

Supplementary file to ‘‘Polarization Uncertainty-Guided Diffusion Model for Color Polarization Image Demosaicking’’

Chenggong Li^{1,2}, Yidong Luo^{3,4}, Junchao Zhang^{1,2,*}, Degui Yang^{1,2,*}

¹School of Automation, Central South University, Changsha, China

²Hunan Provincial Key Laboratory of Optic-Electronic Intelligent Measurement and Control, Changsha, China

³Zhejiang University, Hangzhou, China

⁴School of Engineering, Westlake University, Hangzhou, China

*Corresponding Author

{244603040, junchaozhang, degui.yang}@csu.edu.cn, luoyidong@westlake.edu.cn

In this supplementary file, we first present the detailed structure of the SD branch, followed by a supplementary derivation of the polarization uncertainty model. We then explain how the intensity uncertainty is obtained in the ablation experiments, and finally provide additional visual comparison results.

1 The detailed structure of the SD branch

The SD consists of a variational autoencoder (VAE) and a latent diffusion model (LDM) shown in Fig. 1. Since SD processes natural images, we treat the four direction intensity images as four separate batches and feed them into the SD model sequentially, i.e., $x_b \in \mathbb{R}^{12 \times H \times W} \rightarrow x_b \in \mathbb{R}^{4 \times 3 \times H \times W}$, where the first dimension corresponds to the batch. The data flow of SD can be expressed as:

$$\begin{aligned} z_b &= E(x_b), \\ z_{sd} &= LDM(z_b), \\ x_{sd} &= D(z_{sd}), \end{aligned} \quad (1)$$

where E and D denote the VAE encoder and decoder, respectively, with $z_b \in \mathbb{R}^{4 \times 4 \times \frac{H}{8} \times \frac{W}{8}}$ and $z_{sd} \in \mathbb{R}^{4 \times 4 \times \frac{H}{8} \times \frac{W}{8}}$ representing the image features in the latent space. The LDM obtains the final result using only one step, which can be expressed as:

$$z_{sd} = LDM(z_b) \triangleq \frac{z_b - \beta_T \psi(z_b, T)}{\alpha_T}, \quad (2)$$

where T is the time, α_T and β_T denotes the noise scheduler of the SD, and ψ represents the diffusion U-Net. In the context of CPDM, text prompts are unnecessary for the SD, hence, we remove the text encoder and cross-attention module in the diffusion U-Net, improving efficiency without compromising demosaicking performance.

Instead of training from scratch, we follow Wu et al.’s approach (Wu et al. 2024) by injecting LoRA modules into the VAE and diffusion U-Net, and fine-tune them on simulated datasets to accommodate the CPDM task. LoRA updates only a subset of parameters during training, which can be expressed as:

$$Q_0 + \Delta Q = Q_0 + UL, U \in \mathbb{R}^{d \times r}, L \in \mathbb{R}^{r \times e}, \quad (3)$$

where $Q_0 \in \mathbb{R}^{d \times e}$ denotes the pre-trained weights, U and L are the two low-rank matrices to be trained, with $r \ll \min(d, e)$. Training with LoRA effectively preserves the diffusion prior of the SD, enabling the network to leverage additional knowledge to recover lost pixels. This helps overcome the performance bottleneck caused by limited CPDM datasets and further enhance the models’ generalization capability.

As shown in Eq. (1) and Fig. 1, encoding from image space to the latent space involves lossy information compression, making it difficult to directly use SD to meet the high-fidelity requirements of the CPDM task. Therefore, we aim to selectively leverage the diffusion prior in SD: expecting it to improve only regions with large polarization reconstruction errors, while relying on a conventional network for high-fidelity reconstruction in low-error regions. This forms a key motivation behind our design of the dual-branch architecture and the polarization uncertainty model.

