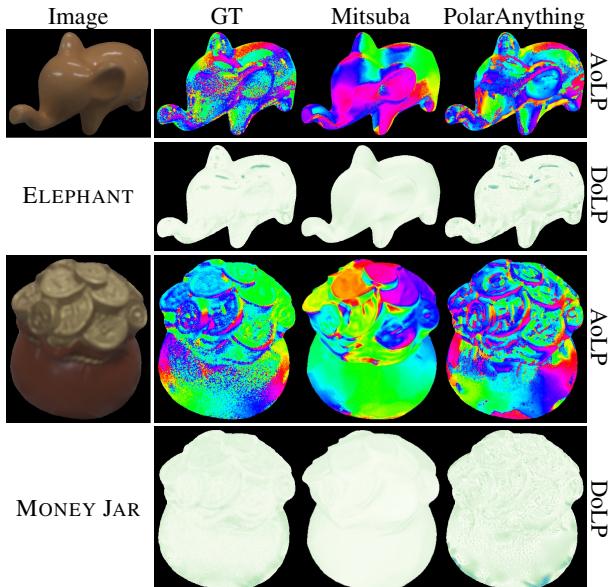


Figure 4. Qualitative comparison with Mitsuba [43] on selected views of ELEPHANT and MONEY JAR.

Comparison with Mitsuba We compare the quality of polarization images synthesized from our generative model PolarAnything and physics-based renderer Mitsuba [43]. For polarization simulation, Mitsuba requires environment maps, surface meshes, BRDF, and physically based rendering (PBR) parameters such as roughness and albedo. Figure 3 illustrates the polarization generation pipelines of Mitsuba and PolarAnything. To reproduce polarization images using Mitsuba, we capture multi-view polarization images of a dielectric object ELEPHANT and a diffuse object MONEY JAR, each set contains approximately 26 views with calibrated camera poses with Metashape [36]. Applying the state-of-the-art neural inverse rendering technique

NeRO [35] on unpolarized components of captured images, we obtain the corresponding shapes, environment maps, and the PBR parameter maps. Given these inputs, we can render polarization images using Mitsuba [43] for each view. On the other hand, PolarAnything only takes the unpolarized component of each polarization image as input and generates the polarized ones. Figure 4 compares PolarAnything against Mitsuba on generated AoLP and DoLP maps. Despite taking only a single RGB image as input, PolarAnything produces AoLP and DoLP much closer to the ground truth, achieving consistent performance across both specular (ELEPHANT) and diffuse (MONEY JAR) surface types. Quantitative evaluation is infeasible in this experiment due to misalignment between the Mitsuba-rendered images and the GT ones, which is attributed to camera calibration errors and inherent inaccuracies in NeRO [35].

Comparison with real polarization images Figure 5 displays the PolarAnything generated polarization properties compared to the real-captured ones. We show that PolarAnything can handle a wide range of materials, including diffuse, dielectric, metallic, and even transparent surfaces. Columns 1 to 4 show the results on our test set. In the 5-th and 6-th columns, we test PolarAnything on public real-world polarization image datasets released by NeRSP [16] and PANDORA [7]. Though the two datasets may have a distribution gap compared to our captured dataset for training, PolarAnything still demonstrates its generalization ability. As shown in the last two columns, PolarAnything is capable of generating realistic scene-level polarization properties, which are challenging for parametric polarization image simulators like Mitsuba [43] due to the scarcity of scene-level 3D assets with consistent geometry and PBR materials. Our method can easily simulate a polarization image with a causal capture of the target scene.

