PolarAnything: Diffusion-based Polarimetric Image Synthesis

Kailong Zhang[1†] Youwei Lyu[1†] Heng Guo[1,2*] Si Li[1] Zhanyu Ma[1] Boxin Shi[3,4]

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

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

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

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

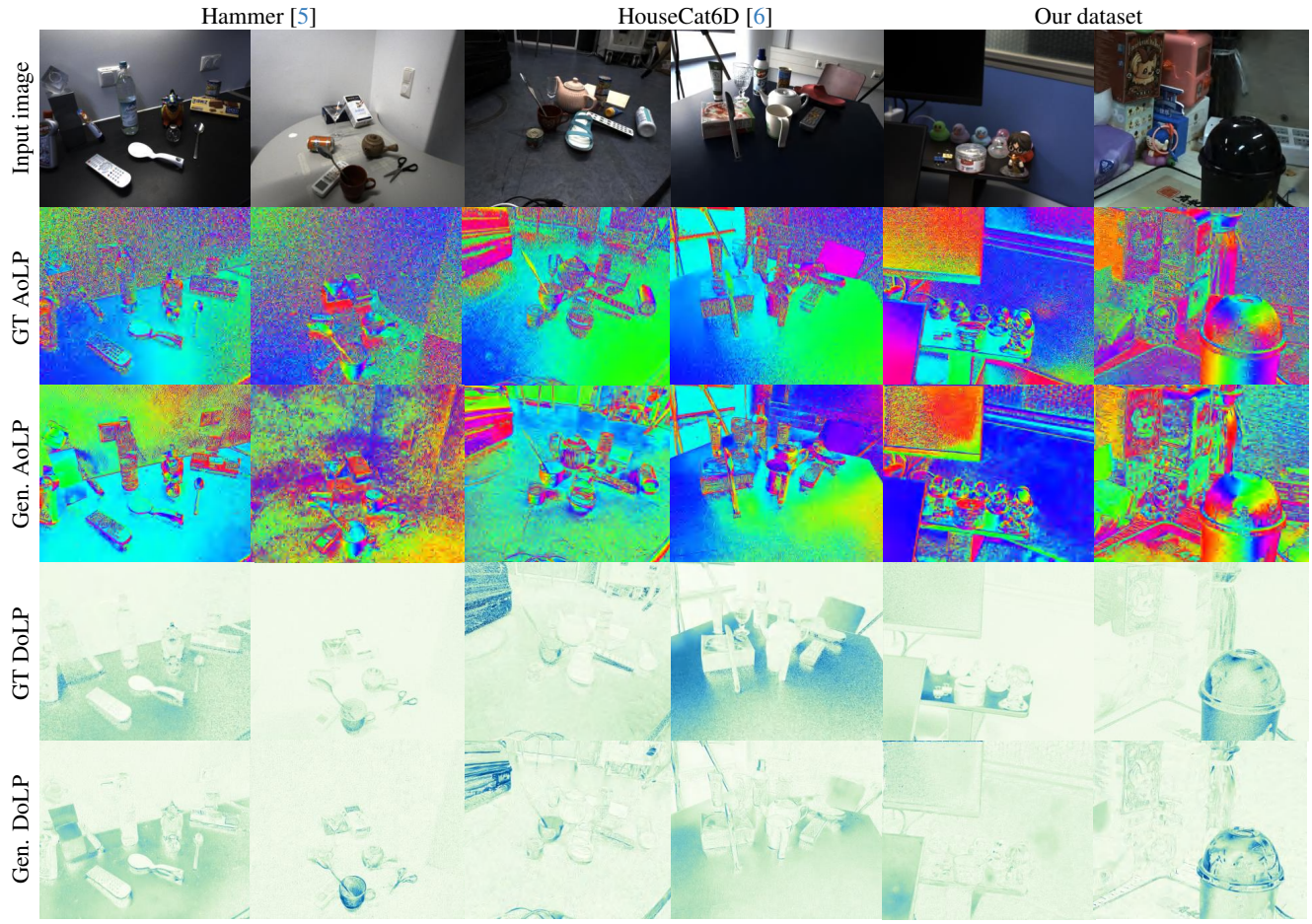

Figure 1. Additional qualitative evaluation of PolarAnything on scene-level polarization image generation, evaluated on Hammer [5], HouseCat6D [6], and our dataset.

