Low-Light Demosaicking and Denoising for Small Pixels Using Learned Frequency Selection

Omar A. Elgendy, *Member, IEEE*, Abhiram Gnanasambandam,, *Student Member, IEEE*, Stanley H. Chan, *Senior Member, IEEE*, and Jiaju Ma, *Member, IEEE*

***Abstract*—Low-light imaging is a challenging task because of the excessive photon shot noise. Color imaging in low-light is even more difficult because one needs to demosaick and denoise simultaneously. Existing demosaicking algorithms are mostly designed for well-illuminated scenarios, which fail to work with low-light. Recognizing the recent development of small pixels and low read noise image sensors, we propose a learning-based joint demosaicking and denoising algorithm for low-light color imaging. Our method combines the classical theory of color filter arrays and modern deep learning. We use an explicit carrier to demodulate the color from the input Bayer pattern image. We integrate trainable filters into the demodulation scheme to improve flexibility. We introduce a guided filtering module to transfer knowledge from the luma channel to the chroma channels, thus offering substantially more reliable denoising. Extensive experiments are performed to evaluate the performance of the proposed method, using both synthetic datasets and real data. Results indicate that the proposed method offers consistently better performance over the current state-of-the-art, across several standard evaluation metrics.**

***Index Terms*—Quanta Image Sensors (QIS), low light, color filter arrays, single-photon imaging, color demosaicking**

I. Introduction

A. Motivations

Since the introduction of CMOS active pixel sensors in the early 90's, pixel pitch of digital image sensors have been continuously shrinking. Today, the mainstream CMOS image sensors (CIS) used on DSLR cameras are in the range of 3.5μm-8.3μm, with smaller ones (down to sub-1μm) used on hand-held cameras [1]. While smaller pixels enjoy higher pixel density per unit space and hence higher spatial resolution, they have reduced full-well capacity that limits the signal-to-noise ratio (SNR). As a result, using these small pixels for low-light conditions where photons are scarce, the acquired images will have very poor SNR.

To tackle this problem, alternative technologies to CMOS image sensors (CIS) have been developed. A few notable options include the single-photon avalanche diodes (SPAD) [2], [3] and the quanta image sensors (QIS) [4]–[7]. These single-photon image sensors are very sensitive to photons, and hence they are good candidates for low-light imaging. In particular, the latest prototype QIS have demonstrated a very promising read noise, dark current, and their size (1.1μm) is comparable to the smaller pixels of the CMOS technology [8], [9]. Many new applications have also been proposed for these sensors [10]–[14].

Regardless of the particular type of image sensors one uses (CIS or QIS), a universal problem to all sensors is the reconstruction of color images under *low-light*. That is, how do we reconstruct the full color image from the raw signal that is generated by an image sensor with a color filter array? If we put aside the low-light condition, color reconstruction is solely a demosaicking problem where solutions are plenty [15]–[19]. In low-light, the problem becomes much more challenging because one needs to jointly denoise and demosaick. While there are some existing methods for this task [20]–[28], these optimization based methods are iterative and they tend to oversmooth the images.

This paper presents an end-to-end learning-based color reconstruction method by grounding the deep neural networks on the known physical properties of the color acquisition process. Our method is a generic solution to both CIS and QIS with a small pixel pitch. The key observation here is that although pixels with a small pitch suffer from low-light and noise, the aliasing is weaker for smaller pixels. If we further use a better sensor such as a QIS, then it is possible to further reduce the noise. Therefore, with the appropriate combination of algorithms and sensors, we propose an algorithm to produce high-quality color imaging under extremely low-light conditions (1.8 photoelectrons per pixel per frame), which could be challenging with traditional methods.

B. Key ideas: Exploiting small pixels and low read noise

We summarize our key ideas in this subsection and discuss the details later. Small pixels and low read noise are two very important factors for color imaging in low-light. Traditional CIS and their corresponding demosaicking algorithms usually emphasize the mitigation of the aliasing (aka the *Moirè* artifacts), for example, using advanced edge-aware demosaicking methods [29] or using a post-processing module to remove the demosaicking artifacts [30], [31]. However, aliasing is less a problem when the pixels are small (Figure 1). This suggests that we can use a simple method for demosaicking, possibly some traditional methods, such as linear demodulation followed by denoising, to reconstruct the color.

To elaborate on the aliasing problem, we show three illustrations in Figure 1. In Figure 1(a), we plot the Nyquist sampling limit of the color filter array as a function of the pixel pitch and illustrate two color spectra associated with the sampling limits.

Omar Elgendy and Jiaju Ma are with Gigajot Technology Inc., Pasadena, CA 91107.

Abhiram Gnansambandam and Stanley H. Chan are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA. Contact author: stanchan@purdue.edu.

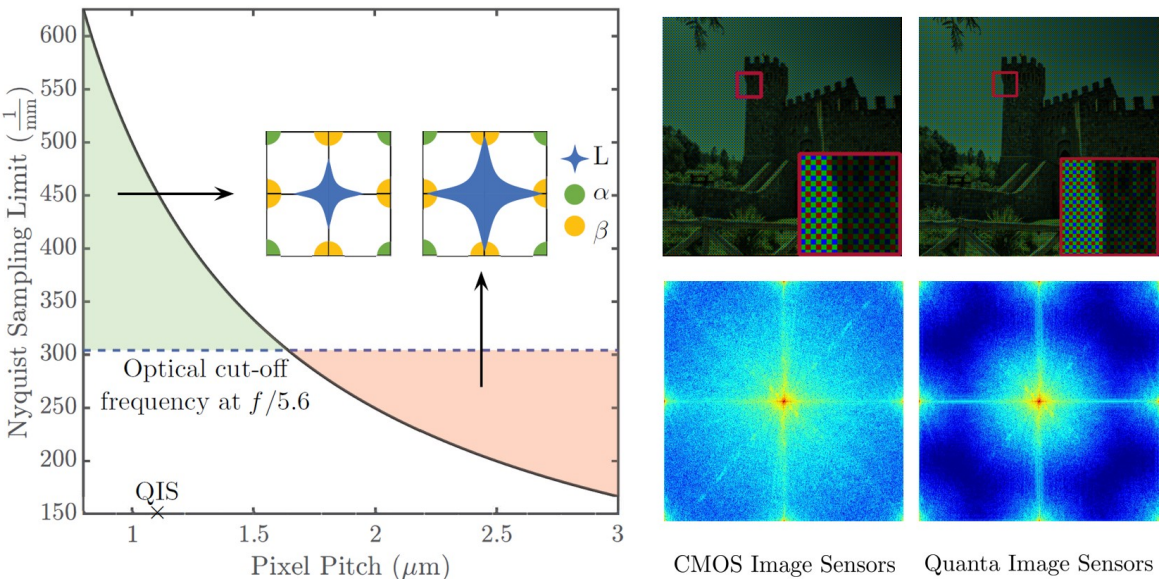


Fig. 1. [Left] Nyquist limit decreases with pixel pitch. At f-number value of $f/5.6$, QIS is diffraction-limited since Nyquist limit exceeds the optical cut-off frequency. [Top Right] An average of 10 QIS frames at photon level of $5.5e^{-}$. The CFA is a Bayer pattern. [Bottom Right] Fourier spectrum of the color. Notice that interference between base-band luminance and chrominance components at $(\pi, 0)$, $(0, \pi)$ and (π, π) is minimal.

Nyquist limit defines the lowest spatial sampling frequency required to prevent aliasing. As the pixels become smaller, we effectively oversample the scene and so the Nyquist limit increases. If we use a $f/5.6$ optical system as an example, then the maximum pixel pitch we can afford is 1.6μm, which is safely above the 1.1μm pitch of the current QIS. This argument is further justified by looking at the synthetic data shown in Figure 1(b) and (c), where we show the raw Bayer CFA data and the corresponding color spectrum. It can be seen that because of the small pixel pitch, QIS can potentially offer a much better spectrum.

A lower read noise is as important as smaller pitch as far as color reconstruction is concerned. If the sensor has a strong read noise, a linear demodulation scheme may not be able to keep the essential level of signal-to-noise for reliable denoising. As we will show in this paper, while we can still reconstruct color images for small pitch CIS, we can get better performance at low light using QIS type of low-noise sensors.

C. Scope and Contributions

There are two main contributions of this paper:

- Frequency-selection. The proposed demosaicking algorithm is rooted in the classical theory of color filter arrays. We develop a color processing module to *demodulate* the color channels by selecting the known carrier frequencies of the color filter array. Existing deep learning-based solutions using generic convolutional neural networks largely do not use the physics of the color filter arrays.
- Guided reconstruction. The proposed algorithm leverages the physics that the luma channel has a much better signal-to-noise ratio than the chroma channels. As such, signal details that are preserved by the luma channel can be used to guide the filtering of the chroma channels. Existing generic convolutional neural networks do not exploit these characteristics of the data.

An assumption we make in this paper is that the microlens of the sensor can converge light onto the sensing site. We also assume that crosstalk is minimal. Thus, the color crosstalk in the imaging scenarios is limited. This is merely a simplification consistent with the demosaicking literature [22]–[28], [32]–[34]. Treatment of the crosstalk can be done in two ways: (i) new color filter arrays as in [21], [35], and (ii) reconstruction algorithm that can perform deconvolution. To keep the focus of the present paper, we decide to leave crosstalk as future work.

To demonstrate the performance of the proposed algorithm, we report findings using both synthetic data and real data. We also present ablation studies that show the importance of each module in the proposed algorithm. The QIS camera used in this study is the *PathFinder* camera developed by Gigajot Tech Inc. [20], [36].

II. Background

The pipeline of our color image reconstruction method is shown in Figure 2. The goal is to take the input from the color filter array, and jointly demosaick and denoise the signal. Because this paper is positioned at the intersection of image sensors and demosaicking algorithms, in the following we highlight a few works relevant to both.

