IMAGE RECONSTRUCTION FOR QUANTA IMAGE SENSORS USING DEEP NEURAL NETWORKS

Joon Hee Choi, Omar A. Elgendy, and Stanley H. Chan

School of ECE and Dept of Statistics, Purdue University, West Lafayette, IN 47907.

ABSTRACT

Quanta Image Sensor (QIS) is a single-photon image sensor that oversamples the light field to generate binary measurements. Its single-photon sensitivity makes it an ideal candidate for the next generation image sensor after CMOS. However, image reconstruction of the sensor remains a challenging issue. Existing image reconstruction algorithms are largely based on optimization. In this paper, we present the first deep neural network approach for QIS image reconstruction. Our deep neural network takes the binary bit stream of QIS as input, learns the nonlinear transformation and denoising simultaneously. Experimental results show that the proposed network produces significantly better reconstruction results compared to existing methods.

Index Terms— Quanta Image Sensor, single-photon imaging, image reconstruction, deep neural networks

1. INTRODUCTION

Quanta Image Sensor (QIS) is a new type of image sensor envisioned to supersede CMOS and CCD [1]. Having a very small full-well capacity ($1 - 250$ photoelectrons) and single-photon sensitivity, QIS is perceived as an ideal candidate for compensating the deterioration of signal-to-noise ratio in small pixels. The sensor has an extremely high readout rate (10k fps as in [2], and 156k fps in [3]), and can potentially be made for very high spatial resolution [1, 4]. However, the QIS data is binary: A pixel has a value 1 if the photon count exceeds certain threshold, and has a value 0 if the photon count is below the threshold. As a result, non-traditional image reconstruction algorithms are need to recover the images, as illustrated in Figure 1.

Existing image reconstruction methods for QIS are largely based on maximum-likelihood (ML) or maximum a-posteriori (MAP) estimation. These optimizations are done using gradient descent [5], dynamic programming [7] or ADMM [8], which are all time consuming. A significantly faster algorithm is the Transform-Denoise method by Chan et al. [6], where the authors use the variance stabilizing transform (VST) to convert the truncated Poisson random variables to Gaussian, and then apply denoising algorithms for smoothing. In this paper, we propose a deep neural network approach for QIS image reconstruction. As shown in Figure 2, the neural network has better performance than Transform-Denoise by a substantial margin.

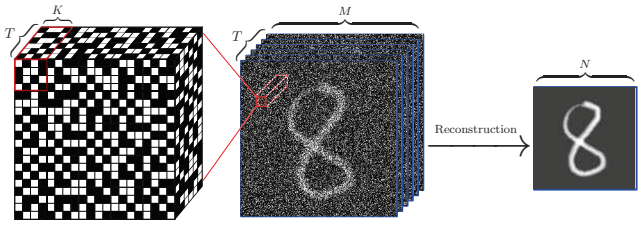
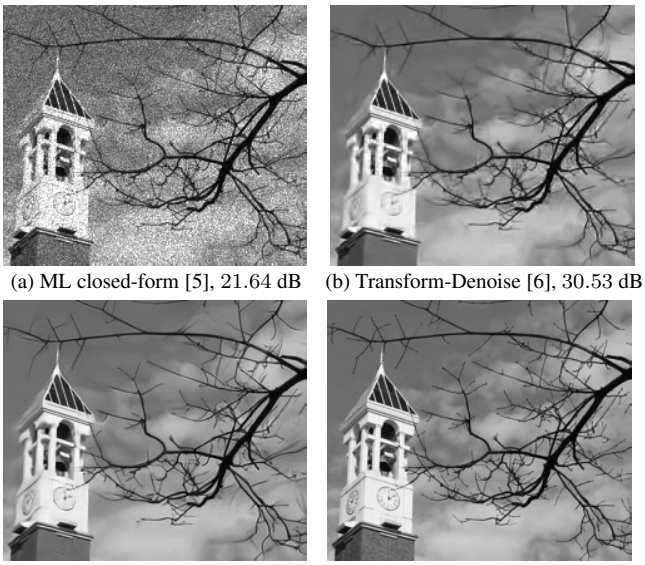


Fig. 1. Image reconstruction of QIS. Given the binary bit planes, the algorithm estimates the gray-scale image shown on the right.



(a) ML closed-form [5], 21.64 dB (b) Transform-Denoise [6], 30.53 dB

(c) Proposed Method, 31.45 dB (d) Ground Truth

Fig. 2. Image Reconstruction using ML [5], TD [6], and our proposed RED-Net method.

Using deep neural networks for image restoration problems is relatively new but has a strong momentum [9–15]. In [16], the authors proposed a neural network to unroll the ISTA iteration with a sparsity prior. However, sparsity priors are generally inferior to discriminative priors learned directly by the neural networks [11]. A simple QIS reconstruction net-

E-mail: {choi240,oelgendy,stanchan}@purdue.edu. This work was supported, in part, by the U.S. National Science Foundation under Grant CCF-1718007.

work is proposed by Rojas et al. [17], where they presented a two-layer neural network to learn the Transform-Denoise pipeline in [6]. However, despite the speed-up offered by the network, the PSNR performance is worse than Transform-Denoise using BM3D as the denoiser.