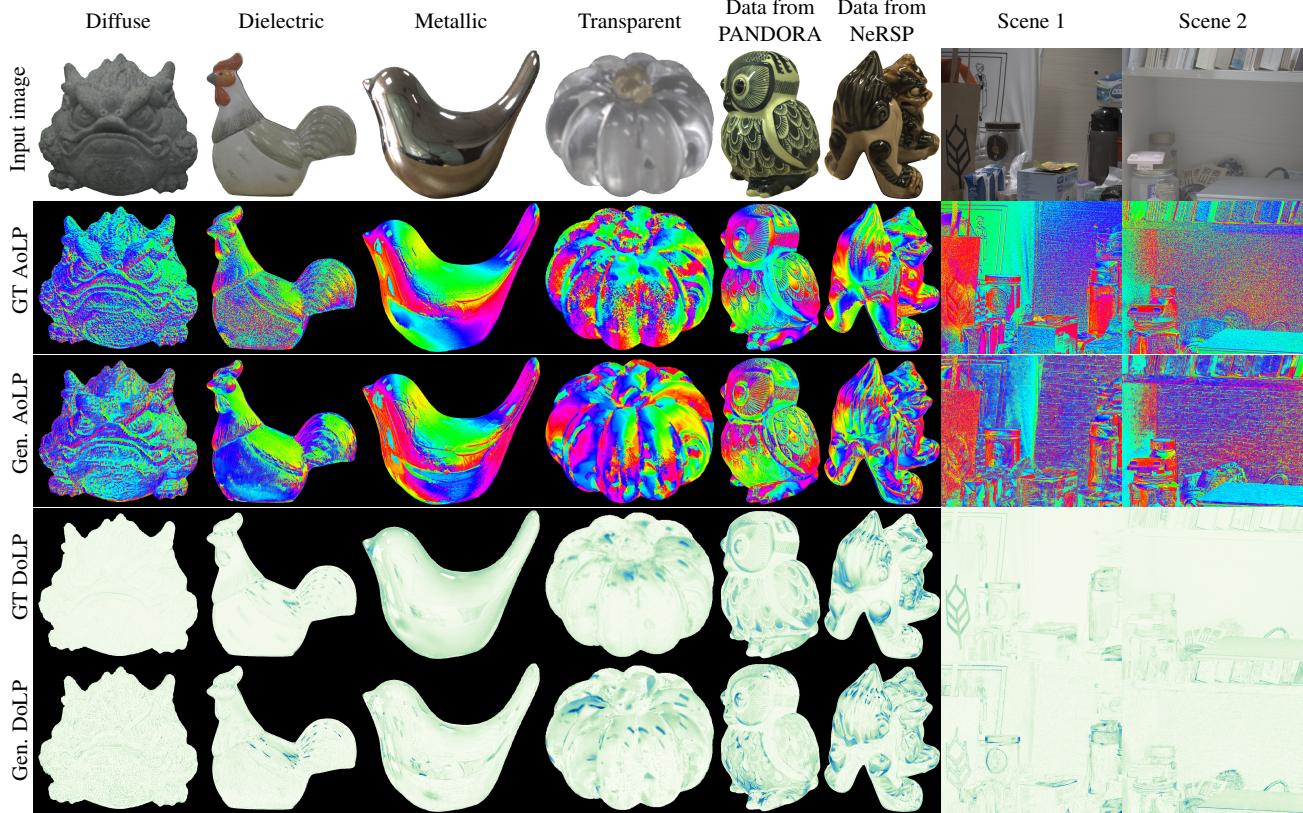


Figure 5. Qualitative evaluation on causal captured images with diverse shape and reflectance. The method works robustly on different materials (Columns 1-4), public polarization test datasets (Columns 5-6), and scene-level polarization synthesis (Columns 7-8).

4.3. Application: Boosting Shape from Polarization

Shape from polarization (SfP) reconstructs 3D surfaces from single- or multi-view polarization images by leveraging the geometric cues encoded in AoLP and DoLP. With PolarAnything, we show 1) existing multi-view SfP methods can be directly applied to RGB images by first converting them into synthesized polarization images, achieving shape reconstruction results comparable to those using real polarization data; and 2) by integrating PolarAnything with Stanford-ORB [30] containing RGB-normal pairs, we introduce **PolarStanford-ORB**, a dataset comprising synthesized polarization images and their corresponding surface normals. This dataset benefits learning-based SfP methods by expanding the training set while mitigating the need for costly polarization image acquisition and ground-truth surface normal collection.

Evaluation on multiview SfP We choose the state-of-the-art multiview SfP method PISR [15] as the baseline. We adopt the test sample SHISA from [16] for comparison, which comprises 6 polarization images with GT camera poses and normal maps. To verify the effectiveness of PolarAnything, we feed PISR [15] with the real-captured

Table 1. Comparisons of the single-view SfP method DeepSfP [2] retrained on different datasets. We evaluate the retrained models on the original test set (“DP”), our captured dataset (“PN”), and the combination of the two datasets (“DP+PN”).