A. Related Work

Classical demosaicking algorithms. Classical demosaicking methods are developed for CIS with well-illuminated images. Under low-light conditions where the noise is heavy, the demosaicking algorithm must also have denoising capabilities. It is important to note that the order of denoising and demosaicking matters [15], [22], [37], [38]. If demosaicking is performed first, then the interpolation will destroy the spatial independence of the noise. This will substantially complicate the denoising process [39]. If denoising is performed first, then

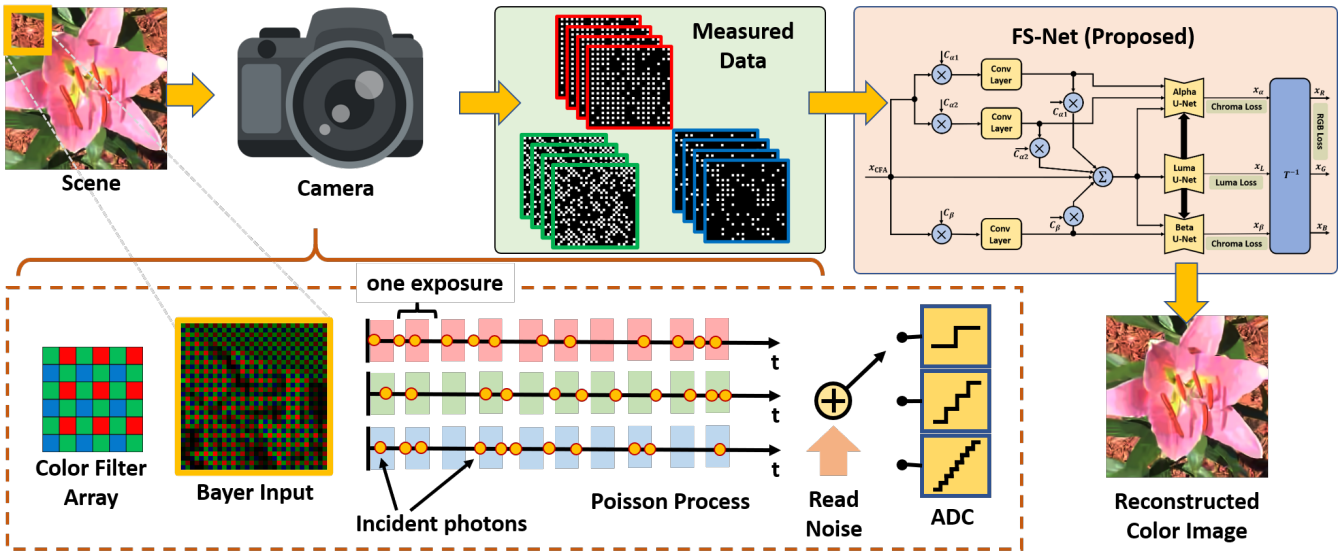


Fig. 2. Color imaging. As the light of the scene object arrives at the camera, the color filter array selects the wavelengths to pass through. This generates the Bayer pattern. Modeling the incident photons as a Poisson process, we obtain three channels of signals as illustrated by the three temporal responses of red, green, and blue pixels. After modeling the read noise of the sensor and performing the analog-to-digital conversion (ADC), we obtain a single frame or a stack of frames based on the camera. A CIS usually captures a single frame with high bit-depth (eg. 14 bits). QIS usually captures a stack of low bit-depth frames. The proposed method is an end-to-end solution known as the **frequency-selection network** (FS-Net). The output of the network is the reconstructed color image.

most of the image priors cannot be used because the mosaicked images do not have natural image statistics [37], [40]–[43]. Joint demosaicking and denoising methods are better options here. However, most of the joint demosaicking and denoising methods are iterative [22]–[28].

Deep neural network based demosaicking. State-of-the-art demosaicking algorithms are largely based on deep neural networks. The idea is to modify a generic deep neural network by adding a space-to-depth layer that converts the raw Bayer image into 4 Bayer channels with quarter resolution. The network processes these down-sampled channels then upscale them to a full-resolution color image using a depth-to-space layer. For example, the Demosaic-Net by Gharbi et al. [40] uses a residual network and a customized dataset within a curriculum training approach. Dong et al. [41] use generative adversarial training. Tan et al. [18], Cui et al. [30], Niu-Ouyang [31] use multi-phase approaches. Ehret et al. [43] study burst reconstruction without ground truth. Wu et al. [17] and Kiku et al. [16] use a guided filter for chrominance reconstruction, however, they do not consider the effect of noise.

Existing deep neural network-based solutions are generic in the sense that there is no consideration of the physics of the color filter arrays. Furthermore, down-sampling the RAW image to quarter resolution leads to a loss of resolution. As we will show in this paper, we can achieve better results by carefully including physical insights of color acquisition into the design and by processing the full-resolution RAW image directly without the need for down-sampling.

B. Quanta Image Sensors

What are QIS? The method proposed in our paper applies to both CIS and QIS. While CIS have been in use in cameras since the 1990s, QIS is a new type of image sensors. QIS was originally proposed in 2005 as a candidate solution for the performance challenges associated with the shrinking pixel sizes [4]. The principle of QIS is to partition a typical CMOS pixel into many tiny cells called "jots", where each jot is a single-photon detector. Depending on the desired level of bit-depth, the sensor can be operated in a single-bit mode or a multi-bit mode. If it operates in single-bit mode, then each jot will output a binary signal to reflect whether the number of photons exceeds a certain threshold. For multi-bit, the sensor quantizes the photon counts into multi-bit signals before the pixel reaches saturation. Readers interested in the early work of QIS image reconstruction and signal processing theory can consult, e.g., [6], [11], [44]–[49].

QIS vs SPAD. QIS is a generic concept according to its inventor [4]. It can be implemented using the existing CIS technology or a single-photon avalanche diode (SPAD). The QIS we use in this paper is the CIS-based sensor. It is different from a SPAD-QIS such as [50]. SPAD-QIS amplifies signal using avalanche multiplication. This requires a high electrical voltage (typically higher than 20V) to accelerate the photoelectron. Along with a larger transistor count and higher operating voltage in the SPAD pixels, SPAD-QIS has a high dark count ($> 10e^{-}$), a large pitch ($> 5\mu$m), a low fill-factor ($< 70\%$), and a low quantum efficiency ($< 50\%$). In contrast, CIS-QIS does not require avalanche multiplication. It has significantly better fill-factor, quantum efficiency, and dark current. SPAD-QIS are excellent candidates for resolving time-stamps, e.g., time-of-flight applications [12], [51]–[55], whereas CIS-QIS offers higher spatial resolution and detection efficiency. A comparison between CIS-QIS and SPAD-QIS can be found in the recent literature, e.g., [20].

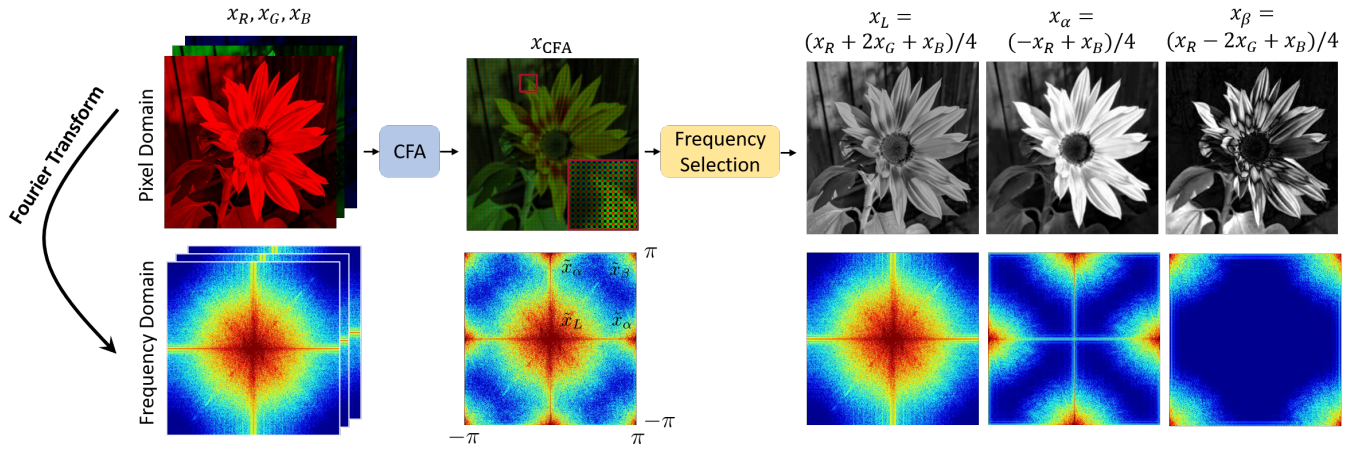


Fig. 3. Objective of classical frequency selection. Given a color filter array (CFA) image $\boldsymbol{x}_{\text{CFA}}$, the frequency selection method is a Fourier domain operation that extracts the corresponding frequency components of the luma $\boldsymbol{x}_L$ and chroma $\boldsymbol{x}_\alpha, \boldsymbol{x}_\beta$ channels from the image.

III. Frequency Selection Demosaicking Network

The proposed method leverages the classical frequency selection with new modifications. In this section, we first provide a background on frequency selection (III.A). Afterward, we present our modifications of learned demosaicking low-pass filters and guided filtering of chroma channels using the luma channel (III.B). We also present the loss functions (III.C) and the training process (III.D).

A. Frequency selection

Consider a color image $\boldsymbol{x}_{rgb} \in \mathbb{R}^{H\times W\times 3}$. We denote the normalized light intensities in the red, green and blue channels at the pixel (m,n) as

$$\boldsymbol{x}_{rgb}(m,n) = \left[x_r(m,n), \quad x_g(m,n), \quad x_b(m,n)\right], \tag{1}$$

where $m = 0,\ldots,H-1$, $n = 0,\ldots,W-1$. The color image $\boldsymbol{x}_{rgb} \in \mathbb{R}^{H\times W\times 3}$ is sub-sampled by the color filter array (CFA) to create a mosaicked image $x_{\text{CFA}} \in \mathbb{R}^{H\times W}$. Assuming that the CFA follows the standard Bayer pattern, it can be shown that the pixel (m,n) of the mosaicked image takes the form (See [56], [57]):

$$\begin{aligned} x_{\text{CFA}}(m,n) = x_L(m,n) + x_\alpha(m,n)\left(e^{j\pi m} + e^{j\pi n}\right) \\ + x_\beta(m,n)e^{j\pi(m+n)}, \end{aligned} \tag{2}$$

where the components x_L, x_α and x_β are defined as a linear transformation of the latent RGB color pixels:

$$\begin{bmatrix} x_L(m,n) \\ x_\alpha(m,n) \\ x_\beta(m,n) \end{bmatrix} = \underbrace{\begin{bmatrix} 1/4 & 1/2 & 1/4 \\ -1/4 & 0 & 1/4 \\ 1/4 & -1/2 & 1/4 \end{bmatrix}}_{\stackrel{\text{def}}{=}\boldsymbol{T}} \begin{bmatrix} x_r(m,n) \\ x_g(m,n) \\ x_b(m,n) \end{bmatrix}. \tag{3}$$

Note that Equation (2) is a forward model. That is, given the luma and chroma components (x_L, x_α, x_β) we can determine x_{CFA}. The inverse problem, which is the demosaicking problem, is to determine (x_L, x_α, x_β) from x_{CFA}.

The starting point of frequency selection is to inspect the Fourier spectrum of x_{CFA}. If we take the 2D discrete Fourier transform of x_{CFA}, we can show that the frequency representation of x_{CFA} is given by

$$\begin{aligned} \widetilde{x}_{\text{CFA}}(\mu,\nu) = \underbrace{\widetilde{x}_L(\mu,\nu)}_{\text{base-band}} + \underbrace{\widetilde{x}_{\alpha_1}(\mu-\pi,\nu) + \widetilde{x}_{\alpha_2}(\mu,\nu-\pi)}_{\text{horizontal/vertical side-band}} \\ + \underbrace{\widetilde{x}_\beta(\mu-\pi,\nu-\pi)}_{\text{diagonal side-band}}, \end{aligned} \tag{4}$$

where μ and ν are the 2D angular frequencies and $\widetilde{(\cdot)}$ denotes the Fourier transform of the argument. Equation (4) suggests that the spectrum of the mosaicked image $\widetilde{x}_{\text{CFA}}$ comprises a linear combination of a luma channel $\widetilde{x}_L$, two alpha chroma channels $\widetilde{x}_{\alpha_1}$ and $\widetilde{x}_{\alpha_2}$, and the beta chroma channels $\widetilde{x}_\beta$. Figure 3 illustrates the ideas of the frequency analysis. Given a color image, we can inspect the image generated by the color filter array. In the frequency domain, the luma channel occupies the center of the spectrum, whereas the chroma channels are located on the sides of the spectrum.

