SLIM: Small and Learnable Image Signal Processing Module for CMOS and Quanta Image Sensors

Stanley H. Chan[1], Yiheng Chi[1], and Preston Rahim[1]

[1] DeepLux Technology Inc., West Lafayette, IN 47906, USA
{stanley.chan, yiheng.chi, preston.rahim}@deeplux.tech

Abstract. While multibit Quanta Image Sensors (QIS) today have demonstrated a superior sub-electron read noise characteristic, at extreme photon-limited conditions they still face the fundamental photon shot noise problem. The image signal processing (ISP) unit of today's multibit QIS is largely identical to those used for CMOS image sensors. In extreme photon-limited conditions, these physics-based ISP struggle to generate high-quality images. Deep learning methods are seen as the potential solution to overcome the low-light bottleneck, but existing neural networks are too large to fit into any camera products. In this paper, we present a learning-based ISP where key components are replaced by a lightweight neural network followed by traditional physics-based filtering steps. The proposed ISP, known as the Small and Learnable ISP Module (SLIM), allows us to jointly demosaick and denoise images at a photon level as low as 1 photon per pixel where traditional ISP fails.

1. Introduction

Single-bit and multi-bit Quanta Image Sensors (QIS) have delivered promising low-light image capturing results with sub-electron read noise. However, as the total amount of photon flux drops, QIS will eventually encounter the fundamental photon shot noise limit. The mainstream image signal processing (ISP) units today have limited capability of handling an excessive amount of shot noise. Many of them are still using the classical signal processing techniques based on engineered heuristics. Recent advancements in deep learning has generated a significant interest in the ISP community where people start to consider upgrading these traditional ISP to a learning-based ISP. Yet, the complexity of learning-based ISP (especially those using deep neural networks) is so high that even a high-end mobile phone processor only performs such operations occasionally when a user needs to restore a single photograph. For mid-grade and low-end products such as laptops, medical devices, cars, and household appliances, pushing artificial intelligence to ISP becomes a pressing demand that will continue for the coming decade.

In this paper, we present an algorithmic solution for two critical steps in the ISP pipeline: low-light denoising and low-light demosaicking. Compared to the traditional ISP and deep learning based ISP, our solution can be seen as a middle-ground solution that balances performance and complexity:

- Compared to traditional ISP that performs rule-based demosaicking and linear filtering (usually edge-aware weighted averaging and median filtering), our proposed solution uses a few shallow layers of neurons to extract high and low level features across different scales. The denoising is performed by a chain of new procedures to construct and select denoising filters. These new procedures alleviate the limitations of traditional ISPs which often fail to identify edges and texture when the input is corrupted by heavy noise.
- Compared to deep-learning based ISP such as [1] that requires training large models end-to-end, our proposed solution is significantly more light-weight. We use simple convolutions and shallow layers of neurons to perform most of the tasks, in contrast to complex models such as vision transformers and self-attention. Moreover, since our design is based on the traditional pipeline where parts are modularized, it makes debugging and interpretation easier.

2. Small and Learnable ISP Module (SLIM)

Figure 1 illustrates the schematic diagrams of a typical ISP and our proposed ISP. The input to the ISP is the Bayer color filter array pattern assuming that the standard pre-processing steps are completed (e.g., gray-level offset, pixel response non-uniformity calibration, dead pixel removal, etc.) The focus of our work is the demosaicking and denoising steps in the raw domain. The output is sent to a downstream ISP module for additional processing of color and edge. In the proposed pipeline, we replace several key steps of the denoising process by learning based methods:

