

# Physics-Guided ISO-Dependent Sensor Noise Modeling for Extreme Low-Light Photography (Supplementary Material)

Yue Cao<sup>1</sup>, Ming Liu<sup>1</sup>(✉), Shuai Liu, Xiaotao Wang, Lei Lei, Wangmeng Zuo<sup>1,2</sup>  
<sup>1</sup>School of Computer Science and Technology, Harbin Institute of Technology, China

<sup>2</sup>Peng Cheng Laboratory, China

cscaoyue@gmail.com, csmliu@outlook.com, 18601200232@163.com, wangxt@126.com,

figoleilei@163.com, wzmzuo@hit.edu.cn

The content of this supplementary material involves:

- Noise Components Analysis in Sec. **A**.
- Comparison between Different Noise Modeling Methods in Sec. **B**.
- Training Detail in Denoising Stage in Sec. **C**.
- More Ablation Studies in Sec. **D**.
- Visualizing ISO dependence in Sec. **E**.
- Low-light Image Denoising Dataset in Sec. **F**.
- Image Denoising Datasets in rawRGB Space in Sec. **G**.
- More Quantization and Qualitative Results in Sec. **H**.

## A. Noise Components Analysis

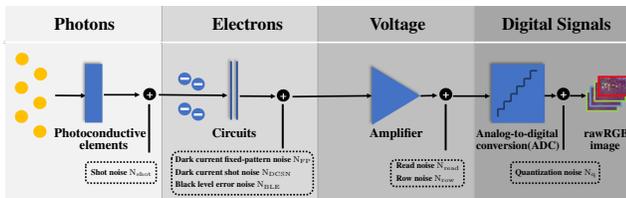


Figure A. Overview of imaging pipeline of CMOS photosensors.

As shown in Fig. **A**, the rawRGB image acquisition process can be summarised in four processes, which are photons incidence, photons converted to electrons, electrons to voltage, and voltage to digital signals [4], and various noises can be induced in these operations.

**Photons Incidence.** Shot noise arises due to the uncertainty generated when photons are incident, which is an unavoidable physical phenomenon [10, 15]. Shot noise is the dominant source of signal-dependent noises, which is generally modeled as a Poisson distribution and is determined by the signal and the camera gain. Some methods [9, 12] approximate the Poisson distribution by the Gaussian distribution.

**Photons to Electrons.** Dark currents are electrons generated within the sensor due to the thermal effect, hence dark current noise is also known as thermal noise [7, 15]. We divide the dark current noise  $N_{DC}$  into three components, *i.e.*, dark current fixed-pattern noise  $N_{FP}$ , dark current shot noise  $N_{DCSN}$  and black level error noise  $N_{BLE}$ .

In a real sensor, slight differences between individual pixels resulting in another source of noise called dark current fixed pattern noise [15] or dark-current nonuniformity [10]. Early sensor experts considered dark current fixed pattern noise suffers from temperature dependency [11, 13]. With the increasingly advanced sensor technology [12, 22] of recent years, mainstream modern sensors are no longer subject to temperature dependence. Therefore, PMN [8] and Starlight [17] consider this temperature-independent FPN.

The presence of dark currents not only leads to dark current FPN  $N_{FP}$  but also to cause the output level of the sensor to be unstable at 0 in a light-free environment, which can cause some images to appear color bias phenomenon [19, 20]. Further, sensor experts define the output value of a sensor in the absence of light as the black level. Therefore, the black level error is the difference between the level when no light arrives at camera and the recorded black level by the camera. While ELLE [19] uses a uniform distribution to model black level error noise  $N_{BLE}$ , ELD models this type of noise by sampling the real black level errors. PMN [8] recalibrates the black level by averaging 400 bias frames taken at different ISO settings.

In addition,  $N_{DCSN}$  is the dark current shot noise representing the randomness of the thermal effect.

**Electrons to Voltage.** Read noise  $N_{read}$  is generated by the uncertainty of the electronic readout and is a device-dependent noise. To model the read noise and dark current shot noise jointly, ELD [20] employs the Tukey lambda

distribution [14], which has a longer tail and can help mitigate the chrominance artifacts in low-light conditions. ELLE [19] and Starlight [17] employ the heteroscedastic Gaussian noise to model the shot noise  $N_{shot}$ , the read noise  $N_{read}$  and dark current shot noise  $N_{DCSN}$  jointly.