The Fourier spectrum in Equation (4) indicates that the CFA is effectively *modulating* the color channels. As such, a natural solution for demosaicking is to *demodulate* x_{CFA} so that we can retrieve (x_L, x_α, x_β). Demodulation is feasible here because we know the CFA and also its *carrier frequency* from basic principles of sampling theorem in 2D domains. Denoting the carrier frequencies for x_{α_1}, x_{α_2} and x_β are $\boldsymbol{\omega}_{\alpha_1}$ and $\boldsymbol{\omega}_{\alpha_2}$, $\boldsymbol{\omega}_\beta$ respectively, then the carriers are defined as (using α_1 as an example):

$$c_{\alpha_1}(m,n) = A_{\alpha_1}\cos\left(\boldsymbol{\omega}_{\alpha_1}^T \begin{bmatrix} m \\ n \end{bmatrix} + \theta_{\alpha_1}\right), \tag{5}$$

where A_{α_1} and θ_{α_1} are the amplitude and the phase offset of the carriers. To demodulate the color, we multiply $x_{\text{CFA}}(m,n)$ with the carriers $c_{\alpha_1}(m,n)$, followed by convolving with a predefined lowpass filter $g(m,n)$:

$$x_{\alpha_1}(m,n) = \left(x_{\text{CFA}}(m,n) \times c_{\alpha_1}(m,n)\right) \circledast g(m,n) \tag{6}$$

The computation for x_{α_2} and x_β is performed in a similar manner. For simplicity, we combine the two α channels by simple averaging:

$$x_\alpha(m,n) = x_{\alpha_1}(m,n) + x_{\alpha_2}(m,n). \tag{7}$$

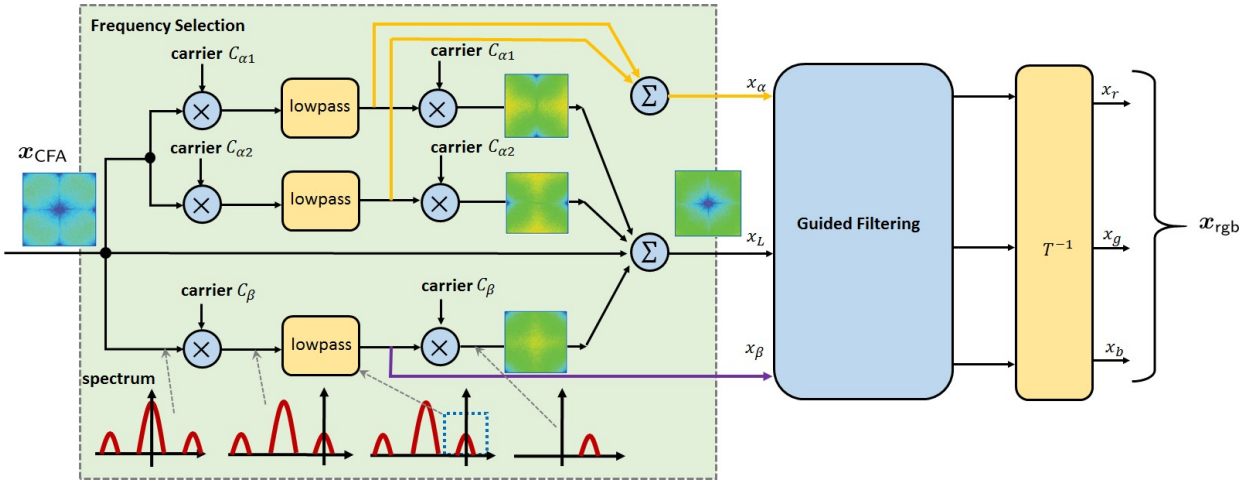


Fig. 4. Implementation of the classical frequency selection. Given the input spectrum $\boldsymbol{x}_{\text{CFA}}$, our goal is to remove the unwanted side-bands. By adopting the classical demodulation scheme, i.e., carrier+lowpass+carrier, we can recover the main lobe. The demodulated signals are x_α, x_β and x_L. After post-processing (typically the luma-denoising) and coordinate transform $\boldsymbol{T}$, we retrieve the RGB signal.

The base-band luma component is recovered by subtracting the re-modulated (i.e., shifted to their original positions) x_{α_1}, x_{α_2} and x_β components from the input CFA image:

$$x_L(m,n) = x_{\text{CFA}}(m,n) - x_{\alpha_1}(m,n) \times c_{\alpha_1}(m,n) - x_{\alpha_2}(m,n) \times c_{\alpha_2}(m,n) - x_\beta(m,n) \times c_\beta(m,n). \quad (8)$$

The demodulation process is pictorially illustrated in Figure 4. The input CFA image is first multiplied by the carriers. In the Fourier domain (the red curves shown at the bottom), the spectrum is shifted according to the carrier frequency. Since we know the CFA, the carrier frequency is deterministic. We then pass the signal through a lowpass filter. Afterward, we multiply the signal with the carriers again to un-shift the spectrum. The whole pipeline is reminiscent of the classical sinusoidal demodulation problem in, e.g., [58, Chatper 8].

To customize frequency selection for our problem, we make the lowpass filters trainable. Specifically, we use three layers of convolutional kernels of size 7×7 to reconstruct the vertical, horizontal, and diagonal chroma channels. During training, filters are regularized by enforcing the ℓ_1 norm of the filter coefficients such that it is equal to unity. This filter learning approach is more flexible than the minimum mean-squared error (MMSE) filter estimation such that those used in the classical literature [33] since it performs the filter estimation jointly with the luma and chroma denoising in an end-to-end training approach.

B. Guided Filtering

The output of the frequency selection stage consists of the luma signal $\boldsymbol{x}_L$, and the two color signals $\boldsymbol{x}_\alpha$ and $\boldsymbol{x}_\beta$. All signals are corrupted by noise because during the frequency selection process, we have only decoupled the colors from the input and have not aggressively removed the noise. The objective of the guided filtering step is to denoise.

The rationale behind the guided filtering step is the different signal-to-noise ratios of $\boldsymbol{x}_L$, $\boldsymbol{x}_\alpha$, and $\boldsymbol{x}_\beta$. The luma signal $\boldsymbol{x}_L$, by definition, is the average of the RGB signals, i.e., $\boldsymbol{x}_L = (\boldsymbol{x}_R + 2\boldsymbol{x}_G + \boldsymbol{x}_B)/4$. This averaging process suppresses the noise more than that of the color channels. Classical papers

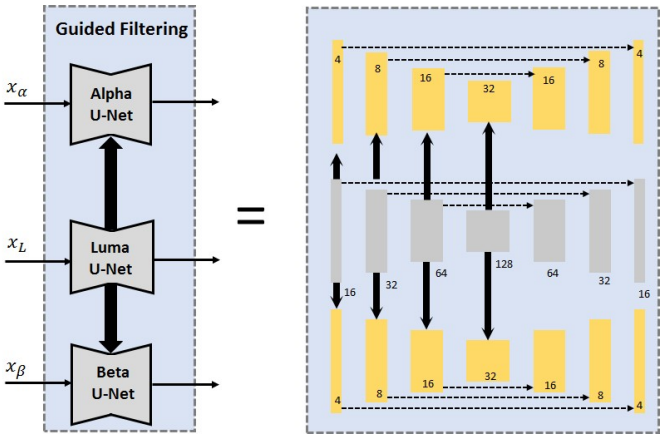


Fig. 5. The proposed guided filtering step consists of three UNets: The luma UNet, and two chroma UNets. Each UNet has a residual connection. Across the networks, we transfer knowledge from the luma channel to the chroma channels by concatenating features.

in color processing have long recognized this phenomenon, e.g., [59], [60]. Instead of independently denoising the three channels or jointly denoising the channels by treating them as a spectral volume, it is more promising to first denoise the luma channel, then use the recovered luma signal to guide the filtering of the chroma signals.

Building upon this intuition, we use three deep networks as shown in Figure 5. The luma denoising network is a standard UNet [61] with the number of layers as shown in the middle of Figure 5. On top of this luma network, we introduce two smaller UNets for the chroma channels. The size of the chroma channel UNet is 4 times less than that of the luma channel UNet. When training the networks, we pull the features generated by the encoder of the luma UNet, and concatenate with corresponding features generated by the chroma UNets. The chroma UNets are benefited from this feature sharing since they can use the high frequency information such as edges and textures from the luma denoiser. The layers of all the three UNets are convolutional, with a kernel size of 3×3. Following [62], [63], the number of

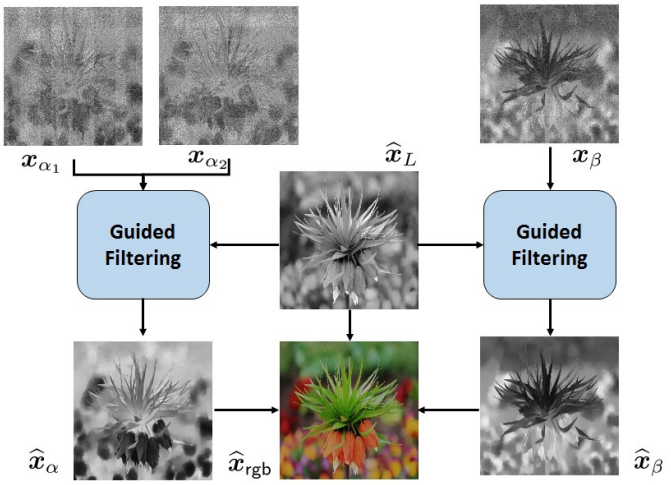


Fig. 6. Visualization of the proposed guided filtering. Given the input $\boldsymbol{x}_{\alpha_1}$, $\boldsymbol{x}_{\alpha_2}$, $\boldsymbol{x}_\beta$ and the luma estimate $\widehat{\boldsymbol{x}}_L$, the guided filter leverages the luma estimate to reconstruct the chroma channels which are otherwise very difficult to recover. The resulting image is $\widehat{\boldsymbol{x}}_{\text{rgb}}$.

feature channels at all image scales is kept fixed to avoid unnecessary enlargement of the network size. We replace the trainable transposed convolution layers in the standard UNet with non-trainable bilinear upsampling layers to reduce the total number of parameters [64].

The effectiveness of the proposed guided filtering step can be visualized in Figure 6. In this example, we generate the input signal using a QIS imaging model with a mean photon arrival rate of 10 photoelectrons per pixel. Given the inputs $\boldsymbol{x}_{\alpha_1}$, $\boldsymbol{x}_{\alpha_2}$ and $\boldsymbol{x}_\beta$, straight-forward denoising on each of these channels independently will be extremely difficult because there is not much signal in the input. With guided filtering, since $\boldsymbol{x}_L$ preserves most of the details of the image, it serves as a strong prior to the chroma channels. As we can observe in Figure 6, the reconstructed chroma channels are substantially better than without the guided filtering.