2 The supplementary derivation of the polarization uncertainty model

2.1 Polarization Uncertainty Model

Inspired by uncertainty analysis in the super-resolution task (Ning et al. 2021; Zhang et al. 2025), we design a polarization uncertainty model tailored for the color polarization demosaicking (CPDM) task. We begin by modelling the reconstruction process from the four direction intensity images, which can be formulated as:

$$x_i = f_\theta(\tilde{x}_i) + \varepsilon\eta, i = 0^\circ, 45^\circ, 90^\circ, 135^\circ, \quad (4)$$

where η is the intensity uncertainty, and ε follows the normal distribution with zero-mean and unit-variance, the subscript i indicates the polarization images at different directions, for example, x_{0° and $f_\theta(\tilde{x}_{0^\circ})$ denote the ground truth and the network’s reconstructed result for the 0° intensity image, respectively. Eq. (4) can be expressed as $x_i \sim N(f_\theta(\tilde{x}_i), \eta^2)$. According to Collett’s literature (Collett 2005), the Stokes

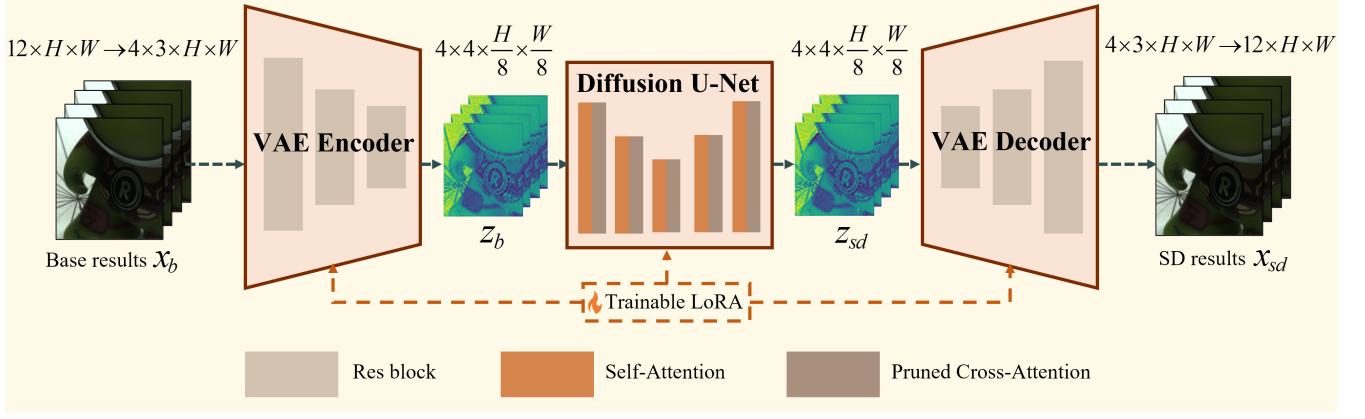


Figure 1: The architecture of the SD branch. The output of the base branch is first split along the channel dimension to separate the four direction intensity images into individual samples, which are then fed into the SD model in a batch-wise manner. The VAE encoder first compresses each image into the latent space, after which the diffusion U-Net refines the latent features. Finally, the VAE decoder reconstructs the refined features back into the image space.