Test set	Training set	Angular error ($^{\circ}$) ↓			Accuracy (%) ↑		
		Mean	Median	RMSE	$\leq 10^{\circ}$	$\leq 20^{\circ}$	$\leq 30^{\circ}$
DP	DeepSfP [2] ^a	19.38	15.21	23.89	34.98	70.74	82.00
	DeepSfP+MSO	19.15	15.87	22.95	32.46	71.63	83.47
	DeepSfP+PSO	19.20	15.71	23.01	33.46	72.06	83.41
PN	DeepSfP [2] ^a	30.69	26.88	35.86	9.80	34.49	58.26
	DeepSfP+MSO	24.22	21.14	28.42	12.95	47.88	72.26
	DeepSfP+PSO	22.93	20.08	27.58	14.39	50.08	77.47
DP+PN	DeepSfP [2] ^a	22.21	18.13	26.88	24.68	57.39	76.07
	DeepSfP+MSO	20.42	17.19	24.31	23.23	61.32	80.67
	DeepSfP+PSO	20.13	16.81	24.15	24.47	62.00	81.92

^a Original training data from DeepSfP [2]

and generated polarization data, respectively, and the qualitative and quantitative results are shown in Fig. 6. Chamfer Distance (CD) is adopted to evaluate the estimated meshes. Our method can generate physically reasonable polarization properties (“Gen. AoLP” and “Gen. DoLP”). The reconstructed 3D shapes from generated images by PolarAnything are comparable to those from real-captured images. More reconstruction results of PISR [15] are provided in the supplementary material.

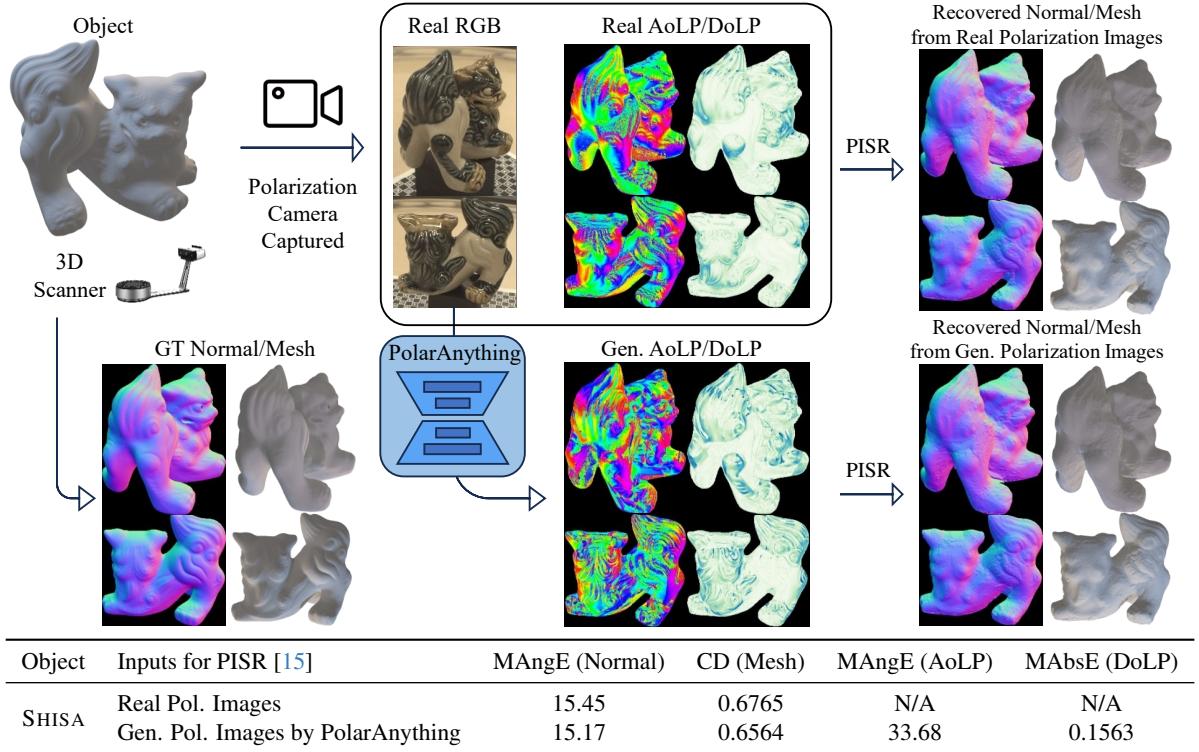


Figure 6. Evaluating PolarAnything on multiview shape from polarization based on PISR [15].