C. Loss Functions

The overall system is trained end-to-end. To perform the training, we define the overall training loss as

$$\mathcal{L}(f_{\boldsymbol{\Theta}}) = \mathcal{L}_{\text{RGB}}(f_{\boldsymbol{\Theta}}) + \eta_1 \mathcal{L}_{\text{luma}}(f_{\boldsymbol{\Theta}}) + \eta_2 \mathcal{L}_{\text{chroma}}(f_{\boldsymbol{\Theta}}), \quad (9)$$

where $f_{\boldsymbol{\Theta}}$ denotes the model parameterized by $\boldsymbol{\Theta}$. There are three terms contributing to the overall loss.

The RGB loss is defined as the sum of a *mean absolute error* (MAE) loss and the *perceptual* loss. The MAE loss is the ℓ_1 difference between the predicted image $f_{\boldsymbol{\Theta}}(\boldsymbol{x}_{\text{CFA}})$ and the true image $\boldsymbol{y}_{rgb}$, whereas the perceptual loss is the ℓ_2 difference between the feature embedding using a pre-trained network [65]. The intuition is that if our network is performing well, then the features of the reconstructed image should be close to that of the clean image. Here, the features are obtained from a VGG-19 network, although other embeddings can also be used.

The luma and the chroma losses are defined as the ℓ_1 loss between the predicted luma and the ground truth luma, and the predicted chroma and the ground truth chroma, respectively. The hyper-parameters η_1, η_2 are chosen empirically to minimize the validation loss.

D. Implementation

The training of the proposed model is based on the WED database [66] which contains 4744 high-quality color images of natural scenes. Color images are mosaicked using Bayer CFA. We simulate shot noise in QIS using Poisson distribution assuming fixed average photon counts per image, which is sampled uniformly in the range $[1e^-, 10e^-]$ for every patch. We also simulate readout noise by a zero-mean Gaussian with the standard deviation set to $0.25e^-$. Images are then normalized to the range $[0, 1]$ by dividing by thrice the average photon count and clipping to $[0, 1]$.

In every training epoch, 128×128 patches are randomly cropped from each image, and data augmentation is performed by random flipping in the horizontal and vertical directions. Pairs of clean and noisy patches are fed into the network with batch size 64 for training. The network is trained for a total of 1000 epochs in mixed precision using NVIDIA GeForce RTX 2080 GPU with 8GBs of dedicated memory. Adam is used for optimization with a learning rate of 10^{-4} and 10^{-5} for the first and second 500 epochs, respectively.

Training hyper-parameters are $\eta_1 = 1$, $\eta_2 = 1$. The trainable lowpass filter $g(m, n)$ is modeled using a 7×7 kernel. The perceptual loss is based on MSE loss computed at 8^{th} and 35^{th} layers of the VGG-19 network. For validation during training, we compute the average PSNR and SSIM on the 18 color images of McMaster dataset [19] every 50 epochs.

IV. Experiment

A. Evaluation Metric

We use the WED [66] database for training, the McMaster dataset [19] for validation, and 24 images from the Kodak dataset and 100 validation images of Div2k dataset [67] for testing. This ensures no overlap between the three different tasks. We simulate QIS noisy images at signal levels $\{1e^-, 2e^-, \ldots, 10e^-\}$, and we assume that the read noise is $0.25e^-$ rms.

Since no single image quality metric can address all questions, we use the following suite of evaluation metrics:

- Peak signal to noise ratio (PSNR), which is the negative log of the mean squared error between the estimated image and the ground truth image. The mean squared error is computed per pixel per color. The is the usual PSNR extended to three color channels.
- Structure Similarity Index Metric (SSIM) [68], used to quantify the visual difference between two images.
- CIE 2000 color error metric, which measures the mean square color difference in the CIELAB color space as suggested by the CIE 2000 standard [69]. CIE 2000 metric captures better the color difference and is used in previous literature in color filter arrays [21], [35]. We use the implementation of skimage color module in [70] to compute the CIE 2000 metric.

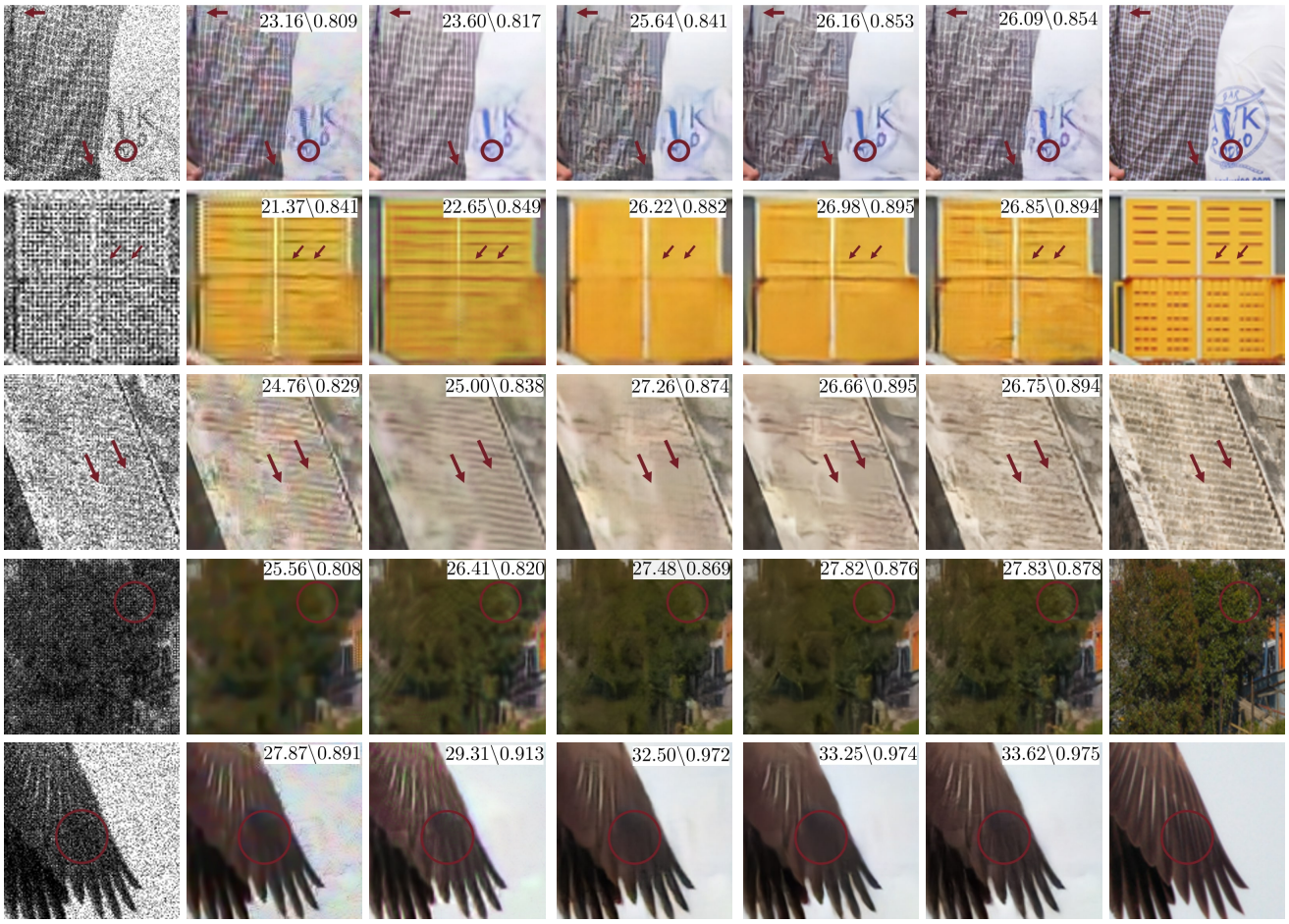


Fig. 7. Synthetic experiments using data from the DIV2k dataset. We simulate the QIS data at an average photoelectrons level of $10e^{-}$ per pixel. Quantitative image quality metrics PSNR\SSIM are shown on the top right of every image.

- Learned Perceptual Image Patch Similarity (LPIPS) metric, proposed by Zhang et al. [71]. This metric measures the perceptual difference between 2 images using deep features. According to [71], this metric shows a good correlation with human perception. To compute the deep features, we use AlexNet.

B. Competing Methods

We compare the proposed method with the following classical and deep learning-based methods. We acknowledge other approaches proposed in the past and recent years. However, due to limited space, we have chosen to focus on a few representative ones.

- Classical frequency selection (FS) by Condat [32], [33]. This method uses a linear demodulation scheme by decorrelating the luma and chroma from the color data. The denoising step uses a follow up work by Condat and Mosaddegh [23] which applies total variation minimization. This is a CIS-based method, so we apply Anscombe transform [44] before the reconstruction to stabilize the noise variance.
- Classical optimization-based algorithm, using the alternating direction method of multiplier (ADMM). We use the implementation in [20]. The backbone optimization algorithm is the Plug-and-Play ADMM (PnP-ADMM) where we use BM3D as the denoiser [60].
- Demosaic-Net by Gharbi et al. [40]. This is a generic deep neural network solution designed for CIS. The method has a standard convolutional structure that takes a raw CFA color image and converts to a full RGB image. We used the code provided by the authors of [40], and trained the network from scratch using QIS data
- MM-Net by Kokkinos and Lefkimmiatis [28]. The method combines an explicit forward degradation model with a deep neural network. The algorithm iterates like classical optimization approaches, but each iteration is executed by a deep network. We used the code provided by the authors of [28], and trained from scratch using QIS data and the pre-trained denoiser provided by the author.