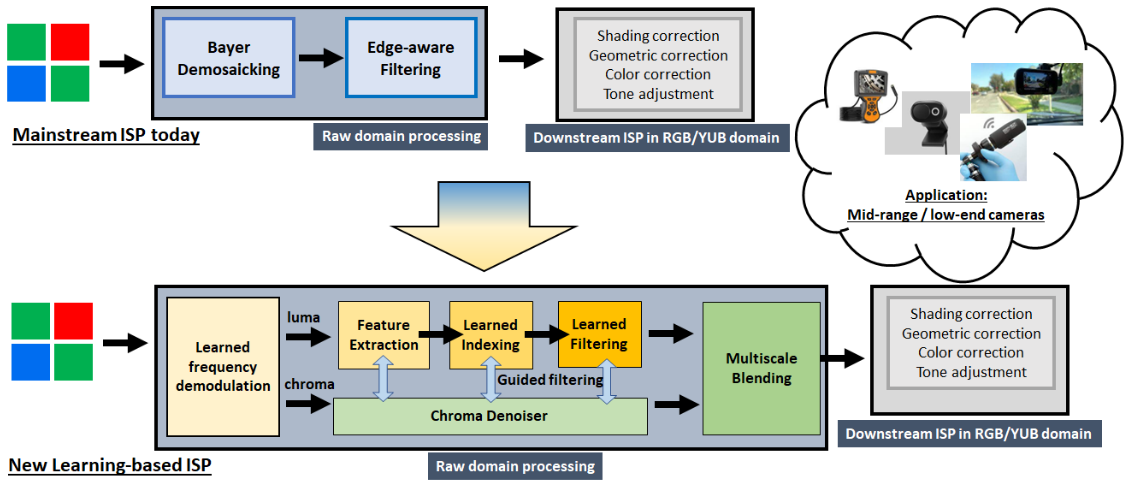

Figure 1: Schematic diagram of the proposed ISP compared to a typical ISP. In our proposed ISP, we replace several key steps of the traditional ISP with learning-based modules. These modules are light-weight shallow neural networks that are trained independently of each other. For end applications, we target mid-grade to low-end cameras such as laptops, dash-cams, endoscopes, and inspection cameras.

- **Learnable frequency selection**. We use a frequency selection module [2] to solve the Bayer color filter array (CFA) demosaicing problem in SLIM. When light goes through the CFA, the color channels are modulated by carrier signals with known frequencies. Given this mosaiced signal, the full-resolution color information can be recovered by signal demodulation operations. Traditional methods use linear demodulation schemes which are not robust to severe noise under low lighting conditions [4]. Recent deep-learning based demosaicing methods usually require excessive computing resources, memory, and execution time, while they often ignore the physics of CFAs. Instead, we propose a learnable light-weight frequency selection module. This module is developed upon the physics of the Bayer pattern CFA and is adaptive to various signal-to-noise ratios under different lighting conditions.
- **Feature extraction**. SLIM extract features from the luma channel of the demodulated image signal before denoising it. Such nonlinear feature extraction is crucial to the later stage of denoising because it accumulates spatial information, which helps reconstruct the image when the input is contaminated by heavy shot noise at low light.
- **Learned indexing**. Instead of using a single filter or a selected filter from a collection, our proposed SLIM uses a learned indexing scheme to calculate a combination of multiple filters and their corresponding strengths (weights) from the extracted features. Such indexing is continuous and back-propagatable. With this learned indexing, SLIM composes sophisticated and nonlinear filters similar to deep networks do, but the computational complexity of SLIM is much lower so it can be empowered by edge devices.
- **Learned filtering**. Like many data-driven methods today, SLIM learns the image filters from vast image data. The training data are synthesized using realistic image formation models at various light levels. When deployed, processing image signals with these predetermined learned filters can save much memory compared to generating full-resolution images using deep restoration networks in one pass.

- **Chroma-Luma decoupling**. The SNR of luma signal is higher than chroma signals due to their derivations. Therefore, we distribute most computations towards luma channel processing. Afterwards, we use the luma features and indices to guide the denoising of the chroma channels.
- **Multi-scale blending**. A pyramid-shape multi-scale structure has been demonstrated effective in image processing literature. SLIM performs a two-level multi-scale blending to perform global-then-local image denoising.