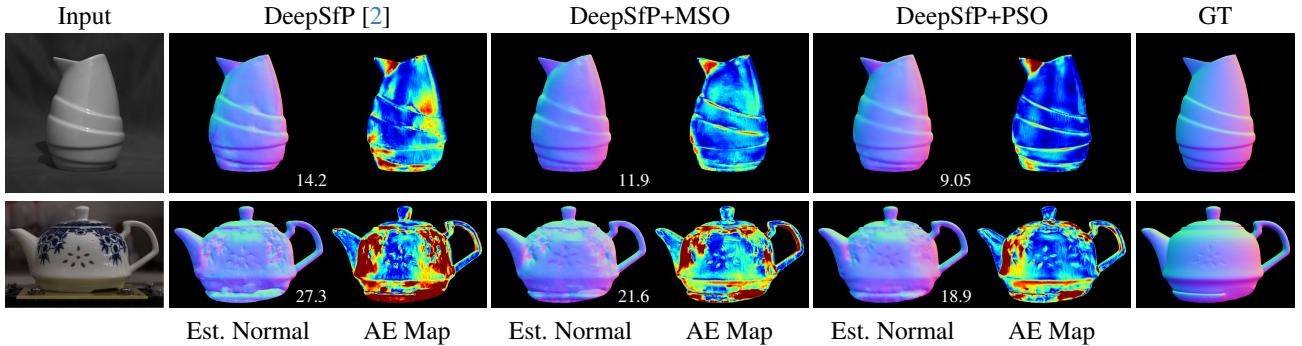


Figure 7. Evaluating PolarAnything on single-view shape from polarization based on DeepSfP [2]. “AE Map” denotes the angular error map between GT and estimated surface normal maps. The numbers at the bottom-right corner denote MAngE.

PolarStanford-ORB for enhancing single-view SfP PolarAnything aims at easily producing polarization images from common RGB image datasets for learning-based tasks, mitigating the barrier of polarization data acquisition. To validate this, we randomly pick 300 RGB images from Stanford-ORB [30] and create PolarStanford-ORB (PSO) by PolarAnything to expand the training dataset of DeepSfP [2], which contains only 236 training data. For a fair comparison, we also render 300 polarization images using Mitsuba with 3D assets from Stanford-ORB [30] named MitsubaStanford-ORB (MSO) for training set expansion. The DeepSfP model is retrained on the three different datasets, *i.e.*, “DeepSfP [2]”, “DeepSfP+MSO”, and

“DeepSfP+PSO”, with the same training strategy proposed in the original paper [2].

These three retrained models are evaluated on two test datasets: DP and PN, where “DP” refers to the real-world test set from DeepSfP [2] containing 21 objects. Since the capture environment of “DP” is similar to the training dataset used in DeepSfP [2], we additionally capture a dataset named “PN” containing polarization images and surface normal pairs of 7 objects to test the generalizability of the three retrained models. The capturing process of “PN” follows DiliGenT [50], and more details are provided in the supplementary material. Figure 7 and Table 1 present the qualitative and quantitative evaluations.