The key contribution of this paper is a new deep neural network based solution for QIS image reconstruction. Different from [16] which assumes a sparsity prior, our network learns the denoiser directly; And compared to [17], our network has a significantly deeper layer to learn the transformation. We present two designs: one mimics the entire Transform-Denoise pipeline, and the other one substitutes part of the Transform-Denoise pipeline. We show that both networks has significantly better performance than the existing Transform-Denoise method.

2. QIS IMAGING MODEL

In this section we provide an overview of the QIS imaging model. A pictorial illustration is shown in Figure 3. We shall focus on a few important highlights of the model. Readers interested in the details can refer to [5], [6] or [18].

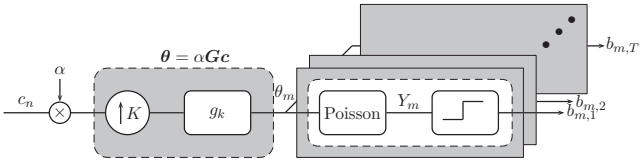


Fig. 3. Image formation process of QIS.

2.1. Spatial-Temporal Oversampling

We model the incoming light intensity as a vector $\boldsymbol{c} = [c_0, \ldots, c_{N-1}]^T$. We assume that c_n is normalized to the range $[0, 1]$ for all n, and use a constant $\alpha > 0$ to model the gain factor.

QIS uses $M \gg N$ jots to *oversample* $\boldsymbol{c}$. The ratio $K \stackrel{\text{def}}{=} M/N$ is the spatial oversampling factor. The oversampling process is modeled by an up-sampling operator and a lowpass filter $\{g_k\}$ as shown in Figure 3. Mathematically, we define the output of the oversampling process as

$$\boldsymbol{\theta} = \alpha \boldsymbol{G} \boldsymbol{c}, \tag{1}$$

where $\boldsymbol{\theta} = [\theta_0, \ldots, \theta_{M-1}]^T$ denotes the light intensity sampled at the M jots, and the matrix $\boldsymbol{G} \in \mathbb{R}^{M \times N}$ is a matrix capturing the upsampling and the lowpass filter $\{g_k\}$.

The lowpass filter $\{g_k\}$ can be arbitrary, e.g., B-spline as mentioned in [5]. However, for efficient reconstruction we shall assume that the filter is box-car. Physically, by using a box-car filter we implicitly assume that the incident light is focused on each jot, which is reasonable to some extent because QIS is equipped with micro-lenses to focus incident light. If $\{g_k\}$ deviates from the box-car but we still use box-car for reconstruction, we say that there is model mismatch, which will be studied in Section 4.

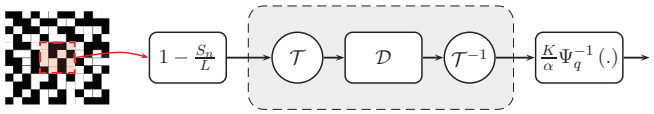


Fig. 4. Transform-Densoise [6]: We apply a pair of transforms $(\mathcal{T}, \mathcal{T}^{-1})$ and a Gaussian denoiser $\mathcal{D}$ for QIS image reconstruction.

2.2. Truncated Poisson Process

The oversampled signal $\boldsymbol{\theta}$ generates a sequence of Poisson random variables according to the distribution

$$\mathbb{P}(Y_{m,t} = y_{m,t}) = \frac{\theta_m^{y_{m,t}} e^{-\theta_m}}{y_{m,t}!}, \tag{2}$$

where $m \in \{0, 1, \ldots, M-1\}$ and $t \in \{0, 1, \ldots, T-1\}$ denote the m-th jot and the t-th independent measurement in time, respectively. Denoting $q \in \mathbb{N}$ as the quantization threshold, the final observed binary measurement $B_{m,t}$ is a truncation of $Y_{m,t}$, i.e., $B_{m,t} = 1$ when $Y_{m,t} \geq q$, and $B_{m,t} = 0$ otherwise. Hence, the distribution of $B_{m,t}$ is

$$\mathbb{P}(B_{m,t} = b_{m,t}) = \begin{cases} \Psi_q(\theta_m), & \text{if } b_{m,t} = 0, \\ 1 - \Psi_q(\theta_m), & \text{if } b_{m,t} = 1. \end{cases} \tag{3}$$

where $\Psi_q : \mathbb{R}^+ \to [0, 1]$ is the upper incomplete Gamma function [19].