Besides, the row-by-row readout manner of CMOS sensors will also lead to row noise, since there exists randomness of the analog-to-digital converter (ADC) between different rows.

**Voltage to Digital Signals.** The quantization error is caused by rounding, as the ADC has a finite accuracy. The probability distribution of the quantization noise is usually considered to be uniform.

In summary, heteroscedastic Gaussian noise and row noise can be implemented by establishing a relationship between the ISO and noise parameters. For the noise caused by dark currents, the dark current FPN is a spatially-dependent and time-invariant fixed-pattern noise, while the black level error is ISO-specific single value. The proposed fine-grained noise model and the analysis of each noise component can guide us in the design of network structure and training strategy.

## B. Comparison of Noise Models

As shown in Tab. A, we compare our noise model (Ours) with 4 physics-based methods (*i.e.*, ELLE [19], ELD [20], SFRN [21], and PMN [8]) and 2 learning-based methods (*i.e.*, NoiseFlow [2] and Starlight [17]). Note that the order is different from Tab. 1 in the main manuscript for a clearer comparison. And the quantization noise is omitted since all methods model it with the same formulation.

ELLE [19] and ELD [20] do not link the black level error noise with the ISO configuration, and fixed-pattern noise is overlooked in their noise models. SFRN [21] and PMN [8] employ the emerging real noise-based synthesis strategy, but they require real bias frames and paired noisy-clean images for noise synthesis, respectively, which also limits the generalization capability of the model to some extent. In contrast to the physical methods described above, we can have multiple types of noise and ISO relationships while avoiding complex manual calibrations. For example, in Sec. 4.3, we compare our ISO-related black level error noise with the image-related ones used by ELD and ELLE. The better performance illustrates that ISO correlation of noise models is crucial. PMN does not model ISO-related read noise, row noise, and dark current shot noise, although it takes into account ISO-related shot noise, black level error noise, and dark current FPN. While SFRN uses real bias frames instead of signal-independent noise, compared to our construction of ISO-dependent noise components, the capacity of bias frames can also limit its generalization capability. Further, our approach can be adopted for two dif-

ferent synthesis strategies, *i.e.*, clean image-based and real noise-based synthesis strategies, which also demonstrates the flexibility of our approach.

As for learning-based methods, the NoiseFlow [2] ignores the complex low-light noise components such as the row noise and fixed-pattern noise. Starlight [17] only considers a single ISO setting and ignores ISO-related noise parameters, which greatly reduces the accuracy and generalization ability of the noise model. In addition, they still require manually calibrating the fixed-pattern noise, which limits the flexibility of starlight [17]. In contrast, our approach takes into account strip artifacts and color bias without manual parameter calibration and is flexible enough to incorporate the emerging real noise-based denoising process.

## C. Training Detail in Denoising Stage

We follow the same training settings with PMN. The denoising network [6] is trained with 8 non-overlapped  $512 \times 512 \times 4$  patches with  $L_1$  loss for 1,800 epochs. We initialize the learning rate to  $2 \times 10^{-4}$ , which is steadily decreased to  $1 \times 10^{-5}$  with the cosine learning rate decay [16]. In specific, we use rotation and flipping augmentations, 1,865 and 720 samples from SID and LLD are used for Sony A7S2 and Nikon D850. We chose ELD as val set and SID as test set, which is consistent with PMN.

## D. More Ablation Studies

**Paired Data Volume.** Our proposed low-light dataset (LLD) is primarily intended for training the learning-based noise models. This part of the experiment is to verify whether the newly captured LLD is in close agreement with the noise distribution of the original public SID dataset. In Tab. B, due to increased paired data volume, performance improvement in denoising network performance can be achieved.