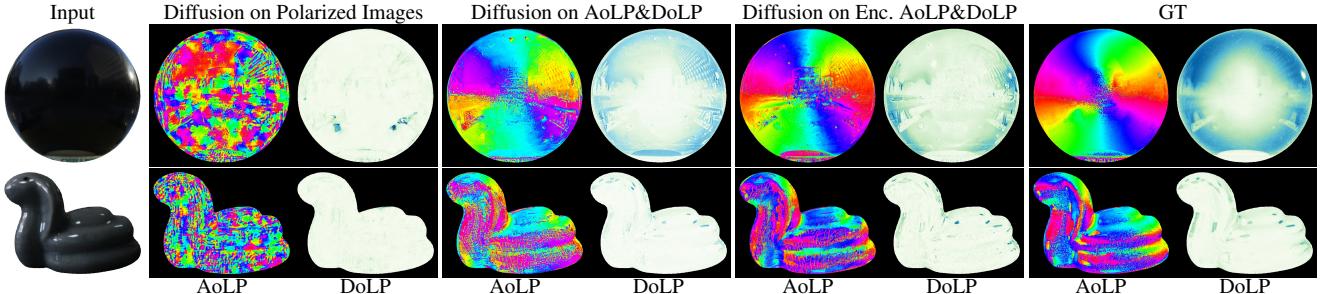


Figure 8. Ablation study. Naively generating polarization images leads to unreliable polarization properties. PolarAnything generates more accurate polarization predictions with encoded AoLP.

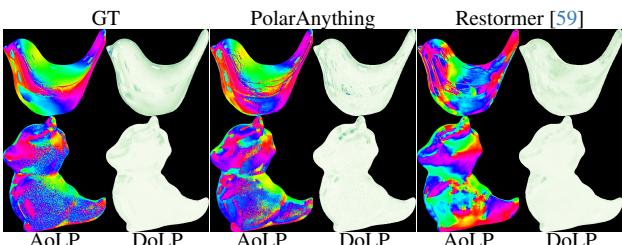


Figure 9. Comparison of Transformer-based and Diffusion-based architectures for polarization synthesis.

The quantitative results for ‘‘DeepSfP+MSO’’ indicate that adding more polarization data enhances the model’s generalization. When replacing the Mitsuba-rendered polarization data with PolarAnything, we observe a similar reduction in MAngE for normal predictions, as shown in ‘‘DeepSfP+PSO’’. This suggests that polarization images from PolarAnything can be effectively used in SfP. With PolarStanford-ORB augmentation, ‘‘DeepSfP+PSO’’ improves the Mean and RMSE metrics on the DeepSfP test set ‘‘DP’’, though the gains are marginal due to the similar distribution between the training and test sets. Testing on ‘‘PN’’, our newly captured real-world dataset with a different distribution, reveals more significant improvements, highlighting PolarAnything’s effectiveness in improving the generalizability.

4.4. Ablation study

To evaluate the effect of polarized information encoding in PolarAnything, we test several diffusion targets against our polarization formulation: a) PolarAnything generating 4 polarization images (*e.g.*, \mathbf{I}_{0° , \mathbf{I}_{45° , \mathbf{I}_{90° , \mathbf{I}_{135°). b) PolarAnything generating AoLP and DoLP maps. c) PolarAnything generating encoded AoLP and DoLP maps. For setting a), we generate the four polarization images concatenated in the height and width dimensions as a 2×2 grid. For b), we also use a single denoising U-Net to produce AoLP and DoLP maps as two channels in the output. All the ablative experiments are conducted with the same training dataset and training strategy as PolarAnything. As shown in Fig. 8 and Table 2, choice a) predicting polarization images only

Table 2. Ablation study of each polarization representation in PolarAnything on our real-world dataset.

Generation Target	PSNR \uparrow	SSIM \uparrow	MAngE \downarrow	MAbsE \downarrow
Polarization images	23.23	0.9165	45.67	0.1233
AoLP&DoLP	40.57	0.9904	29.46	0.1100
Encoded AoLP&DoLP	41.74	0.9927	25.33	0.1075

retains RGB radiance information but hardly recovers polarization properties, due to the absence of physical constraint on polarization. Both quantitative and qualitative results show that choice b) produces worse results than our method. Since the range of AoLP values is within $[-90^\circ, 90^\circ]$, directly regressing AoLP values can disrupt its inherent periodicity making the network challenging to learn.

We also test different network architectures trained on the same training dataset to validate our motivation for employing diffusion models in polarization simulation. The synthesis results using Restormer [59] as the baseline are shown in Fig. 9. Compared to PolarAnything, Restormer [59] struggles with AoLP and DoLP estimation due to its reliance on large-scale datasets, and the scarcity of polarization data limits its effectiveness in estimating polarization properties. In contrast, our chosen diffusion models, with their strong zero-shot capabilities, demonstrate superior performance in generating polarization images from a single RGB input.