Method	# Parameters	Description
Deep-Learning ISP		
MIRNet 2020 [9]	32,787,000	Denoising or super-resolution or image enhancement
Restormer 2022 [8]	26,127,000	Denoising or deraining or deblurring
DRUNet 2021 [10]	32,640,000	Denoising or demosaicing
PyNet 2020 [5]	47,554,000	ISP (demosaicking + denoising + white balance + color correction)
Traditional Filter-Bank ISP		
RAISR 2016 [7]	26,000	Superresolution
BLADE 2017^{+} [3]	28,000	Denoising
Proposed Learned Physics-based ISP		
SLIM (proposed)	126,000	ISP (demosaicking + denoising)

Table 1: Comparison between the number of parameters used in various ISPs. The number of parameters is measured in terms of the size of the filters and the number of filters. $^{+}$ Existing mainstream ISP are rule-based edge-aware filters. While they do not explicitly build the filter bank, the number of filters they use are in the same order of magnitude compared to BLADE.

Results and Conclusion

Table 1 shows the number of parameters used by existing deep learning models and SLIM. Figure 2 shows a comparison between a proprietary ISP on one of the mid-grade cameras and the proposed ISP. The input Bayer images are simulated at photon levels from 1 ppp to 100 ppp, assuming a 0.19e- read noise, 0.02e-/s dark current, 12-bit analog-digital converter, 80% quantum efficiency, and a uniform sensor response. We train the learning-based modules *independently* using simulated data across a wide range of photon levels. Such an independence implies that the whole ISP does *not* need to be trained end-to-end. Thus, updating one module does not interfere with another module, hence making the debugging at the same level as a traditional ISP (and much more convenient than a deep-learning ISP). To summarize, SLIM demonstrates the potential as a viable solution for the next generation learning-based ISP where the photon level is low.

References

[1] Y. Chi, A. Gnanasambandam, V. Koltun, and S. H. Chan. Dynamic low-light imaging with quanta image sensors. In *Proc. European Conf. Computer Vision*, pages 122–138, 2020.

[2] O. A. Elgendy, A. Gnanasambandam, S. H. Chan, and J. Ma. Low-light demosaicking and denoising for small pixels using learned frequency selection. *IEEE Trans. Computational Imaging*, 7:137–150, 2021.

[3] P. Getreuer, I. Garcia-Dorado, J. Isidoro, S. Choi, F. Ong, and P. Milanfar. BLADE: Filter learning for general purpose computational photography. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–11, 2018.

[4] A. Gnanasambandam, O. Elgendy, J. Ma, and S. H. Chan. Megapixel photon-counting color imaging using quanta image sensor. *OSA Optics Express*, 27(12):17298–17310, 2019.

[5] B.-H. Kim, J. Song, J. C. Ye, and J. Baek. Pynet-ca: enhanced pynet with channel attention for end-to-end mobile image signal processing. In *Proc. European Conf. Computer Vision*, pages 202–212, 2020.

[6] J. Ma, D. Zhang, O. A. Elgendy, and S. Masoodian. A 0.19e- rms read noise 16.7mpixel stacked quanta image sensor with 1.1 um-pitch backside illuminated pixels. *IEEE Electron Device Letters*, 42(6):891–894, 2021.

[7] Y. Romano, J. Isidoro, and P. Milanfar. RAISR: Rapid and accurate image super resolution. *IEEE Trans. Computational Imaging*, 3(1):110–125, 2016.

[8] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, pages 5728–5739, 2022.

[9] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao. Learning enriched features for real image restoration and enhancement. In *Proc. European Conf Computer Vision*, pages 492–511, 2020.

[10] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021.