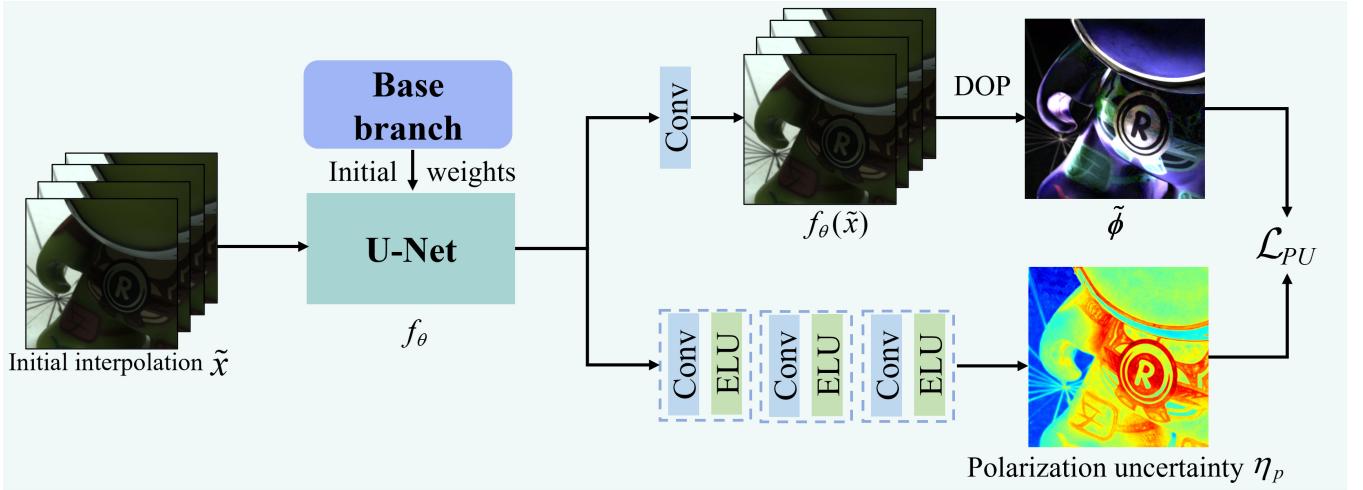


Figure 2: The architecture of the uncertainty estimation network. The network's backbone shares the same architectures as the base branch, but augmented with an additional estimation head to output polarization uncertainty.

parameters are defined as:

$$\begin{aligned} S_0 &= \frac{1}{2}(x_{0^\circ} + x_{45^\circ} + x_{90^\circ} + x_{135^\circ}) \\ S_1 &= x_{0^\circ} - x_{90^\circ} \\ S_2 &= x_{45^\circ} - x_{135^\circ} \\ DOP &= \frac{\sqrt{S_1^2 + S_2^2}}{S_0} \\ AOP &= \frac{1}{2} \arctan\left(\frac{S_2}{S_1}\right) \end{aligned} \quad . \quad (5)$$

Substituting Eq. (4) into Eq. (5), we can easily derive the form of $S_{0,1,2}$ using the properties of the normal distribution:

$$S_0 \sim N(\tilde{S}_0, \eta^2), S_1 \sim N(\tilde{S}_1, 2\eta^2), S_2 \sim N(\tilde{S}_2, 2\eta^2), \quad (6)$$

where

$$\begin{aligned} \tilde{S}_0 &\triangleq \frac{1}{2}(f_\theta(\tilde{x}_{0^\circ}) + f_\theta(\tilde{x}_{45^\circ}) + f_\theta(\tilde{x}_{90^\circ}) + f_\theta(\tilde{x}_{135^\circ})) \\ \tilde{S}_1 &\triangleq f_\theta(\tilde{x}_{0^\circ}) - f_\theta(\tilde{x}_{90^\circ}), \tilde{S}_2 \triangleq f_\theta(\tilde{x}_{45^\circ}) - f_\theta(\tilde{x}_{135^\circ}) \end{aligned} \quad . \quad (7)$$

Since f_θ has already reconstructed the intensity image (S_0) effectively, we can assume that the estimate of S_0 is accurate, and thus use \tilde{S}_0 as an approximation of the ground truth S_0 , i. e. $\tilde{S}_0 \triangleq S_0$. This conclusion is also adopted in the polarization noise model (Hwang et al. 2025). Subsequently, we can derive:

$$\frac{S_1}{S_0} \sim N\left(\frac{\tilde{S}_1}{\tilde{S}_0}, \frac{2\eta^2}{\tilde{S}_0^2}\right), \frac{S_2}{S_0} \sim N\left(\frac{\tilde{S}_2}{\tilde{S}_0}, \frac{2\eta^2}{\tilde{S}_0^2}\right). \quad (8)$$

Based on Eqs. (5) and (8), the DOP (denoted as ϕ) follows a Rice distribution of the form:

$$\phi \sim \text{Rice}\left(\tilde{\phi}, \frac{\sqrt{2}\eta}{\tilde{S}_0}\right), \quad (9)$$

and we denote $\frac{\sqrt{2}\eta}{S_0}$ as η_p , which represents the polarization uncertainty we aim to model in this paper. Compared to the intensity uncertainty η , η_p provides more direct polarization information for the subsequent fusion process.