5. Conclusion

We propose PolarAnything, a novel diffusion-based polarization image generation framework that generates full polarization properties. Compared to polarization image renderer Mitsuba [43], PolarAnything only takes as input a single RGB image rather than complex PBR components, which greatly reduces the barrier of large-scale polarization dataset creation. Extensive experiments show that PolarAnything generates photorealistic images while faithfully recovering physical polarization information. As demonstrated in the supplementary material, the synthesized data supports downstream tasks such as shape-from-polarization, glass segmentation, and reflection control.

Acknowledgement

This work was supported by Hebei Natural Science Foundation Project No. 242Q0101Z, Beijing-Tianjin-Hebei Basic Research Funding Program No. F2024502017, National Natural Science Foundation of China (Grant No. 62472044, U24B20155, 62225601, U23B2052, 62136001), Beijing Natural Science Foundation Project No. L242025. We thank openbayes.com for providing computing resource.

References

- [1] Gary A Atkinson and Edwin R Hancock. Recovery of surface orientation from diffuse polarization. *IEEE Transactions on Image Processing*, 2006.
- [2] Yunhao Ba, Alex Gilbert, Franklin Wang, Jinfa Yang, Rui Chen, Yiqin Wang, Lei Yan, Boxin Shi, and Achuta Kadambi. Deep shape from polarization. In *Proc. of European Conference on Computer Vision (ECCV)*, 2020.
- [3] 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 Trans. on Graph.*, 2020.
- [4] Seung-Hwan Baek, Daniel S Jeon, Xin Tong, and Min H Kim. Simultaneous acquisition of polarimetric SVBRDF and normals. *ACM Trans. on Graph.*, 2018.
- [5] Tim Brooks and Jonathan T. Barron. Learning to synthesize motion blur. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [6] Guangcheng Chen, Li He, Yisheng Guan, and Hong Zhang. Perspective phase angle model for polarimetric 3d reconstruction. In *Proc. of European Conference on Computer Vision (ECCV)*, 2022.
- [7] Akshat Dave, Yongyi Zhao, and Ashok Veeraraghavan. Pandora: Polarization-aided neural decomposition of radiance. In *Proc. of European Conference on Computer Vision (ECCV)*, 2022.
- [8] Kenji Enomoto, TJ Rhodes, Brian Price, and Gavin Miller. Polar matte: Fully computational ground-truth-quality alpha matte extraction for images and video using polarized screen matting. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [9] Xiao Fu, Wei Yin, Mu Hu, Kaixuan Wang, Yuexin Ma, Ping Tan, Shaojie Shen, Dahua Lin, and Xiaoxiao Long. GeoWizard: Unleashing the diffusion priors for 3d geometry estimation from a single image. In *Proc. of European Conference on Computer Vision (ECCV)*, 2024.
- [10] Daoyi Gao, Yitong Li, Patrick Ruhkamp, Iuliia Skobleva, Magdalena Wysocki, HyunJun Jung, Pengyuan Wang, Arturo Guridi, and Benjamin Busam. Polarimetric pose prediction. In *Proc. of European Conference on Computer Vision (ECCV)*, 2022.
- [11] Ruisheng Gao, Yutong Liu, Zeyu Xiao, and Zhiwei Xiong. Diffusion-based light field synthesis. *arXiv preprint arXiv:2402.00575*, 2024.
- [12] Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A. Wilson, and Paul E. Debevec. Circularly polarized spherical illumination reflectometry. *ACM Trans. on Graph.*, 2010.
- [13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Proc. of Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2014.
- [14] Chen Guangcheng, He Yicheng, He Li, and Zhang Hong. PISR: Polarimetric neural implicit surface reconstruction for textureless and specular objects. In *Proc. of European Conference on Computer Vision (ECCV)*, 2024.
- [15] Chen Guangcheng, He Yicheng, He Li, and Zhang Hong. Pisr: Polarimetric neural implicit surface reconstruction for textureless and specular objects. In *Proc. of European Conference on Computer Vision (ECCV)*, 2024.
- [16] Yufei Han, Heng Guo, Koki Fukai, Hiroaki Santo, Boxin Shi, Fumio Okura, Zhanyu Ma, and Yunpeng Jia. Nersp: Neural 3d reconstruction for reflective objects with sparse polarized images. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [17] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Proc. of Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.
- [18] Tomoki Ichikawa, Matthew Purri, Ryo Kawahara, Shohei Nobuhara, Kristin Dana, and Ko Nishino. Shape from sky: Polarimetric normal recovery under the sky. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [19] Tomoki Ichikawa, Yoshiki Fukao, Shohei Nobuhara, and Ko Nishino. Fresnel microfacet brdf: Unification of polariradiometric surface-body reflection. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [20] Tomoki Ichikawa, Shohei Nobuhara, and Ko Nishino. SPI-DeRS: Structured polarization for invisible depth and reflectance sensing. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [21] Kei Ikemura, Yiming Huang, Felix Heide, Zhaoxiang Zhang, Qifeng Chen, and Chenyang Lei. Robust depth enhancement via polarization prompt fusion tuning, 2024.
- [22] LUCID Vision Labs Inc. Triton 5.0 mp polarization model (imx250myr). <https://thinklucid.com/product/triton-5-mp-polarization-camera/>, 2018. Accessed: 2025-03-04.
- [23] HyunJun Jung, Patrick Ruhkamp, Guangyao Zhai, Nikolas Brasch, Yitong Li, Yannick Verdie, Jifei Song, Yiren Zhou, Anil Armagan, Slobodan Ilic, Ales Leonardis, and Benjamin Busam. Is my depth ground-truth good enough? HAMMER - highly accurate multi-modal dataset for dense 3d scene regression. *CoRR*, 2022.
- [24] HyunJun Jung, Guangyao Zhai, Shun-Cheng Wu, Patrick Ruhkamp, Hannah Schieber, Pengyuan Wang, Giulia Rizzoli, Hongcheng Zhao, Sven Damian Meier, Daniel Roth, Nassir Navab, et al. HouseCat6D—a large-scale multi-modal category level 6d object perception dataset with household objects in realistic scenarios. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [25] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Depth sensing using geometrically constrained po-