1. Scene-level polarization image synthesis

To further assess the generalization ability of our model, we conducted scene-level evaluations on publicly available datasets, including HAMMER [5], HouseCat6D [6], as well as our own captured data. Figure 1 shows that PolarAnything is capable of synthesizing correct polarization information across various light conditions and objects effectively, demonstrating the generalization performance of our method.

2. Failure case

As shown in Fig. 2, we show failure cases of our method, where the generated Angle of Linear Polarization (AoLP) maps are different from the real-captured ones on surfaces

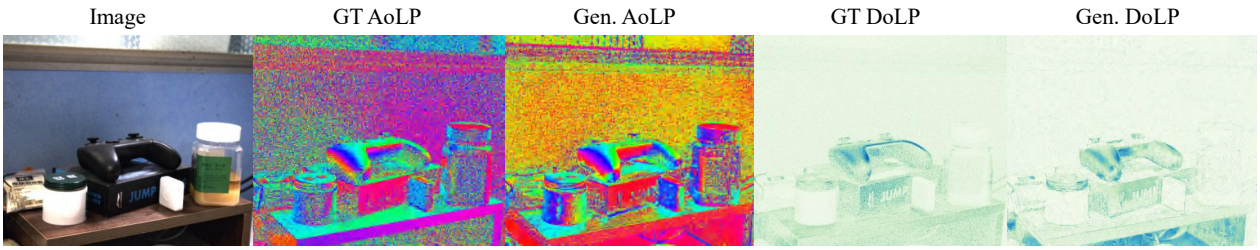

Figure 2. Failure cases of PolarAnything. Due to low DoLP on the background scene, the polarization information is not reliable on both synthesized and real-captured polarization images, considering the sensor noise. Under such a case, the AoLP maps from ours and GT could be different.

with diffuse reflectances. We conduct analysis as follows.

We first present the formula of AoLP Φ and Degree of Linear Polarization (DoLP) $\mathbf{P}$:

$$\Phi = \frac{1}{2} atan2 \left(\frac{S_1}{S_2} \right), \tag{1}$$

$$\mathbf{P} = \frac{\sqrt{S_1^2 + S_2^2}}{S_0}, \tag{2}$$

where S_0, S_1, and S_2 are components of the Stokes vector used to describe the state of polarization of light. Specifically: S_0 represents the total intensity of the light, which is the sum of all light components, including both polarized and unpolarized light. S_1 represents the difference in intensities between light polarized horizontally and vertically. It captures the linear polarization information along the horizontal and vertical axes. S_2 represents the difference in intensities between light polarized at +45° and -45°, capturing the linear polarization information at these diagonal angles.

For rough surfaces, the DoLP is typically small and approaches zero. Following (2), we have $S_1 \approx S_2 \approx 0$, indicating minimal polarization. However, real-world measurements are influenced by noise. To account for this, we introduce two noise terms, ϵ_1 and ϵ_2, both following a Gaussian distribution: $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0, \sigma^2)$.

Following (1), the AoLP under noise can be denoted as

$$\Phi = \frac{1}{2} atan2 \left(\frac{S_1 + \epsilon_1}{S_2 + \epsilon_2} \right). \tag{3}$$

As $S_1 \approx S_2 \approx 0$, AoLP exhibits significant variance under noise, making the AoLP estimates unstable and unreliable on rough surfaces. Under such cases, both real-captured and our synthesized polarization images are noisy, which could be the possible reason why our results are different from the real-captured ones, as shown in Fig. 2.

3. Captured setup of PolarAnything dataset

The complete shooting process of our dataset is shown in Fig. 3. The camera we selected is the Triton 5.0 MP Polarization Model (Sony IMX250MYR) [4]. For object selection, we chose approximately 100 objects representing a

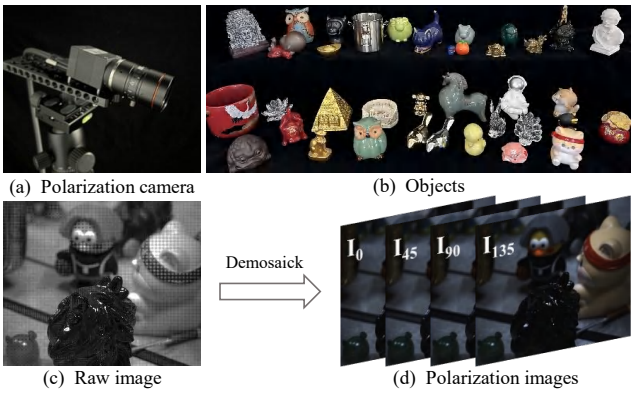

Figure 3. Capture process of PolarAnything dataset.