The probability density function (PDF) of Eq. (9) can be expressed as:

$$P(\phi|\tilde{\phi}, \eta_p) = \frac{\phi}{\eta_p^2} \exp\left(-\frac{(\phi^2 + \tilde{\phi}^2)}{2\eta_p^2}\right) I_0\left(\frac{\phi\tilde{\phi}}{\eta_p^2}\right), \quad (10)$$

where I_0 denotes the modified Bessel function of the first kind, order zero. When the order of the Bessel function k is non-negative, the modified Bessel function I_k has the following asymptotic form for large variable $v \gg |k^2 - \frac{1}{4}|$ (Bowman 2012):

$$I_k(v) \rightarrow \frac{1}{\sqrt{2\pi v}} \exp(v). \quad (11)$$

Since DOP is an image with dominant signal and network can reliably recover its main component, the ratio of signal to uncertainty $\frac{\phi\tilde{\phi}}{\eta_p^2} \gg \frac{1}{4}$. Therefore, Eq. (10) can be simplified using Eq. (11) as:

$$\begin{aligned} P(\phi|\tilde{\phi}, \eta_p) &= \frac{\phi}{\eta_p^2} \exp\left(-\frac{(\phi^2 + \tilde{\phi}^2)}{2\eta_p^2}\right) I_0\left(\frac{\phi\tilde{\phi}}{\eta_p^2}\right) \\ &\approx \frac{\phi}{\eta_p^2} \exp\left(-\frac{(\phi^2 + \tilde{\phi}^2)}{2\eta_p^2}\right) \frac{\eta_p}{\sqrt{2\pi\phi\tilde{\phi}}} \exp\left(\frac{\phi\tilde{\phi}}{\eta_p^2}\right) \\ &= \frac{\sqrt{\phi}}{\eta_p \sqrt{2\pi\tilde{\phi}}} \exp\left(-\frac{\phi^2 + \tilde{\phi}^2 - 2\phi\tilde{\phi}}{2\eta_p^2}\right) \\ &= \sqrt{\frac{\phi}{2\pi\phi\eta_p^2}} \exp\left(-\frac{(\phi - \tilde{\phi})^2}{2\eta_p^2}\right) \end{aligned} \quad . \quad (12)$$

To obtain η_p , we use an uncertainty estimation network to directly predict it through supervised learning shown in Fig. 2. The uncertainty estimation network shares the same architecture as the base branch, but the two are independent. In the fusion process, we only use the uncertainty predicted by this network, discarding its reconstructed image output. The optimization objective is to minimize the negative log likelihood of Eq. (12). Inspired by Ning's paper (Ning et al. 2021), we incorporate Jeffrey's prior (Figueiredo 2001) $P(\eta_p) \propto \frac{1}{\eta_p}$, and the optimization objective can be written as:

$$\begin{aligned} \mathcal{L}_{PU} &= -\ln(P(\phi|\tilde{\phi}, \eta_p)P(\eta_p)) \\ &= -\ln\left(\sqrt{\frac{\phi}{2\pi\phi\eta_p^2}} \exp\left(-\frac{(\phi - \tilde{\phi})^2}{2\eta_p^2}\right)\right) \\ &= 2\ln\eta_p + \frac{1}{2}\ln 2\pi + \frac{1}{2}(\ln\tilde{\phi} - \ln\phi) + \frac{1}{2\eta_p^2} \left\| \tilde{\phi} - \phi \right\|_2^2 \\ &= \frac{1}{2}(\ln\tilde{\phi} - \ln\phi) + 2s + \frac{1}{2} \exp(-2s) \left\| \tilde{\phi} - \phi \right\|_2^2 \end{aligned} \quad (13)$$

where s denotes the log polarization uncertainty $s = \ln\eta_p$, and we omit the constant terms that do not contribute to the gradient updates.

2.2 Derivation of intensity uncertainty

In the ablation study in the main text, we compare intensity uncertainty derived from the four direction polarization intensity image x_i , ($i = 0^\circ, 45^\circ, 90^\circ, 135^\circ$) and total intensity image S_0 , respectively. According to Eqs. (4) and (6), both four direction images and S_0 follows a normal distribution and their uncertainty are theoretically consistent. Therefore,

Table 1: Complexity analysis on 4090 GPU with $12 \times 512 \times 512$ inputs.