**Noisy-Clean Pair Sources.** In training the denoising network stage, half data in each mini-batch comes from the darkshading correction (*i.e.*,  $(D - k \cdot N_{FP} - N_{BLE}(ISO), I)$  pairs) and the other comes from the zero-mean part of our noise model (*i.e.*,  $(I + N_{HG} + N_{row} + N_q, I)$  pairs). In other words, the ratio of darkshading correction pairs  $r = 50\%$ . It is worth noting that darkshading correction pairs require paired noisy-clean images for synthesis, while zero-mean part pairs of our noise model require only clean images for synthesis. To assess the contribution of each set, experiments are conducted by employing different darkshading correction pairs ratios of total paired images per mini-batch for training the denoising network. The noisy images with the darkshading correction, *i.e.*,  $D^{test} - k \cdot N_{FP} - N_{BLE}(ISO)$ , are fed into the de-

Table A. Comparison between different noise modeling methods.  $I$  denotes a clean image for noisy image synthesis,  $S$  means that the noise is sampled from real bias frames. In this table, all hand-calibrated variables are highlighted in blue, and the learnable parameters are highlighted in red. Note that the variables marked with \* are kept fixed for different ISO configurations.

Method	Category	$N_{\text{shot}}$	$N_{\text{read}}$	$N_{\text{DCSN}}$	$N_{\text{FP}}$	$N_{\text{BLE}}$	$N_{\text{row}}$
ELLE [19]	Physics	$\mathcal{N}(0, \beta_1 \cdot I + \beta_2)$			-	$U(a^*, b^*)$	$\mathcal{N}(0, \sigma_r^2)$
ELD [20]	Physics	$\beta_1 \mathcal{P}(I/\beta_1) - I$	$TL(\lambda^*; 0, \sigma_{TL})$		-	$S$	$\mathcal{N}(0, \sigma_r^2)$
SFRN [21]	Physics	$\beta_1 \mathcal{P}(I/\beta_1) - I$	$S$				
PMN [8]	Physics	$\beta_1 \mathcal{P}(I/\beta_1) - I$	-	-	$N_{\text{FP}} \cdot \text{ISO}$	$N_{\text{BLE}}(\text{ISO})$	-
NoiseFlow [2]	Learn	$\mathcal{N}(0, \beta_1 \cdot I + \beta_2)$			-	-	-
Starlight [17]	Learn	$\text{DNN}^*(\mathcal{N}(0, \beta_1^* \cdot I + \beta_2^*))$			$\text{DNN}^*(N_{\text{FP}}^*)$	-	$\text{DNN}^*(\mathcal{N}(0, \sigma_r^{2*}))$
Ours	Learn	$\mathcal{N}(0, \beta_1 \cdot I + \beta_2)$			$N_{\text{FP}} \cdot k$	$N_{\text{BLE}}(\text{ISO})$	$\mathcal{N}(0, \sigma_r^2)$

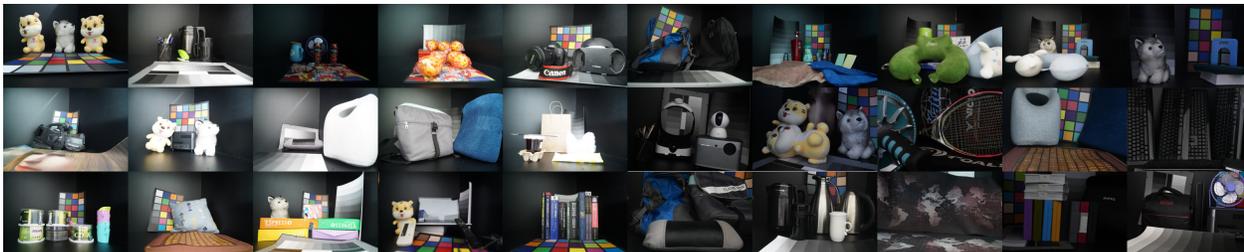


Figure B. Thirty scenarios in our low-light image denoising dataset (LLD).

Table B. PSNR/SSIM performance with ablation studies about different paired data volume settings.

Paired Data	Index	SID			ELD	
		100	250	300	100	200
SID [6]	PSNR	42.06	39.60	36.85	44.47	41.97
	SSIM	0.955	0.938	0.923	0.968	0.928
SID [6]+LLD	PSNR	42.42	39.62	36.88	44.56	42.33
	SSIM	0.955	0.939	0.924	0.968	0.937

Table C. PSNR/SSIM performance with ablation studies about different darkshading correction pairs ratios  $r$  of total paired images per mini-batch for training the denoising network.