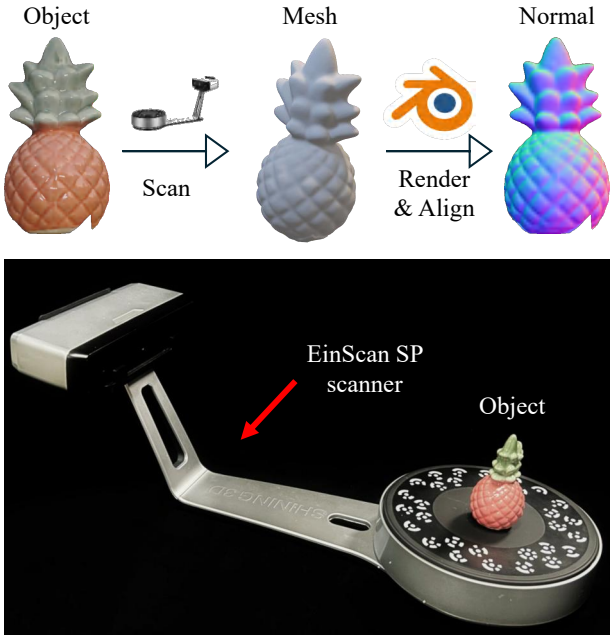

Figure 4. Process of obtaining the "GT" surface normal map.

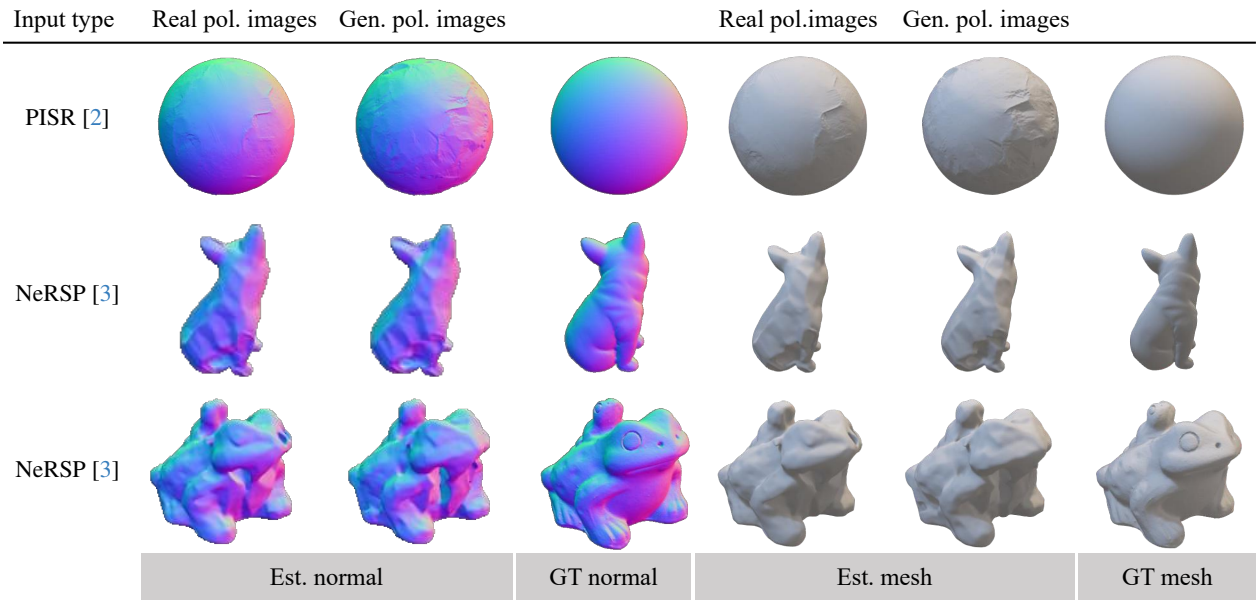

Figure 5. Evaluating PolarAnything on multiview shape from polarization methods PISR [2] and NeRSP [3] on data from [3].

variety of materials. During the shooting process, we aimed to include at least 2-3 different objects in each frame to maximize the amount of useful information captured. Our dataset includes a total of 1,146 polarization images. The final four polarization images (corresponding to the polarization angles of $0°$, $45°$, $90°$, and $135°$) are demosaicked from the captured raw images, using a simple bilinear interpolation algorithm.