Method	Params (M)	FLOPs (G)	Inference time (s)
DCPM	143.44	13060.63	0.42
PIDSR	7.45	168.15	0.14
Ours	899.19	1884.22	0.36

we present the derivation of S_0 only, and the derivation of the four direction case is analogous.

The probability density function (PDF) of S_0 can be expressed as:

$$P(S_0|\tilde{S}_0, \eta) = \frac{1}{\sqrt{2\pi\eta}} \exp\left(-\frac{(S_0 - \tilde{S}_0)^2}{2\eta^2}\right), \quad (14)$$

here, we aim to estimate the uncertainty of S_0 and therefore do not apply the approximation in Eq. (8), which is introduced to derive the distribution of DOP with increased uncertainty. Similar to Eq. (13), the optimization objective is the negative log-likelihood of Eq. (14), i.e.:

$$\mathcal{L}_{IU} = 2\ln\eta + \frac{1}{2\eta^2} \left\| \tilde{S}_0 - S_0 \right\|^2. \quad (15)$$

We still use the estimation network shown in Fig. 2 to estimate the intensity uncertainty. The loss for the four direction intensity image follows the same form as Eq. (15), except that in implementation, the loss for the four direction case is averaged over the batch, whereas for S_0 , the four directions are first averaged to compute S_0 , and then the loss is calculated.

3 More visual comparison

We present additional visual comparison in Fig. 3. As an initial interpolation method, Polanalyser (Maeda 2019) performs poorly in recovering polarization characteristics. NLCSR (Luo et al. 2024) produces over-smoothed results with severe loss of details in S_0 . CPDNet (Wen et al. 2019), TCPDNet (Nguyen et al. 2022), DCPM (Li et al. 2025) and PIDSR (Zhou et al. 2025), which trained with limited data priors, achieve partial reconstruction but still exhibit significant errors in DOP and AOP. In contrast, our method leverages the diffusion prior to break through the limitations of existing data distributions and incorporates polarization uncertainty estimation for targeted reconstruction, achieving the clearest and most accurate polarization characteristics. Notably, due to noise present in the ground truth, our method even produces visually superior results compared to the ground truth.

To fully demonstrate the generalization of our method, we evaluate it on real-world mosaic scenes, with visual comparisons shown in Fig. 4. Our methods outperforms the baseline in both noise suppression and texture preservation, thanks to the diffusion model's strong generative capability for recovering missing pixels.

4 Complexity analysis

As shown in Table 1, our method uses SD's priors within its huge parameters for better recovery, while SD's latent dif-



Figure 3: Visual comparisons for CPDM of different methods. our method achieves the most outstanding polarization reconstruction performance in terms of AOP and DOP.