$r$	Index	SID			ELD	
		100	250	300	100	200
100%	PSNR	42.84	40.53	37.57	46.12	44.34
	SSIM	0.960	0.946	0.933	0.982	0.974
75%	PSNR	42.81	40.60	37.44	46.22	44.55
	SSIM	0.959	0.944	0.931	0.982	0.973
50%	PSNR	<b>43.36</b>	<b>41.02</b>	<b>37.80</b>	<b>46.74</b>	<b>44.95</b>
	SSIM	<b>0.961</b>	<b>0.948</b>	<b>0.935</b>	<b>0.986</b>	<b>0.977</b>
25%	PSNR	42.97	40.59	37.54	46.44	44.72
	SSIM	0.960	0.946	0.934	0.985	0.976
0%	PSNR	42.75	40.41	37.15	46.24	44.51
	SSIM	0.956	0.932	0.916	0.984	0.973

noising network during the test phase to obtain the denoising results. As shown in Tab. C, the darkshading correction pairs ratio setting  $r$  of total paired images 50% is adopted for training the deep denoisers in our experiments.

## E. Visualizing ISO dependence

As shown in the Fig. Fig. D, when training our *noise model* and *denoiser* with the  $N_{\text{BLE}}(\text{ISO})$  calibrated by PMN, the PSNR will drop by  $\sim 0.1$  dB, showing the effectiveness of our learned ISO dependency.

## F. Low-light Image Denoising Dataset

For the noisy images, we adopt 24 different ISO settings, *i.e.*, 50, 64, 80, 160, 200, 250, 320, 400, 500, 640, 800, 1000, 1250, 1600, 2000, 3200, 4000, 5000, 6400, 8000, 10000, 12800, 16000 and 25600. Our method models the noise in the rawRGB space, and we exclusively employed unprocessed 14-bit raw data. As shown in Fig. B, 30 different sets of indoor scenes are adopted in our LLD. Sony Imaging Edge Desktop Remote and Nikon Camera Control Pro 2 software are employed for remote shooting with Sony A7S2 and Nikon D850. Empirically, the misalignment of long and short exposure images captured with the remote shutter is less severe than with the manual shutter. The exposure time for the noisy images is set to four groups (*i.e.*, 1/10s, 1/30s, 1/1600s, and 1/3200s) and the corresponding exposure time for the clean images is set to three groups (2s, 4s, and 8s), which ensures that the exposure ratio is in the range of [100, 300]. The exposure ratio with the same definition as ELD/SFRN, *i.e.*,  $\rho = \frac{T_{\text{long}} \times \text{ISO}_{\text{long}}}{T_{\text{short}} \times \text{ISO}_{\text{short}}}$ , and  $T_{\text{short}}$  is chosen from the four groups such that  $\rho \in [100, 300]$ . We capture only one short exposure noisy image using a specific ISO setting within one scene. ELD and LLD employ identical camera devices (*i.e.*, Sony A7S2 and Nikon D850), as well as low/high ISO and long/short exposure settings for capturing the noisy-clean pairs, so the ELD dataset can serve as a suitable validation or testing set for the LLD dataset. By contrast, ELD [20] chooses three ISO levels (*i.e.*, 800, 1600, and 3200) and two exposure ratios (*i.e.*, 100 and 200) for noisy images in 10 scenes, resulting in 60 rawRGB image pairs in total, which limits ELD to the benchmark rather than the training set for learning-based