4. Captured setup of PN dataset

The PN dataset mentioned in Sec.4.3 was acquired as follows: (a) For obtaining the polarization images, we used a polarized camera [4] under ambient lighting conditions to capture them directly. (b) For capturing the "GT surface normal map", we followed the methodology of DiliGenT [9]. As illustrated in Fig. 4, we first scanned the surface geometry using an EinScan SP scanner[1]. Then, we imported the scanned mesh and the captured images into Blender [1] for alignment. Once the calibration was completed, we exported the object's normal map as the corresponding GT.

5. More results about SfP

Besides the results of PISR [2] shown in Sec.4.3, we present more 3D reconstruction results from other MVSfP methods including PISR [2] and NeRSP [3]. As shown in Fig. 5, we add additional results on BALL, DOG, and FROG. The reconstruction surface normals and meshes using data from PolarAnything and real-captured polarization images are comparable, demonstrating that PolarAnything is capable of producing physically plausible polarization properties and benefiting downstream tasks.

[1]https://www.einscan.com/einscan-sp/

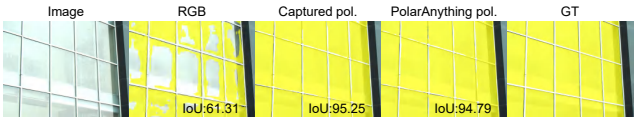

Figure 6. Glass segmentation results of PGSNet on the RGB image, the captured polarization images, and generated data by PolarAnything.

6. Application: Glass Segmentation

To assess whether synthetic polarization images could effectively support downstream tasks, we conducted an evaluation on glass segmentation using PGSNet [7], a model specifically designed to leverage polarization cues. We tested three types of input: (1) captured polarization images, (2) RGB-only images without polarization information, and (3) polarization images generated by PolarAnything. As shown in Fig. 6, the model achieves comparable performance when using synthetic and real polarization inputs, both of which significantly outperform the RGB-only setting. These results suggest that the generated polarization images not only capture meaningful polarization features but could also serve as a practical alternative when real polarization data is unavailable.

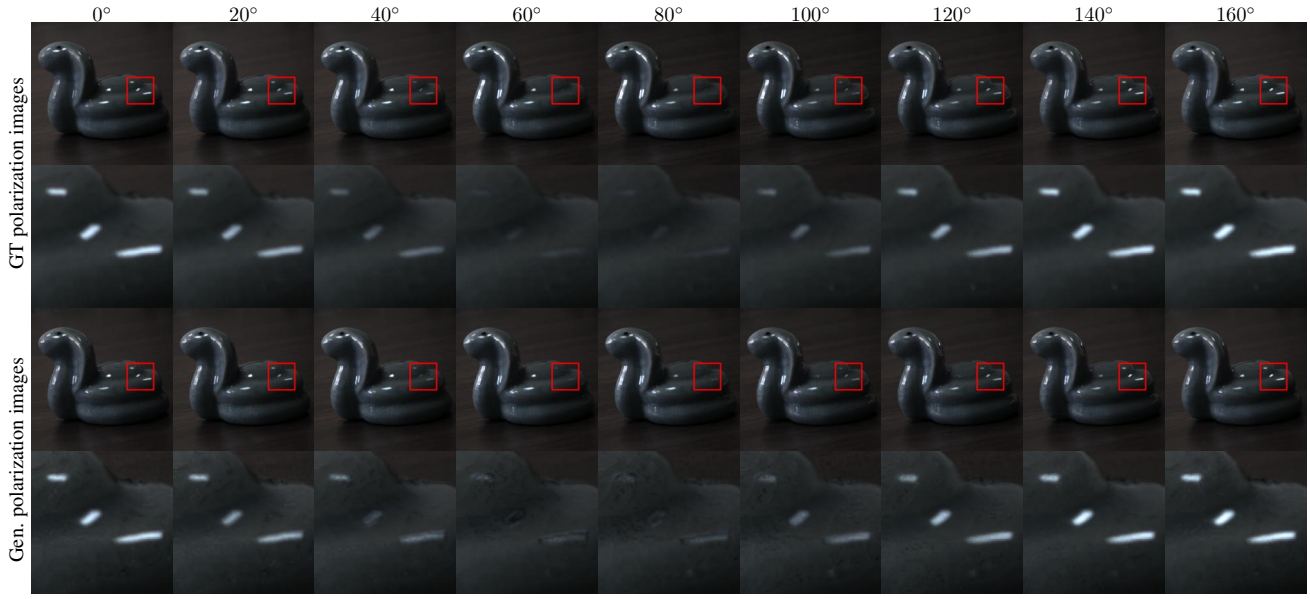

Figure 7. Polarization images synthesized at different angles. Top two rows use ground-truth polarization; bottom two rows use PolarAnything predictions. Reflections are clearly reduced in highly reflective regions.