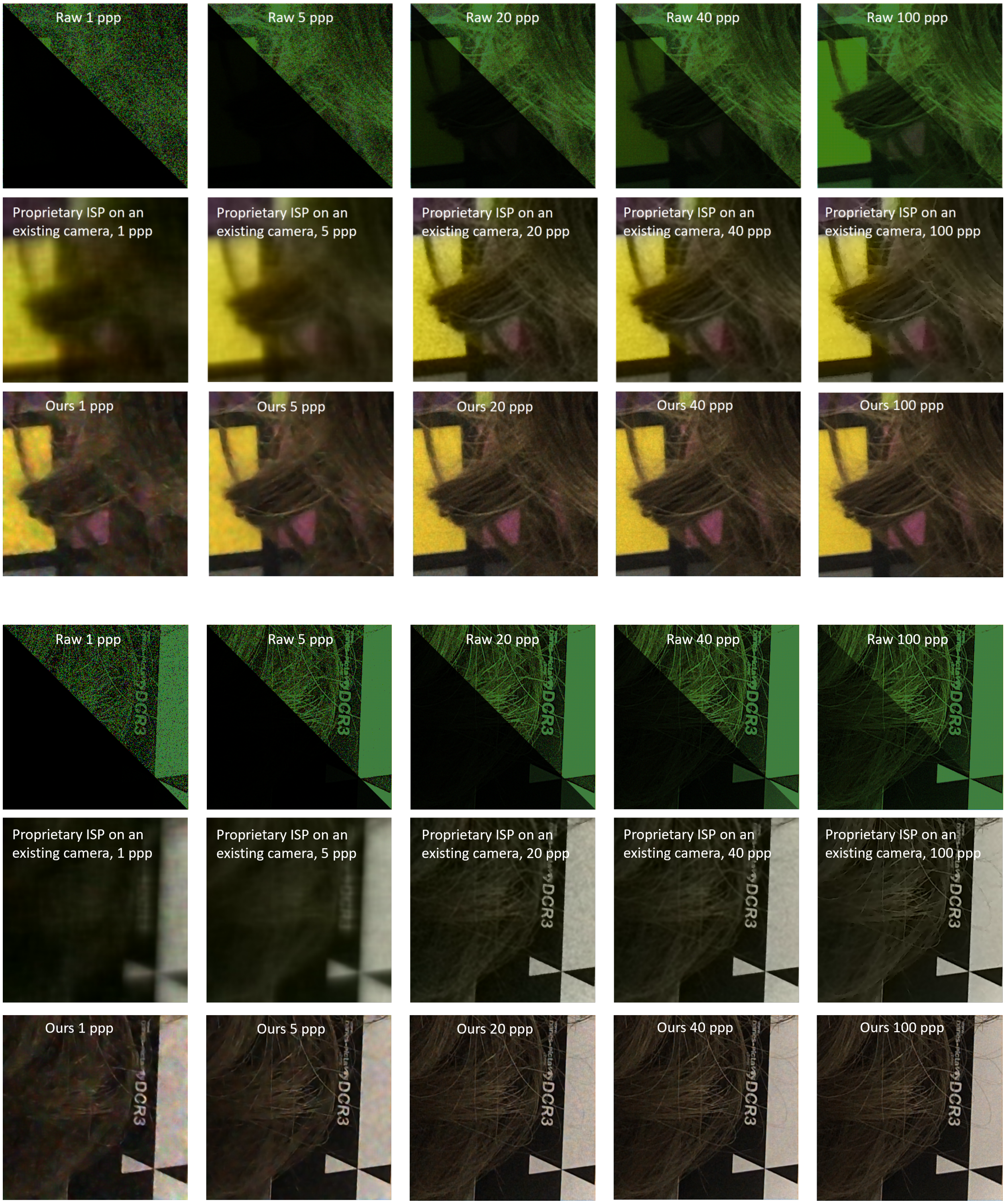

Figure 2: Denoising and demosaicking results of our proposed ISP, compared to a proprietary ISP used in existing image sensors. The noise level is indicated by the number of photons per pixel (ppp). The noise model follows a published QIS specification [6].