C. Results of Synthetic Experiments

In this section, we report the results of synthetic experiments. We first offer some visual comparisons. Figure 7 shows a few snapshots of several testing images taken from the Div2K dataset. A few observations we can see from these images:

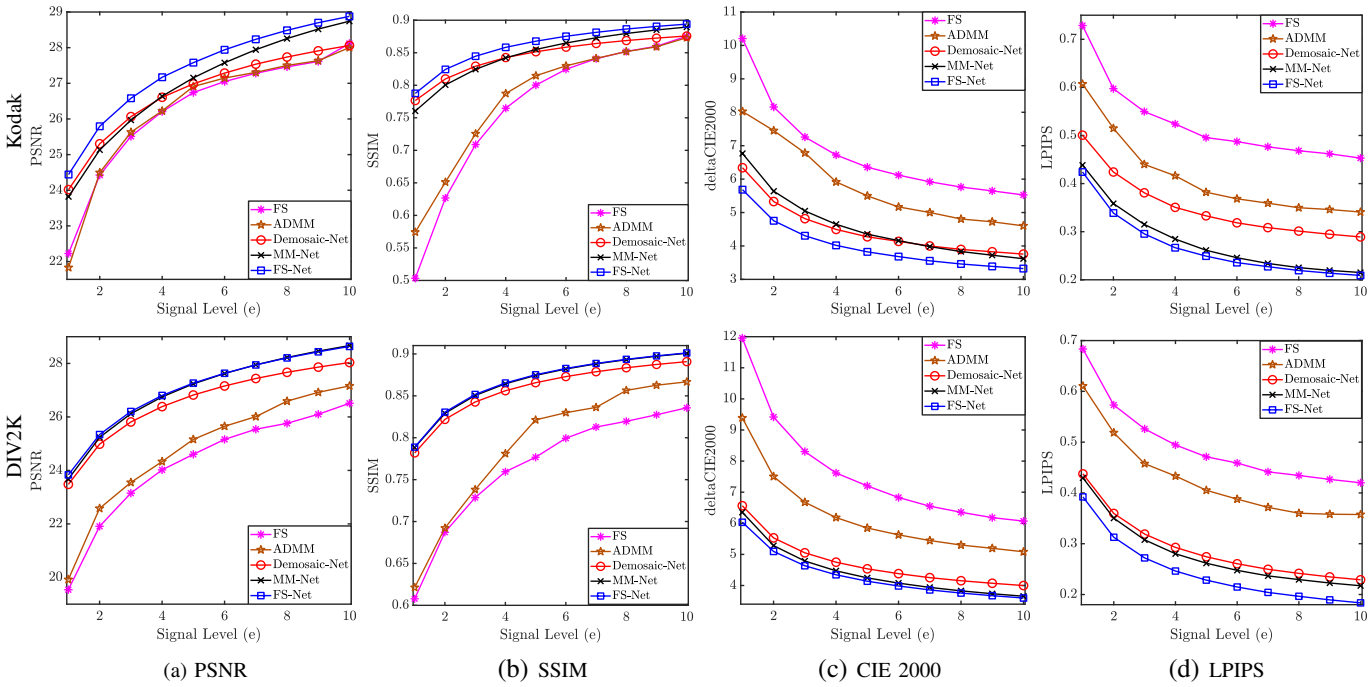


Fig. 8. Performance of synthetic QIS data with read noise = $0.25e^-$ at different signal levels. The first row shows the results of using 24 images in the Kodak dataset, whereas the second row shows 100 images of the DIV2K dataset.

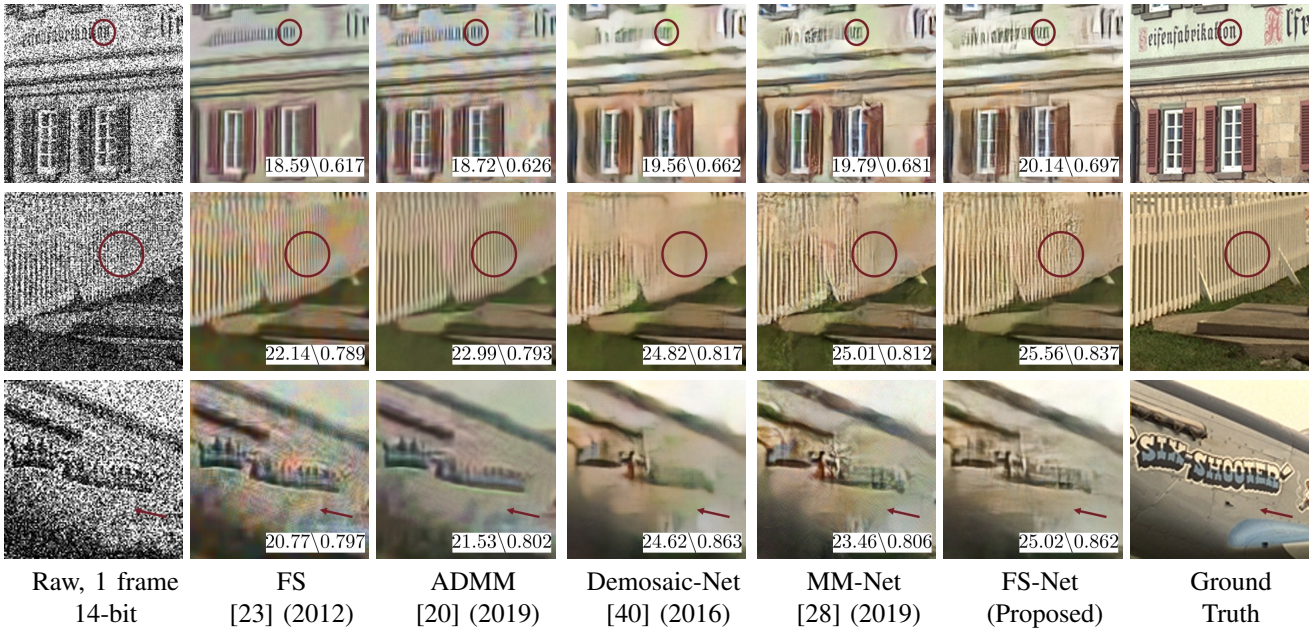


Fig. 9. Visual evaluation using the Kodak dataset (synthetic dataset). The QIS data is simulated at $2e^-$. Observe the strong color noise in classical methods such as FS, and over-smoothing in learning-based methods. Quantitative image quality metrics PSNR\SSIM are shown on the top right of every image.

- Details: It is evident from the first three rows of the images that the classical FS [23] and the ADMM approach [20] have hallucinated the details whereas deep learning approaches tend to over smooth the details. The proposed method gives a more faithful recovery of details.
- Color: The proposed method has less false colors than classical methods. This is especially obvious when we compare the stair-case image and the feather image, where color bleeding of the classical methods is severe.

A quantitative comparison is shown in Figure 8, where we compared the PSNR, SSIM, CIE 2000, and LPIPS curves as a function of the photon level. We separately consider the Kodak dataset and the Div2K dataset so that we can see the generalization capabilities of the methods. We make some observations:

First of all, the performance of the competing methods is consistent for both datasets. In particular, we observe that learning-based methods are generally better than classical methods across different photon levels. The gap appears smaller at stronger photon levels ($10e^-$). Learning-based methods are very competitive, especially between MM-Net and the proposed method, using PSNR and SSIM. However,

Raw, 1 frame 14-bit | Demosaic-Net [40] (2016) | MM-Net [28] (2019) | FS-Net (Proposed) | Ground Truth

Fig. 10. Visual evaluation of CIS data using the Div2K synthetic dataset. The CIS data is simulated at $2e^-$ average signal level and $2e^-$ read noise.

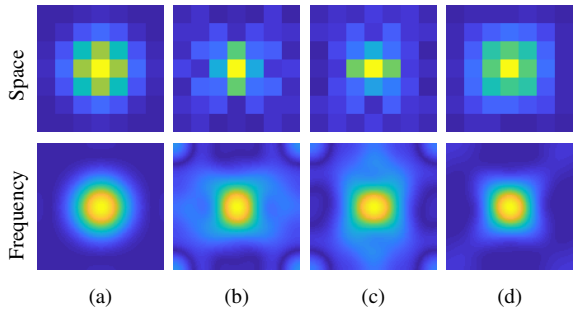


Fig. 11. (a) Initial Gaussian estimate of low-pass filters and the learned filters by FSNet: (b) $g_{\alpha 1}$, (c) $g_{\alpha 2}$ and (d) g_{β}. First and second rows show the spatial and frequency representations, respectively.

as we have seen in the visual comparisons in Figure 7, similar PSNR values can be drastically different visual appearance. Further evidence can be seen from the CIE 2000 metric and the LPIPS metric. In both metrics, the proposed method has a more obvious gap compared to MM-Net and Demosaic-Net.

Second, as photon level drops, the gap between the proposed method and the competing methods becomes larger. This is particularly evident in the Kodak dataset where the proposed method is almost 0.5dB higher than MM-Net and Demosaic-Net in terms of PSNR. The visual comparison in Figure 9 confirms these numbers. The proposed FS-Net has a better recovery of both details and colors.

We also simulated a higher readout noise with standard deviation of $2e^-$, which is similar to what we can get from a standard CIS, and an average signal level of $2e^-$. We ran the same networks used in the previous experiment: Demosaic-Net, MM-Net, and FS-Net on a sample image from Div2K dataset. Figure 10 shows that FS-Net can still give more details and less color noise compared to other networks.

As shown in Figure 11, we show the spatial and spectral representation of the isotropic Gaussian initialization for the low-pass filters as well as the learned filters for the three chrominance channels $\alpha 1$, $\alpha 2$ and β. We notice that the network tends to maximize the spectral support of the chrominance filters to make sure that no chrominance information is lost. Although a wider spectrum implies more noise, this noise is removed afterward in the guided denoising phase. At the same time, the filter response is zeroed-out at the positions of the luminance channel to avoid aliasing.

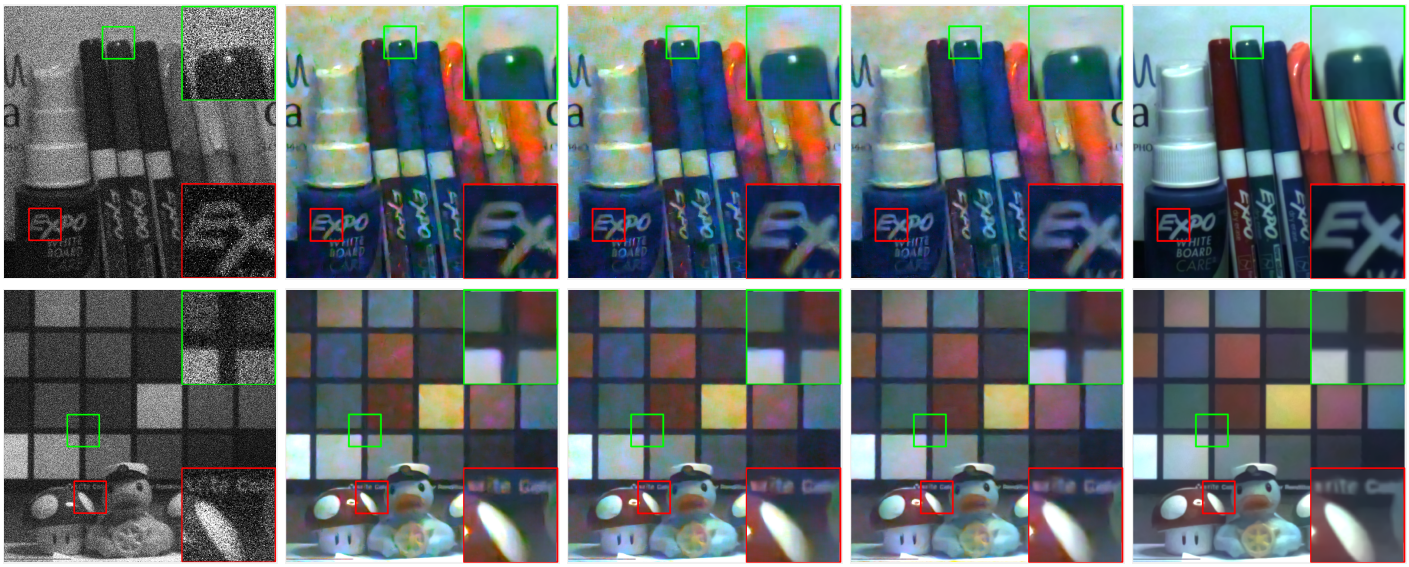
D. Real QIS Data

In this section, we report our experimental results on real QIS data. To conduct the experiments, a QIS camera module developed by Gigajot Tech Inc was used. The sensor has a resolution of 1024×1024 and is equipped with a Bayer color filter array. During the experiments, short exposure times of 74us were used to limit the number of photoelectrons per pixel per frame. Eight (8) "ground truth" images were captured with a longer exposure time of 740us to minimize the impact of photon shot noise, i.e., the average signal level is approximately 10 times the short-exposure images. We then perform the standard linear demosaicking algorithm in [29] to recover the color, and then a simple BM3D denoising algorithm to remove the residual noise. We chose to use the same camera (instead of using a DSLR camera) to obtain the ground truths because this can minimize the impact generated by sensor characteristics and focus the comparison on the color reconstruction.