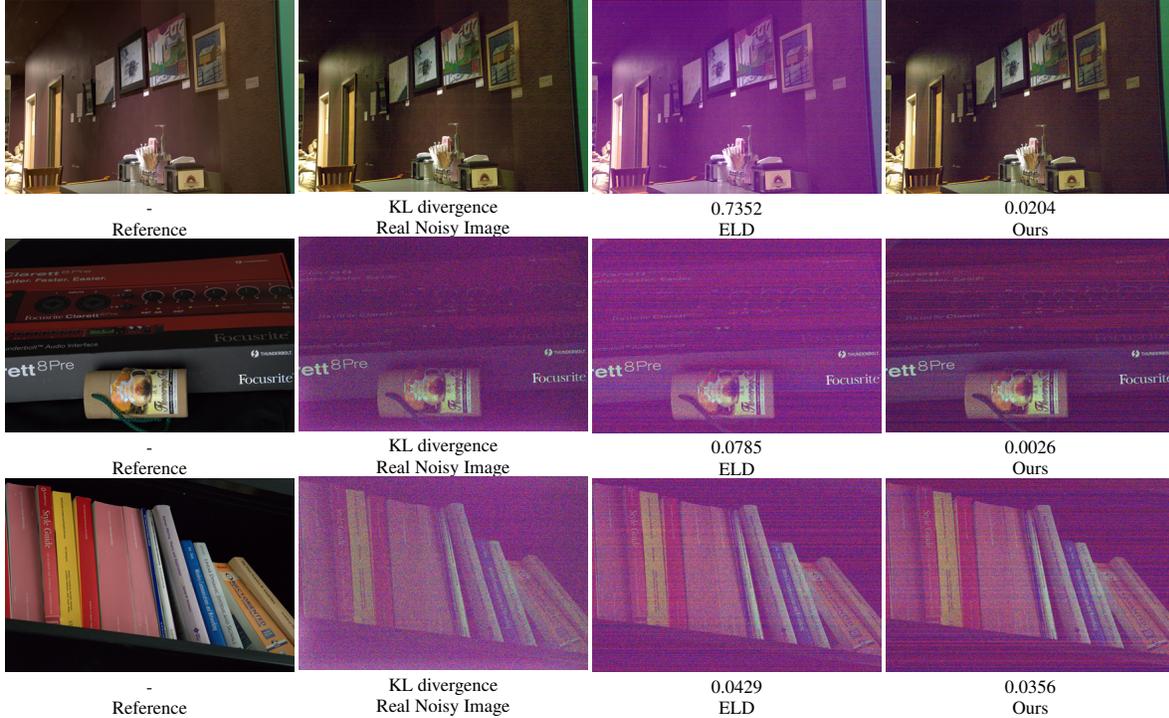


Figure C. Comparison of different noise synthesis methods between ELD [20] and Ours.

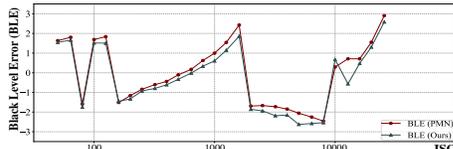


Figure D.  $N_{BLE}(ISO)$  between PMN and Ours on Sony A7S2.

methods.

## G. Image Denoising Datasets in rawRGB Space

As shown in the Tab. D, we present the descriptions of relevant image denoising datasets in rawRGB space, *e.g.*, number of scenes and noisy-clean pairs, reference synthesis method and assumed noise model.

## H. More Quantization and Qualitative Results

**Comparison of Different Noise Synthesis Methods.** As shown in the Fig. C, we compare the synthesized noise obtained by our method and ELD with the real noise. ELD captures the row noise pattern well, but the color bias is less consistent with the real noisy images. In contrast, our method takes into account the ISO-related black level error noise and therefore has a color bias closer to the real noisy images. In terms of KL divergence, our noise model produces synthetic noise that matches the real noise closer than ELD, *i.e.*, lower KL divergence.

**Comparison of Denoising Results.** As shown in Figs. E

and F, we compare our physics-guided ISO-dependent noise model in two noisy image synthesis schemes, *i.e.*, *clean image-based* synthesis strategy (denoted by Ours) and *real noise-based* synthesis strategy (denoted by Ours\*), with 3 learning-based methods, 3 physics-based methods, and 3 real noise-based methods. Neglecting black level error noise, Poisson-Gaussian noise [9], NoiseFlow [2] and CA-GAN [5] leads to significant color bias in the denoising results. Starlight [17] only considers a single ISO setting and therefore has very limited generalization capabilities. In contrast, our method provides the clearest texture and most accurate color denoising result.

## References

- [1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, and Michael S Brown. Ntire 2020 challenge on real image denoising: Dataset, methods and results. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 496–497, 2020. 5
- [2] Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown. Noise flow: Noise modeling with conditional normalizing flows. In *IEEE International Conference on Computer Vision*, pages 3165–3173, 2019. 2, 3, 4
- [3] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1692–1700, 2018. 5
- [4] European Machine Vision Association et al. Standard for characterization of image sensors and cameras. *EMVA Stan-*

Table D. Descriptions of the six image denoising databases in rawRGB space.