- larization normals. *International Journal of Computer Vision*, 2017.
- [26] Agastya Kalra, Vage Taamazyan, Sreeth Krishna Rao, Kartik Venkataraman, Ramesh Raskar, and Achuta Kadambi. Deep polarization cues for transparent object segmentation. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [27] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [28] Yuhki Kondo, Taishi Ono, Legong Sun, Yasutaka Hirasawa, and Jun Murayama. Accurate polarimetric brdf for real polarization scene rendering. In *Proc. of European Conference on Computer Vision (ECCV)*, 2020.
- [29] Naejin Kong, Yu-Wing Tai, and Joseph S. Shin. A physically-based approach to reflection separation: From physical modeling to constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.
- [30] Zhengfei Kuang, Yunzhi Zhang, Hong-Xing Yu, Samir Agarwala, Elliott Wu, Jiajun Wu, et al. Stanford-ORB: a real-world 3d object inverse rendering benchmark. In *Proc. of Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2023.
- [31] Chenyang Lei, Xuhua Huang, Mengdi Zhang, Qiong Yan, Wenxiu Sun, and Qifeng Chen. Polarized reflection removal with perfect alignment in the wild. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [32] Chenyang Lei, Chenyang Qi, Jiaxin Xie, Na Fan, Vladlen Koltun, and Qifeng Chen. Shape from polarization for complex scenes in the wild. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [33] Chenhao Li, Taishi Ono, Takeshi Uemori, Hajime Mihara, Alexander Gatto, Hajime Nagahara, and Yusuke Moriuchi. Neisf: Neural incident stokes field for geometry and material estimation. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [34] Feiran Li, Heng Guo, Hiroaki Santo, Fumio Okura, and Yasuyuki Matsushita. Learning to synthesize photorealistic dual-pixel images from rgbd frames. In *Proc. of IEEE International Conference on Computational Photography (ICCP)*, 2023.
- [35] Yuan Liu, Peng Wang, Cheng Lin, Xiaoxiao Long, Jiepeng Wang, Lingjie Liu, Taku Komura, and Wenping Wang. Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images. *ACM Trans. on Graph.*, 2023.
- [36] Agisoft LLC. Agisoft photoscan standard (version 1.2.6). <https://www.agisoft.com/downloads/installer/>, 2016. Accessed: 2025-03-04.
- [37] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *Proc. of International Conference on Learning Representations (ICLR)*, 2019.
- [38] Youwei Lyu, Zhaopeng Cui, Si Li, Marc Pollefeys, and Boxin Shi. Physics-guided reflection separation from a pair of unpolarized and polarized images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [39] Youwei Lyu, Lingran Zhao, Si Li, and Boxin Shi. Shape from polarization with distant lighting estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [40] Haiyang Mei, Bo Dong, Wen Dong, Jiaxi Yang, Seung-Hwan Baek, Felix Heide, Pieter Peers, Xiaopeng Wei, and Xin Yang. Glass segmentation using intensity and spectral polarization cues. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [41] Miyazaki, Tan, Hara, and Ikeuchi. Polarization-based inverse rendering from a single view. In *Proc. of International Conference on Computer Vision (ICCV)*, 2003.
- [42] Miki Morimatsu, Yusuke Monno, Masayuki Tanaka, and Masatoshi Okutomi. Monochrome and color polarization demosaicking using edge-aware residual interpolation. In *Proc. of IEEE International Conference on Image Processing (ICIP)*, 2020.
- [43] Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. Mitsuba 2: a retargetable forward and inverse renderer. *ACM Trans. on Graph.*, 2019.
- [44] Simeng Qiu, Qiang Fu, Congli Wang, and Wolfgang Heidrich. Linear polarization demosaicking for monochrome and colour polarization focal plane arrays. *Computer Graphics Forum*, 2021.
- [45] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proc. of the International Conference on Machine Learning (ICML)*, 2021.
- [46] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [47] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *Proc. of SIGGRAPH*, 2022.
- [48] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Lit, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Raphael Gontijo-Lopes, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. Photorealistic text-to-image diffusion models with deep language understanding. In *Proc. of Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.
- [49] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. Laion-5b: an open large-scale dataset for training next generation image-text models. In *Proc. of Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.
- [50] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation

- for non-lambertian and uncalibrated photometric stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [51] William AP Smith, Ravi Ramamoorthi, and Silvia Tozza. Height-from-polarisation with unknown lighting or albedo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
- [52] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proc. of the International Conference on Machine Learning (ICML)*, 2015.
- [53] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *Proc. of International Conference on Learning Representations (ICLR)*, 2021.
- [54] Pratul P. Srinivasan, Tongzhou Wang, Ashwin Sreelal, Ravi Ramamoorthi, and Ren Ng. Learning to synthesize a 4d rgbd light field from a single image. In *Proc. of International Conference on Computer Vision (ICCV)*, 2017.
- [55] Chaoran Tian, Weihong Pan, Zimo Wang, Mao Mao, Guofeng Zhang, Hujun Bao, Ping Tan, and Zhaopeng Cui. DPS-Net: Deep polarimetric stereo depth estimation. In *Proc. of International Conference on Computer Vision (ICCV)*, 2023.
- [56] Lezhong Wang, Jeppe Revall Frisvad, Mark Bo Jensen, and Siavash Arjomand Bigdeli. Stereodiffusion: Training-free stereo image generation using latent diffusion models. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2024.
- [57] Patrick Wieschollek, Orazio Gallo, Jinwei Gu, and Jan Kautz. Separating reflection and transmission images in the wild. In *Proc. of European Conference on Computer Vision (ECCV)*, 2018.
- [58] Chongjie Ye, Lingteng Qiu, Xiaodong Gu, Qi Zuo, Yushuang Wu, Zilong Dong, Liefeng Bo, Yuliang Xiu, and Xiaoguang Han. StableNormal: Reducing diffusion variance for stable and sharp normal. *ACM Trans. on Graph.*, 2024.
- [59] 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 *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [60] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proc. of International Conference on Computer Vision (ICCV)*, 2023.
- [61] Jinyu Zhao, Yusuke Monno, and Masatoshi Okutomi. Polarimetric multi-view inverse rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [62] Chu Zhou, Minggui Teng, Yufei Han, Chao Xu, and Boxin Shi. Learning to dehaze with polarization. In *Proc. of Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- [63] Yi Zhou, Connell Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.