Figure 12 shows the snapshots of two scenarios. We only compare Demosaic-Net and MM-Net because the synthetic experiments showed that they are better than classical methods. When compared with these methods, we observe that they have more color reconstruction error. For example, in the "Expo" image, Demosaic-Net and MM-Net have severe color bleeding occurring near the cap of the green marker (See the zoom-in). They also have more noise issues showing up in the text regions. In the "Duck" image, we observe similar problems where the color bleeding is severe in the background color chart.

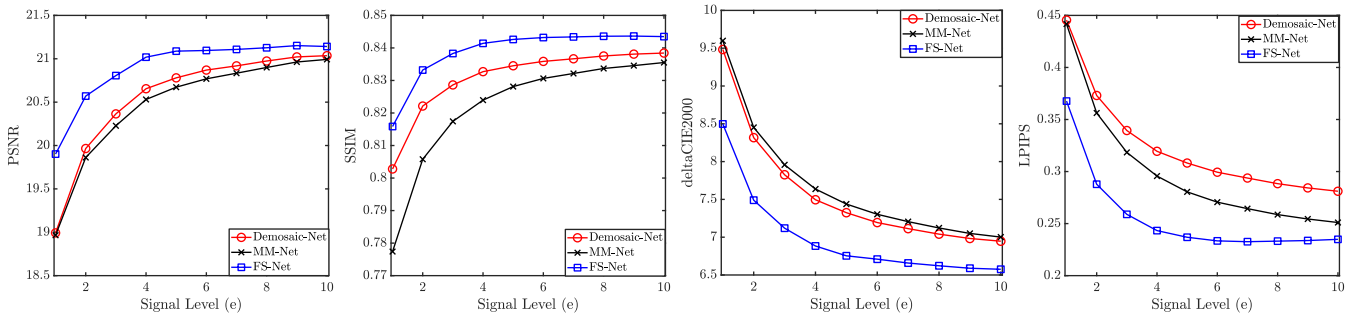
Figure 13 shows the quantitative comparisons using the "ground-truth" images. Two observations are worth mentioning. First, while Demosaic-Net and MM-Net offer very competitive results in the synthetic data, they become worse when it goes to the real data. Note that the training data we use to train Demosaic-Net, MM-Net, and our proposed FS-Net are identical. Thus there is no bias as far as the training data is concerned. We believe that this phenomenon is due to network designs. The results suggest that our physics-inspired color demodulation plus the guided filter is more generalizable to real data where there is uncertainty in the noise model and sensor imperfectness.

The second observation is the inconsistency of different metrics. For example, Demosaic-Net performs similarly with MM-Net in PSNR and CIE2000, but it has better SSIM,



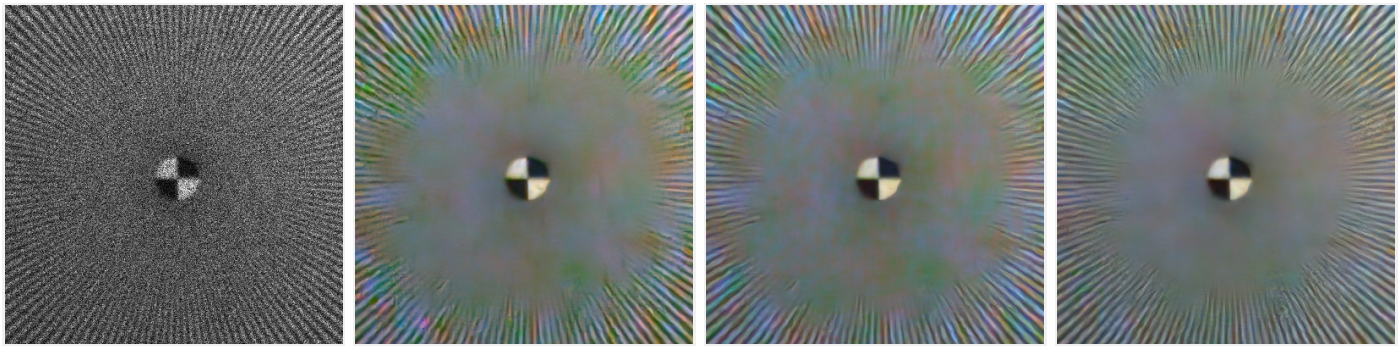
(a) Raw, 1 frame, 14-bit (b) Demosaic-Net (c) MM-Net (d) FS-Net (e) Ground Truth

Fig. 12. Real data reconstruction results. (a) Input, collected at $1.8e^{-}$ (1st row) and 5.5e (2nd row), (b) Demosaic-Net, (c) MM-Net, (d) Proposed FS-Net, and (e) Ground truth, obtained by very long-exposure, with post-processing using demosaicking and denoising.



(a) PSNR (b) SSIM (c) CIE 2000 (d) LPIPS

Fig. 13. Reconstruction results using real data. The performance evaluation is based on PSNR, SSIM, CIE2000, and LPIPS metrics.



Noisy Demosaic-Net MM-Net FS-Net

Fig. 14. Grayscale resolution chart captured with Integration time 224 μsec, lens aperture f/1.8 and average signal level per frame is $8.4e^{-}$. For ideal demosaicking, the result should not have any false colors.

and worse LPIPS. This suggests that no single metric can provide a complete description of the methods. Nevertheless, the proposed FS-Net has consistently much better results than Demosaic-Net and MM-Net in all evaluation metrics.

To further evaluate the performance of the methods, we captured a real resolution chart using a QIS and reconstruct the image. Figure 14 shows the visual comparison. It is evident from the figures that Demosaic-Net and MM-Net have more false colors. Besides, the resolution that can be recovered by these two methods is less than that of the proposed FS-Net.

Finally, in Figure 15 we show the results using 2-bit and 3-bit data. As we can see, the performance of the proposed method remains competitive in these low bit-depth situations.

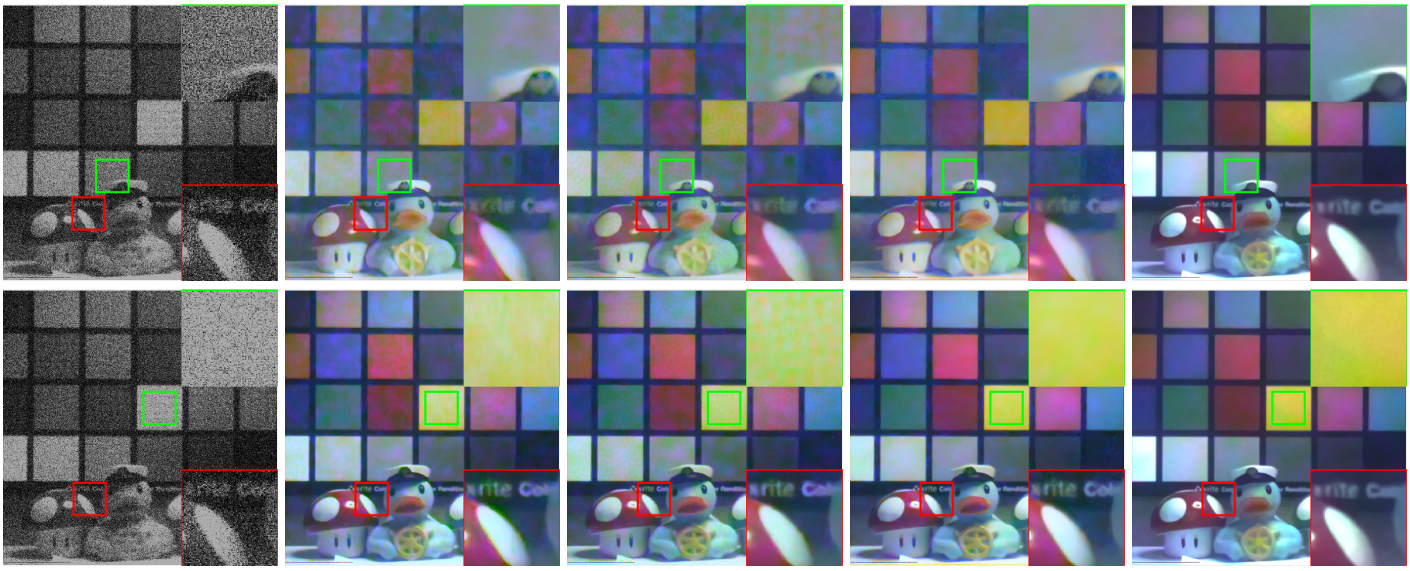


Fig. 15. First row: 10 frames of 2-bit QIS data, average signal per frame: $1.86e^{-}$. 2nd Row: 10 frames of 3-bit QIS data, average signal per frame:$3.8e^{-}$

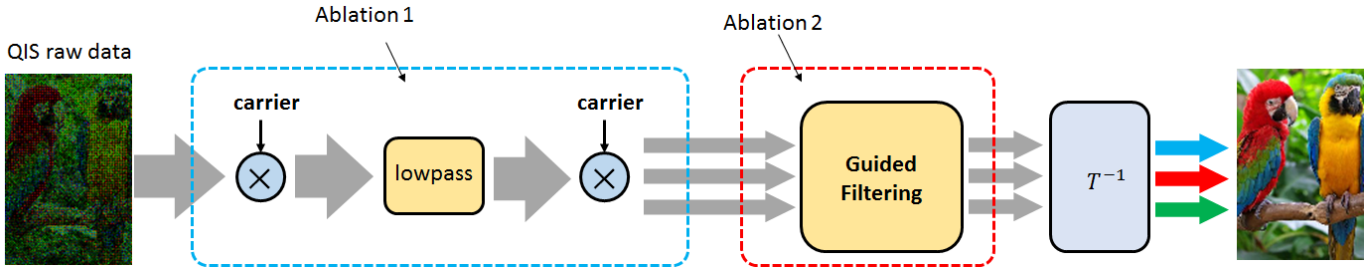


Fig. 16. Ablation study overview. There are two studies. Ablation study 1 concerns about the effectiveness of the demodulation performed by the frequency selection. Ablation study 2 evaluates the significance of the guided filtering step.

TABLE I
COMPARISONS OF NETWORK COMPLEXITY AND SIZE

Method	Demosiac-Net	MM-Net	FS-Net
Complexity (GMAC)	2.20	12.44	2.76
# Parameters	524k	380k	546k

E. Computational Complexity

To evaluate the network complexity, we compute the theoretical total number of multiply-add operations (MAC) for each network using code in [72]. We also compare the total number of parameters for each network to evaluate the required memory for the network weights.