Database	Shooting devices	# Scenes	# Pairs	Reference synthesis method	Assumed noise model
DND [18]	Sony A7R, Olympus OMD E-M10, Sony RX100 IV and Nexus 6P	50	50	low/high ISO and long/short exposure	$N_{shot}, N_{FP}, N_{BLE}, N_{DCSN}, N_{read}, N_{row}, N_q$
SIDD [3]	Google Pixel, iPhone 7, Samsung Galaxy S6 Edge, Motorola Nexus 6 and LG G4	10	320	Multiple overlays	$N_{shot}, N_{DCSN}, N_{read}, N_{row}, N_q$
SIDDPlus [1] (Validation)	Google Pixel 2, Google Pixel 3 XL, LM-G710-LGE-lge, iPhone 7, iPhone 11 and HTC U12	32	1024	Multiple overlays	$N_{shot}, N_{DCSN}, N_{read}, N_{row}, N_q$
SID [18]	Sony A7S2	280	1865	long/short exposure	$N_{shot}, N_{FP}, N_{BLE}, N_{DCSN}, N_{read}, N_{row}, N_q$
ELD [20]	Sony A7S2, Nikon D850, Canon EOS70D, Canon EOS700D	10	60	low/high ISO and long/short exposure	$N_{shot}, N_{FP}, N_{BLE}, N_{DCSN}, N_{read}, N_{row}, N_q$
LLD	Sony A7S2, Nikon D850	30	720	low/high ISO and long/short exposure	$N_{shot}, N_{FP}, N_{BLE}, N_{DCSN}, N_{read}, N_{row}, N_q$

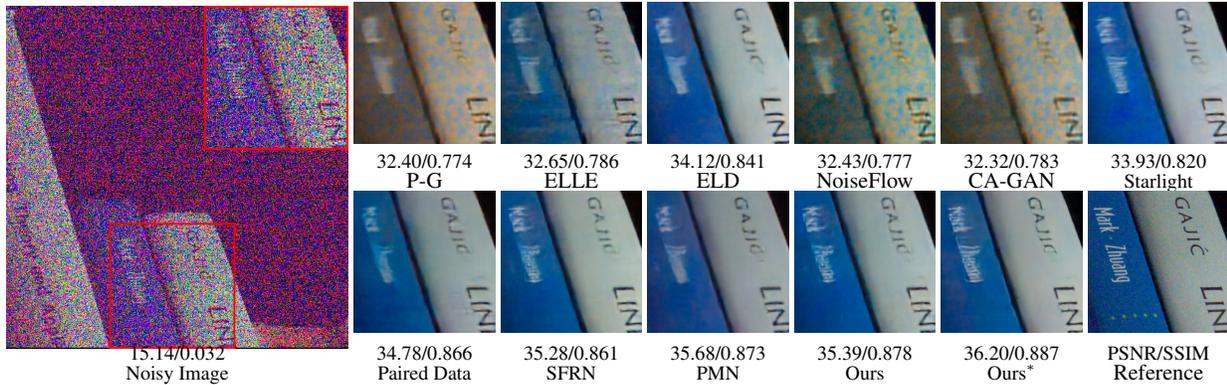


Figure E. Denoising results of different methods on real noisy images from SID [6].

- ard, 1288, 2010. 1
- [5] Ke-Chi Chang, Ren Wang, Hung-Jin Lin, Yu-Lun Liu, Chia-Ping Chen, Yu-Lin Chang, and Hwann-Tzong Chen. Learning camera-aware noise models. In *European Conference on Computer Vision*, pages 343–358. Springer, 2020. 4
- [6] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018. 2, 3, 5
- [7] Yuan Chen, Steven M Guertin, Mihail Petkov, Duc N Nguyen, and Frank Novak. A chip and pixel qualification methodology on imaging sensors. In *IEEE International Reliability Physics Symposium. Proceedings*, pages 435–439. IEEE, 2004. 1
- [8] Hansen Feng, Lizhi Wang, Yuzhi Wang, and Hua Huang. Learnability enhancement for low-light raw denoising: Where paired real data meets noise modeling. In *ACM International Conference on Multimedia*, pages 1436–1444, 2022. 1, 2, 3
- [9] Alessandro Foi, Mejdj Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transac-*