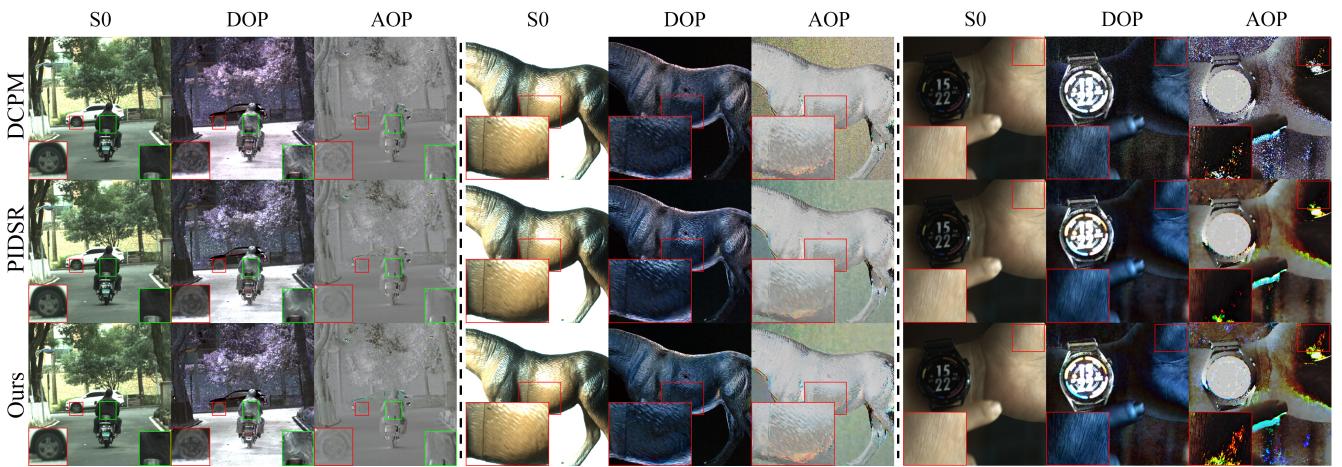


Figure 4: Comparison with strong baseline on real-world scenes. Our method produces results with less noise and richer texture.

fusion ($8\times$ downsample) and efficient architecture keep inference time at the same level as PIDS. In Section 3.1, we pruned SD by cutting the text encoder and cross-attention, reducing parameters by 30% and improving inference by 10% without performance drop. For training, we used LoRA (rank 4), only 1% of the parameters are trained. The above analysis shows that our method is suitable for consumer-grade GPUs.

5 Limitations

The proposed PUGDiff is a meaningful attempt to apply generative models to CPDM. It still has limitations, e.g., the wheel in S_0 (Fig. 4) is blurry due to information loss from SD’s latent-space sampling, even with compensation from the base branch. On the other hand, although our diffusion branch requires only one sampling step, its large parameter size and computational demands make it challenging to deploy on edge devices. To address these issues, we will continue to research more efficient VAE methods that reduce polarization information loss and explore model pruning techniques to accelerate practical deployment.

References

- Bowman, F. 2012. *Introduction to Bessel functions*. Courier Corporation.
- Collett, E. 2005. *Field guide to polarization*, volume 15. SPIE press Bellingham.
- Figueiredo, M. 2001. Adaptive sparseness using Jeffreys prior. *Advances in neural information processing systems*, 14.
- Hwang, I.; Choi, K.; Ha, H.; and Kim, M. H. 2025. Benchmarking Burst Super-Resolution for Polarization Images: Noise Dataset and Analysis. *arXiv preprint arXiv:2503.18705*.
- Li, C.; Luo, Y.; Wu, C.; Zhang, J.; Yang, D.; and Zhao, D. 2025. Demosaicking customized diffusion model for snapshot polarization imaging. *Optics & Laser Technology*, 188: 112868.
- Luo, Y.; Zhang, J.; Shao, J.; Tian, J.; and Ma, J. 2024. Learning a non-locally regularized convolutional sparse representation for joint chromatic and polarimetric demosaicking. *IEEE Transactions on Image Processing*.
- Maeda, R. 2019. Polanalyser: Polarization Image Analysis Tool.
- Nguyen, V.; Tanaka, M.; Monno, Y.; and Okutomi, M. 2022. Two-step color-polarization demosaicking network. In *2022 IEEE International Conference on Image Processing (ICIP)*, 1011–1015. IEEE.
- Ning, Q.; Dong, W.; Li, X.; Wu, J.; and Shi, G. 2021. Uncertainty-driven loss for single image super-resolution. *Advances in Neural Information Processing Systems*, 34: 16398–16409.
- Wen, S.; Zheng, Y.; Lu, F.; and Zhao, Q. 2019. Convolutional demosaicing network for joint chromatic and polarimetric imagery. *Optics letters*, 44(22): 5646–5649.
- Wu, R.; Sun, L.; Ma, Z.; and Zhang, L. 2024. One-step effective diffusion network for real-world image super-resolution. *Advances in Neural Information Processing Systems*, 37: 92529–92553.
- Zhang, L.; You, W.; Shi, K.; and Gu, S. 2025. Uncertainty-guided Perturbation for Image Super-Resolution Diffusion Model. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 17980–17989.
- Zhou, S.; Zhou, C.; Lyu, Y.; Guo, H.; Ma, Z.; Shi, B.; and Sato, I. 2025. PIDS: Complementary Polarized Image Demosaicing and Super-Resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 16081–16090.