Table I shows the comparison between Demosaic-Net, MM-Net with 2 iterations, and FS-Net. We notice that DemosiacNet has the least complexity since it processes a down-scaled version of the image, and then do up-sampling at the end of the network. However, this down-scaling leads to a loss of resolution as we saw in visual comparison. MM-Net has the least number of parameters since it reuses the same network for multiple iterations. However, it has the highest complexity. FS-Net has comparable complexity and network size to Demosaic-Net while achieving superior image quality.

F. Ablation Study

In this section, we report the ablation study results. All our ablation analysis is based on synthetic experiments. The photon level is set as $5e^{-}$. We use WED dataset for training, and Kodak dataset for evaluation. Figure 16 summarizes our ablation study. There are two evaluations: (i) The effectiveness of the frequency selection method; (ii) the significance of the guided filtering.

1) Ablation 1: Frequency selection: Our first ablation study concerns the frequency selection. We want to see what will happen if we do not use the frequency selection. However, when the frequency selection is removed, there is essentially no decoupling of the color and so the guided filtering will become invalid. Therefore, if we remove the frequency selection step, the best alternative is just to train an end-to-end network that performs the entire demosaicking and denoising together.

To this end, we compare with Demosaic-Net by Gharbi et al. [40] which is designed to solve the present problem. As we have seen in the synthetic experiments Figure 7, Figure 8 and Figure 9, as well as the real experiments Figure 12 and Figure 14, the proposed frequency selection is evidently better than Demosaic-Net, both visually and quantitatively. We acknowledge that some benefit could be attributed to the other components of the method, e.g., guided filtering

which will be studied in ablation study 2. However, since removing frequency selection is equivalent to abandoning the entire proposed method, the superior performance of the proposed method over Demosaic-Net does provide supports to the proposed method.

2) Ablation 2: Guided filtering: In the proposed method, the denoising step is performed by an explicit guided filtering. This involves using one UNet for the luma channel, and two smaller UNets for the chroma channels. As we have seen in Figure 6, this guided filtering is very effective because the luma signals have substantially better signal-to-noise ratio than the chroma signals. In this ablation study, we want to see what will happen if we replace the guided filtering step with other alternative denoising steps. We compare with several alternative schemes as shown in Figure 17:

- UNet: We use a single UNet with a similar capacity to replace the entire guided filtering module. The UNet takes the noisy luma-chroma signals and denoise them before the color conversion. This alternative scheme can be regarded as asking the network to learn the best denoising protocol instead of using an explicit guided filtering step.
- DnCNN: We replace the UNet with another popular denoising module DnCNN and the denoiser. The goal is to see if other denoisers would make any difference.
- GF-1: We replace the frequency selection step by a standard bilinear interpolation step, and train the network to denoise such an input.
- GF-2: The proposed.

In all these configurations, the loss functions are identical to that of the proposed scheme. That is, we use RGB-loss, luma-loss, and chroma-loss. All networks are trained using the same set of QIS data to ensure fairness.

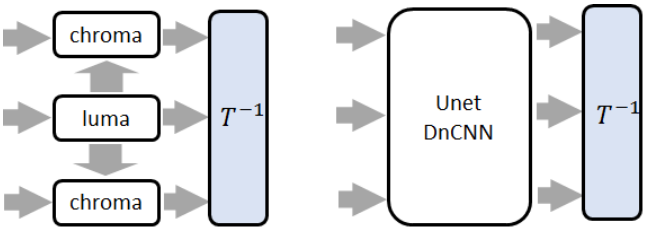


Fig. 17. Ablation study 2. We replace the guided filtering step with various denoising modules. The input arrows denote the input channels which are either RGB or LAB.

The results of this experiment are shown in Table II. In this table, we observe that the proposed guided filtering scheme offers the most competitive result. For example, we achieve 27.58dB while the baseline Demosaic-Net only achieves 26.98dB. Compared to a single UNet which achieves 27.48dB, the proposed guided filtering still offer 0.1dB improvement. In terms of the CIE2000 metric, the improvement is quite substantial. We obtain 3.824 whereas the single UNet obtains 4.064.

V. Conclusion

This paper presents a new color reconstruction method for small pixels. When pixels are small, the problem is unique in the sense that the typical aliasing problems of large CMOS pixels become less influential. This allows us to exploit the classical frequency selection approaches where the color can be demodulated using the known carrier frequencies. However, frequency selection alone has limitations when the photon level is extremely low because the shot noise is strong. We overcome this challenge by integrating the frequency selection method with a deep neural network-based guided filtering step, where we use the luma channels to assist the denoising of the chroma channels. Experimental results based on CIS and QIS provide strong evidence that this physics-based design offers more competitive performance in low-light. Future research should focus on compressing the network capacity while preserving the image reconstruction quality.

Acknowledgment

The authors are grateful for the technical discussion and support from colleagues at Gigajot Technology including Saleh Masoodian, Yu-Wing Chung, and Eric Fossum. The project is partially funded by the NSF Small Business Innovative Research (SBIR) program. A. Gnanasambandam and S. Chan are supported, in part, by the US National Science Foundation under grant CCF-1718007.

References

[1] "Pixel pitch." https://letmaik.github.io/pixelpitch/.
[2] N. A. W. Dutton, L. Parmesan, A. J. Holmes, L. A. Grant, and R. K. Henderson, "320 × 240 oversampled digital single photon counting image sensor," in *Proc. Symp VLSI Circuits*, pp. 1–2, Jun. 2014.
[3] N. A. W. Dutton, I. Gyongy, L. Parmesan, S. Gnecchi, N. Calder, B. R. Rae, S. Pellegrini, L. A. Grant, and R. K. Henderson, "A SPAD-based QVGA image sensor for single-photon counting and quanta imaging," *IEEE Trans. Electron Devices*, vol. 63, pp. 189–196, Jan. 2016.
[4] E. R. Fossum, "Gigapixel digital film Sensor (DFS) proposal," *Nanospace Manipulation of Photons and Electrons for Nanovision Systems*, 2005.
[5] E. R. Fossum, "Some thoughts on future digital still cameras," in *Image sensors and signal processing for digital still cameras*, 2006.
[6] E. R. Fossum, "Modeling the performance of single-bit and multi-bit quanta image sensors," *IEEE J. Electron Devices Society*, vol. 1, no. 9, pp. 166–174, 2013.
[7] E. R. Fossum, "Multi-bit quanta image sensors," in *Proc. Int. Image Sens. Workshop*, pp. 292–295, 2015.
[8] E. R. Fossum, J. Ma, and S. Masoodian, "Quanta image sensor: concepts and progress," in *Proc. SPIE 9858, Advanced Photon Counting Techniques X*, vol. 9858, pp. 985804–985804–14, 2016.
[9] E. R. Fossum, J. Ma, S. Masoodian, L. Anzagira, and R. Zizza, "The quanta image sensor: Every photon counts," *MDPI Sensors*, vol. 16, no. 8, 2016.
[10] I. Gyongy, N. Dutton, and R. Henderson, "Single-photon tracking for high-speed vision," *Sensors*, vol. 18, p. 323, January 2018.
[11] A. Gnanasambandam, O. Elgendy, J. Ma, and S. H. Chan, "Megapixel photon-counting color imaging using Quanta Image Sensor," *Optics Express*, vol. 27, pp. 17298–17310, June 2019.
[12] S. Ma, S. Gupta, A. C. Ulku, C. Brushini, E. Charbon, and M. Gupta, "Quanta burst photography," *ACM Transactions on Graphics (TOG)*, vol. 39, Jul. 2020.
[13] A. Gnanasambandam and S. H. Chan, "Image classification in the dark using Quanta Image Sensors," in *Proceedings of ECCV*, 2020. https://arxiv.org/abs/2006.02026.
[14] Y. Chi, A. Gnanasambandam, V. Koltun, and S. H. Chan, "Dynamic low-light imaging with Quanta Image Sensors," in *Proceedings of ECCV*, 2020. https://arxiv.org/abs/2007.08614.
[15] D. Menon and G. Calvagno, "Color image demosaicking: An overview," *Signal Processing: Image Communication*, vol. 26, no. 8, pp. 518 – 533, 2011.

TABLE II
ABLATION STUDY 2. COMPARISON BETWEEN VARIOUS FILTERING OPTIONS.

Method	demodulation	denoising	denoiser input	output	PSNR	SSIM	CIE	LPIPS
Demosaic-Net	end-to-end	end-to-end	RGGB	RGB	26.98dB	0.852	4.274	0.333
Proposed-DnCNN	Freq. Selection	DnCNN	Lab	RGB	26.96dB	0.850	4.466	0.334
Proposed-UNet	Freq. Selection	UNet	Lab	RGB	27.48dB	0.866	4.064	0.255
Proposed-GF-1	Bilinear Interp.	GF	Lab	RGB	27.45dB	0.864	3.942	0.263
Proposed-GF-2	Freq. Selection	GF	Lab	RGB	**27.58dB**	**0.868**	**3.824**	**0.249**

[16] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Beyond color difference: Residual interpolation for color image demosaicking," *IEEE Trans. Image Process.*, vol. 25, pp. 1288–1300, March 2016.

[17] J. Wu, R. Timofte, and L. Van Gool, "Demosaicing based on directional difference regression and efficient regression priors," *IEEE Trans. Image Process.*, vol. 25, pp. 3862–3874, Aug 2016.

[18] R. Tan, K. Zhang, W. Zuo, and L. Zhang, "Color image demosaicking via deep residual learning," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 793–798, IEEE, 2017.

[19] L. Zhang, X. Wu, A. Buades, and X. Li, "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *Journal of Electronic Imaging*, vol. 20, no. 2, pp. 1 – 17, 2011.

[20] A. Gnanasambandam, O. Elgendy, J. Ma, and S. H. Chan, "Megapixel photon-counting color imaging using quanta image sensor," *Opt. Express*, vol. 27, pp. 17298–17310, Jun 2019.

[21] O. A. Elgendy and S. H. Chan, "Color filter arrays for quanta image sensors," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 652–665, 2020.

[22] P. Chatterjee, N. Joshi, S. B. Kang, and Y. Matsushita, "Noise suppression in low-light images through joint denoising and demosaicing," in *CVPR 2011*, pp. 321–328, June 2011.

[23] L. Condat and S. Mosaddegh, "Joint demosaicking and denoising by total variation minimization," in *2012 19th IEEE International Conference on Image Processing*, pp. 2781–2784, Sep. 2012.

[24] F. Heide, M. S. Y.-T. Tsai, M. Rouf, D. Pajkak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, J. Kautz, and K. Pulli, "FlexISP: A flexible camera image processing framework," *ACM Trans. Graph.*, vol. 33, pp. 231:1–231:13, Nov. 2014.

[25] D. Khashabi, S. Nowozin, J. Jancsary, and A. W. Fitzgibbon, "Joint demosaicing and denoising via learned nonparametric random fields," *IEEE Trans. Image Process.*, vol. 23, pp. 4968–4981, Dec 2014.

[26] T. Klatzer, K. Hammernik, P. Knobelreiter, and T. Pock, "Learning joint demosaicing and denoising based on sequential energy minimization," in *2016 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–11, May 2016.