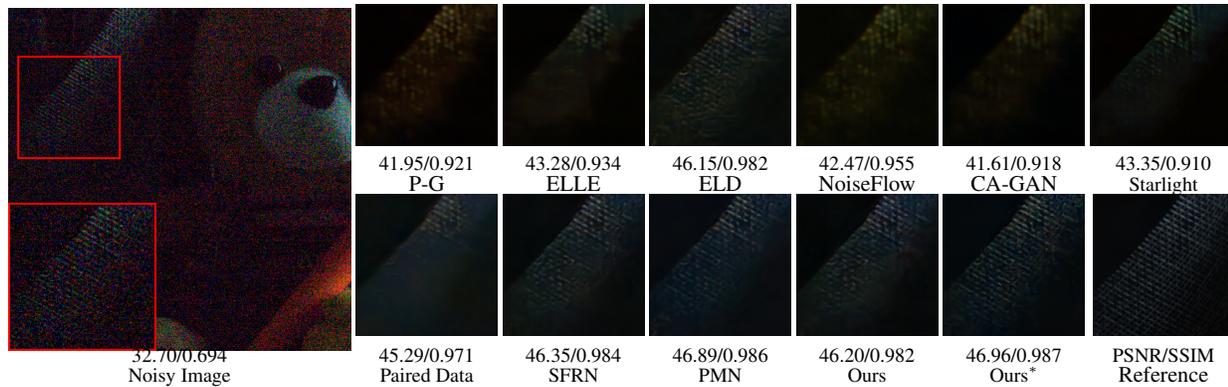


Figure F. Denoising results of different methods on real noisy images from ELD [20].

- tions on Image Processing, 17(10):1737–1754, 2008. 1, 4
- [10] Ryan D Gow, David Renshaw, Keith Findlater, Lindsay Grant, Stuart J McLeod, John Hart, and Robert L Nicol. A comprehensive tool for modeling CMOS image-sensor-noise performance. *IEEE Transactions on Electron Devices*, 54(6):1321–1329, 2007. 1
- [11] Gerald C Holst. Cmos/ccd sensors and camera systems. *CMOS/CCD sensors and camera systems/Gerald C. Holst*, 2007. 1
- [12] Yeonseong Hwang, Seongjoo Lee, and Minkyu Song. Design of a cmos image sensor with a 10-bit two-step single-slope a/d converter and a hybrid correlated double sampling. In *2014 10th Conference on Ph. D. Research in Microelectronics and Electronics (PRIME)*, pages 1–4. IEEE, 2014. 1
- [13] Bernd Jähne. Emva 1288 standard for machine vision: Objective specification of vital camera data. *Optik & Photonik*, 5(1):53–54, 2010. 1
- [14] Brian L Joiner and Joan R Rosenblatt. Some properties of the range in samples from tukey’s symmetric lambda distributions. *Journal of the American Statistical Association*, 66(334):394–399, 1971. 2
- [15] Mikhail Konnik and James Welsh. High-level numerical simulations of noise in ccd and cmos photosensors: review and tutorial. *arXiv preprint arXiv:1412.4031*, 2014. 1
- [16] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 2
- [17] Kristina Monakhova, Stephan R Richter, Laura Waller, and Vladlen Koltun. Dancing under the stars: video denoising in starlight. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 16241–16251, 2022. 1, 2, 3, 4
- [18] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1586–1595, 2017. 5
- [19] Jing Wang, Yitong Yu, Songtao Wu, Chang Lei, and Kuanhong Xu. Rethinking noise modeling in extreme low-light environments. In *IEEE International Conference on Multimedia and Expo*, pages 1–6. IEEE, 2021. 1, 2, 3
- [20] Kaixuan Wei, Ying Fu, Yinqiang Zheng, and Jiaolong Yang. Physics-based noise modeling for extreme low-light photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1, 2, 3, 4, 5, 6
- [21] Yi Zhang, Hongwei Qin, Xiaogang Wang, and Hongsheng Li. Rethinking noise synthesis and modeling in raw denoising. In *IEEE International Conference on Computer Vision*, pages 4593–4601, 2021. 2, 3
- [22] Mei Zou, Ji-qing Zhang, Sheng-you Zhong, Zheng-fen Li, and Li-bin Yao. Low-light-level cmos imaging sensor with ctia and digital correlated double sampling. *Analog Integrated Circuits and Signal Processing*, 101(3):449–461, 2019. 1