7. Application: Reflection Control

Polarization images have been proven effective in analyzing material properties, surface geometry, and reflectance characteristics, due to their sensitivity to the polarization state of light. One particularly valuable application is the suppression of specular reflections, especially in scenes containing highly reflective surfaces. This capability is physically grounded in the relationship between the polarization angle Θ, the Angle of Linear Polarization (AoLP, denoted as Φ), the Degree of Linear Polarization (DoLP, denoted as $\mathbf{P}$), and the unpolarized RGB image $\mathbf{I}_{\text{RGB}}$, as previously introduced in the main paper and formulated by the following equation:

$$\mathbf{I}_{\Theta} = \frac{\mathbf{I}_{\text{RGB}}}{2} \left(1 + \mathbf{P} \cos\left(2\Theta - 2\Phi\right)\right), \tag{4}$$

With AoLP and DoLP predicted by PolarAnything, combined with the RGB image, we can synthesize polarization images at arbitrary angles Θ. This enables flexible control over the polarization state without requiring additional physical captures. We demonstrate this functionality on a highly reflective ceramic object. As shown in Fig. 7, the synthesized images at specific angles can significantly reduce specular highlights.

These results show that our method can effectively reduce reflections by using the predicted polarization parameters to synthesize polarization images. This ability broadens the use of polarization techniques, making it possible to control reflections dynamically. The control can be performed flexibly, without the need to set specific polarization angles during image capture.

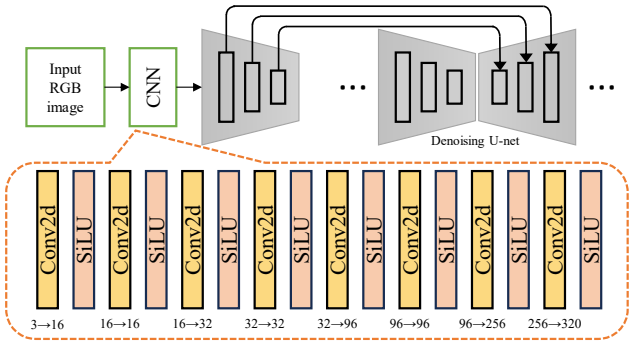

Figure 8. Feature extraction module for condition information.

8. More details about implementation

As described in the main text, we fine-tuned the diffusion model Stable Diffusion v1.5 [8] and added a CNN module for feature extraction. The architecture of the CNN network is shown in Fig. 8. During model training, making the original U-net parameters learnable was found to be beneficial for the network's learning process. The results are presented in Fig. 9 and Table 1.

In the VAE input module, all input images are normalized to the range of [-1, 1], which is consistent with the input range expected by the VAE. The normalization is performed based on the maximum pixel value of the current image. We found that this dynamic normalization method significantly improves the network's generalization ability.

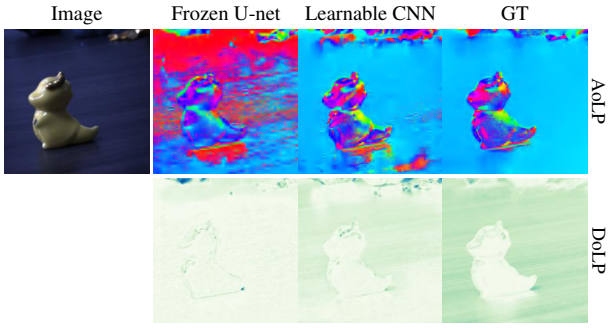

Figure 9. Results on different training strategies.

Table 1. Results produced by different training strategies.

Training Strategy	PSNR↑	SSIM↑	MAngE↓	MAbsE↓
Frozen U-net	36.82	0.9842	30.12	0.1922
Learnable CNN	**41.74**	**0.9927**	**25.33**	**0.1075**

References

[1] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.