[27] H. Tan, X. Zeng, S. Lai, Y. Liu, and M. Zhang, "Joint demosaicing and denoising of noisy bayer images with ADMM," in *Proc IEEE Int Conf Image Process (ICIP)*, Sept 2017.

[28] F. Kokkinos and S. Lefkimmiatis, "Iterative joint image demosaicking and denoising using a residual denoising network," *IEEE Trans. Image Process.*, vol. 28, pp. 4177–4188, Aug 2019.

[29] H. S. Malvar, L. wei He, and R. Cutler, "High-quality linear interpolation for demosaicing of bayer-patterned color images," in *Proc IEEE Int Conf Acoust Speech Signal Process*, vol. 3, pp. iii–485–8 vol.3, May 2004.

[30] K. Cui, Z. Jin, and E. Steinbach, "Color image demosaicking using a 3-stage convolutional neural network structure," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 2177–2181, Oct 2018.

[31] Y. Niu and J. Ouyang, "Cross-channel correlation preserved three-stream lightweight cnns for demosaicking," *CoRR*, vol. abs/1906.09884, 2019.

[32] L. Condat, "A simple, fast and efficient approach to denoisaicking: Joint demosaicking and denoising," in *2010 IEEE International Conference on Image Processing*, pp. 905–908, Sep. 2010.

[33] L. Condat, "A new color filter array with optimal properties for noiseless and noisy color image acquisition," *IEEE Trans. Image Process.*, vol. 20, pp. 2200–2210, Aug 2011.

[34] K. Hirakawa and P. J. Wolfe, "Spatio-spectral color filter array design for optimal image recovery," *IEEE Trans. Image Process.*, vol. 17, pp. 1876–1890, Oct 2008.

[35] L. Anzagira and E. R. Fossum, "Color filter array patterns for small-pixel image sensors with substantial cross talk," *J. Opt. Soc. Am. A*, vol. 32, pp. 28–34, Jan 2015.

[36] J. Ma, S. Masoodian, D. Starkey, and E. R. Fossum, "Photon-number-resolving megapixel image sensor at room temperature without avalanche gain," *Optica*, vol. 4, pp. 1474–1481, December 2017.

[37] S. H. Park, H. S. Kim, S. Lansel, M. Parmar, and B. A. Wandell, "A case for denoising before demosaicking color filter array data," in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, pp. 860–864, Nov 2009.

[38] Q. Jin, G. Facciolo, and J.-M. Morel, "A review of an old dilemma: Demosaicking first, or denoising first?," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, pp. 514–515, 2020.

[39] T. Seybold, C. Keimel, M. Knopp, and W. Stechele, "Towards an evaluation of denoising algorithms with respect to realistic camera noise," in *2013 IEEE International Symposium on Multimedia*, pp. 203–210, Dec 2013.

[40] M. Gharbi, G. Chaurasia, S. Paris, and F. Durand, "Deep joint demosaicking and denoising," *ACM Trans. Graph.*, vol. 35, pp. 191:1–191:12, Nov. 2016.

[41] W. Dong, M. Yuan, X. Li, and G. Shi, "Joint demosaicing and denoising with perceptual optimization on a generative adversarial network," *CoRR*, vol. abs/1802.04723, 2018.

[42] T. Huang, F. F. Wu, W. Dong, G. Shi, and X. Li, "Lightweight deep residue learning for joint color image demosaicking and denoising," in *2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 127–132, Aug 2018.

[43] T. Ehret, A. Davy, P. Arias, and G. Facciolo, "Joint demosaicking and denoising by fine-tuning of bursts of raw images," in *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[44] S. H. Chan, O. A. Elgendy, and X. Wang, "Images from bits: Non-iterative image reconstruction for quanta image sensors," *MDPI Sensors*, vol. 16, no. 11, p. 1961, 2016.

[45] S. H. Chan and Y. M. Lu, "Efficient image reconstruction for gigapixel Quantum Image Sensors," in *IEEE Global Conf. on Signal and Info. Process.*, 2014.

[46] F. Yang, Y. M. Lu, L. Sbaiz, and M. Vetterli, "Bits from photons: Oversampled image acquisition using binary poisson statistics," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1421–1436, 2011.

[47] O. A. Elgendy and S. H. Chan, "Optimal threshold design for quanta image sensor," *IEEE Trans. Comput. Imaging*, vol. 4, pp. 99–111, March 2018.

[48] A. Gnanasambandam, J. Ma, and S. H. Chan, "High Dynamic Range imaging using Quanta Image Sensors," in *Intl. Image Sensors Workshop*, 2019.

[49] J. H. Choi, O. A. Elgendy, and S. H. Chan, "Image reconstruction for Quanta Image Sensors using deep neural networks," in *ICASSP*, 2018.

[50] N. A. W. Dutton, I. Gyongy, L. Parmesan, and R. K. Henderson, "Single photon counting performance and noise analysis of CMOS SPAD-based image sensors," *MDPI Sensors*, vol. 16, p. 1122, Jul. 2016.

[51] M. O'Toole, F. Heide, D. B. Lindell, K. Zang, S. Diamond, and G. Wetzstein, "Reconstructing transient images from single-photon sensors," in *CVPR*, 2017.

[52] D. B. Lindell, M. O'Toole, and G. Wetzstein, "Single-photon 3D imaging with deep sensor fusion," *ACM Trans. Graphics*, vol. 37, no. 4, pp. 1–12, 2018.

[53] C. Callenberg, A. Lyons, D. den Brok, R. Henderson, M. B. Hullin, and D. Faccio, "EMCCD-SPAD camera data fusion for high spatial resolution time-of-flight imaging.," in *Computational Optical Sensing and Imaging*, 2019.

[54] G. Gariepy, N. Krstajić, R. Henderson, C. Li, R. R. Thomson, G. S. Buller, B. Heshmat, R. Raskar, J. Leach, and D. Faccio, "Single-photon sensitive light-in-fight imaging," *Nature communications*, vol. 6, no. 1, pp. 1–7, 2015.

[55] A. Gupta, A. Ingle, and M. Gupta, "Asynchronous single-photon 3D imaging," in *ICCV*, October 2019.

[56] E. Dubois, "Frequency-domain methods for demosaicking of Bayer-sampled color images," *IEEE Signal Processing Letters*, vol. 12, pp. 847–850, Dec 2005.

[57] D. Alleysson, S. Susstrunk, and J. Herault, "Linear demosaicing inspired by the human visual system," *IEEE Trans. Image Process.*, vol. 14, pp. 439–449, April 2005.

[58] A. V. Oppenheim, *Signals and Systems*. Prentice Hall, 2 ed., 1997.

[59] Nai-Xiang Lian, V. Zagorodnov, and Yap-Peng Tan, "Edge-preserving image denoising via optimal color space projection," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2575–2587, 2006.

[60] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space," in *2007 IEEE International Conference on Image Processing*, vol. 1, pp. I – 313–I – 316, 2007.

[61] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), (Cham), pp. 234–241, Springer International Publishing, 2015.

[62] S. Laine, T. Karras, J. Lehtinen, and T. Aila, "High-quality self-supervised deep image denoising," in *Advances in Neural Information Processing Systems 32*, pp. 6970–6980, Curran Associates, Inc., 2019.

[63] K. Imai and T. Miyata, "Gated texture cnn for efficient and configurable image denoising," *arXiv preprint arXiv:2003.07042*, 2020.

[64] T. M. V. Srinivasan, S. Gül, C. Hellge, and W. Samek, "Multi-kernel prediction networks for denoising of burst images," in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 2404–2408, IEEE, 2019.

[65] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision – ECCV 2016* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), (Cham), pp. 694–711, Springer International Publishing, 2016.

[66] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Trans. Image Process.*, vol. 26, pp. 1004–1016, Feb 2017.

[67] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[68] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[69] G. Sharma, W. Wu, and E. N. Dalal, "The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research & Application*, vol. 30, no. 1, pp. 21–30, 2005.

[70] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors, "scikit-image: image processing in Python," *PeerJ*, vol. 2, p. e453, 6 2014.

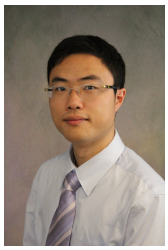
[71] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[72] "Flops-counter." https://github.com/sovrasov/flops-counter.pytorch/.

Omar A. Elgendy (S'12–M'19) received the B.Sc. degree in Electronics and Communications Engineering (Distinction with honors) in 2010 and the M.Sc. degree in Engineering Mathematics in 2015, from Faculty of Engineering, Cairo University, and the Ph.D. degree in Electrical and Computer Engineering from Purdue University, West Lafayette, IN, in 2019. He is currently an image processing specialist at Gigajot Tech Inc., Pasadena, CA.

Dr. Elgendy is a recipient of the Best Paper Award of IEEE International Conference on Image Processing 2016, and the Bilsland dissertation fellowship in 2018 for his PhD. His research interests include statistical image processing with applications to image reconstruction using single-photon imaging sensors, statistical pattern classification, deep learning, and large-scale optimization.

Abhiram Gnanasambandam received the B.Tech. degree in Electrical Engineering from the Indian Institute of Technology Madras in 2017. He is currently working towards the Ph.D. degree from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. His resesarch interest includes computational imaging, image processing, deep learning, and computer vision.



Stanley H. Chan (S'06–M'12–SM'17) received the B.Eng. degree (with first class honor) in Electrical Engineering from the University of Hong Kong in 2007, the M.A. degree in Mathematics from the University of California at San Diego in 2009, and the Ph.D. degree in Electrical Engineering from the University of California at San Diego in 2011. From 2012 to 2014, he was a postdoctoral research fellow at Harvard John A. Paulson School of Engineering and Applied Sciences. He is currently an associate professor in the School of Electrical and Computer Engineering and the Department of Statistics at Purdue University, West Lafayette, IN.

Dr. Chan is a recipient of the Best Paper Award of IEEE International Conference on Image Processing 2016, IEEE Signal Processing Cup 2016 Second Prize, Purdue College of Engineering Exceptional Early Career Teaching Award 2019, Purdue College of Engineering Outstanding Graduate Mentor Award 2016, and Eta Kappa Nu (Beta Chapter) Outstanding Teaching Award 2015. He is also a recipient of the Croucher Foundation Fellowship for Postdoctoral Research 2012-2013 and the Croucher Foundation Scholarship for PhD Studies 2008-2010. His research interests include computational imaging and machine learning.

He is an Associate Editor of the IEEE Transactions on Computational Imaging since 2018, an Associate Editor of the OSA Optics Express in 2016-2018, and an Elected Member of the IEEE Signal Processing Society Technical Committee in Computational Imaging since 2015.

Jiaju Ma received his Ph.D. in Engineering Sciences in 2017 from Thayer School of Engineering at Dartmouth College as the awardee of Charles F. and Ruth D. Goodrich Prize in recognition of his outstanding academic achievements. He co-founded Gigajot Technology, Inc. and has been leading the technological advancement of Quanta Image Sensor and Low-Noise CMOS Image Sensors at Gigajot as Chief Technology Officer since 2017. He has authored and coauthored over 20 technical publications and holds more than 20 patents and patent applications.