The goal of image reconstruction is to reconstruct the underlying image $\boldsymbol{c}$ from the binary measurements $\mathcal{B} = \{B_{m,t} \mid m = 0, \ldots, M-1, \text{and } t = 0, \ldots, T-1\}$ as shown in Figure 1. With the box-car kernel assumption, one can show that the ML solution has a closed-form [6]:

$$\widehat{c}_n = \frac{K}{\alpha} \Psi_q^{-1}\left(1 - \frac{S_n}{L}\right), \tag{4}$$

where $S_n \stackrel{\text{def}}{=} \sum_{t=0}^{T-1} \sum_{k=0}^{K-1} B_{Kn+k,t}$ is the spatial-temporal binning of the binary measurements, and $L \stackrel{\text{def}}{=} KT$ is the combined spatial-temporal oversampling factor.

2.3. Transform-Denoise Approach

Our proposed deep neural network shares some similarity with the Transform-Denoise in [6]. In Transform-Denoise, the key observation is that the random variable S_n in (4) is binomial. The binomial random variable in QIS has spatially varying variance. Thus, one needs to stabilize its variance using variance stabilizing transform (VST). The VST used in Transform-Denoise is the Anscombe binomial transform [20]:

$$Z_n = \mathcal{T}(S_n) \stackrel{\text{def}}{=} \sqrt{L + \frac{1}{2}} \sin^{-1}\left(\sqrt{\frac{S_n + \frac{3}{8}}{L + \frac{3}{4}}}\right). \tag{5}$$

After VST, standard Gaussian denoisers can be used to smooth the image. The final result is obtained by an inverse VST. The overall Transform-Denoise pipeline is shown in Figure 4.

3. PROPOSED METHOD

3.1. Network Structure

The structure of our proposed neural network is shown in Figure 5. We call our network the QISNet. On the network level, QISNet has the same structure as the very deep Residual Encoder-Decoder Network "RED-Net" architecture [21], which was originally proposed for denoising. In this network structure, there is a sequence of N convolutional layers and N deconvolutional layers. The convolutional layers extract the features from the input image, and the deconvolutional layers recover the details lost during the convolutional steps. As mentioned in [22], the deconvolutional layers are necessary for image restoration tasks because the convolutional layers tend to oversmooth the image.

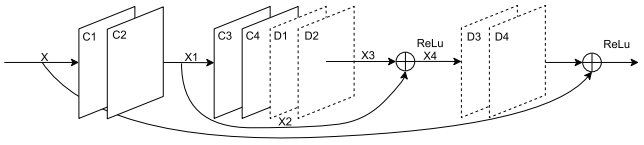


Fig. 5. The proposed QISNet consists of 15 convolutional layers followed by 15 deconvolutional layers.

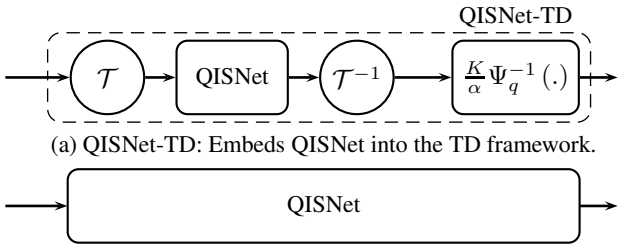
What makes QISNet different from RED-Net is that RED-Net cannot be directly applied to the QIS image reconstruction problem as RED-Net is designed for i.i.d. Gaussian noise. The QIS data, as discussed, is binary following from the truncated Poisson distribution. Therefore, in order to apply the network to QIS, modifications are needed.

Our modification is based on the Transform-Denoise pipeline. The insight is that while individual bits of the QIS data follow a truncated Poisson distribution, the average of the bits within a small spatial-temporal block $1 - \frac{S_n}{L}$ is a Binomial random variable. If we further assume that the blocks do not overlap, then $1 - \frac{S_n}{L}$ can be regarded as an noisy pixel where the distribution is independent (but not identical) Binomial. As a result, if we feed $1 - \frac{S_n}{L}$ into the network, then a denoising network will be sufficient.

3.2. Two Designs for QISNet

Knowing that the input data to the QIS image reconstruction is independent Binomial, we can now design different combinations of the networks for the reconstruction task. Here we present two designs.

The first design is to use the neural network to replace the Gaussian denoiser in Transform-Denoise. We call this design QISNet-TD (See Figure 5(a)). The idea of QISNet-TD is that since the performance of Transform-Denoise depends heavily on the denoiser, we should use a good denoiser. However, we cannot simply put a pre-trained Gaussian noise network denoiser for this task because the pipeline involves other components. We train the network while forcing it to learn the presence of $\mathcal{T}$, $\mathcal{T}^{-1}$ and $\frac{K}{\alpha}\Psi_q^{-1}(\cdot)$.



(a) QISNet-TD: Embeds QISNet into the TD framework.

(b) QISNet: Use QISNet to learn the entire framework.

Fig. 6. The two proposed designs.

The second design is to use the QISNet to replace the entire Transform-Denoise pipeline (See Figure 5(b)). This design is slightly more aggressive as we ask the neural network to learn the denoiser, the nonlinear functions $\mathcal{T}$ and $\mathcal{T}^{-1}$, and $\frac{K}{\alpha}\Psi_q^{-1}(\cdot)$. The difference between QISNet-TD and QISNet is the transforms $\mathcal{T}$ and $\mathcal{T}^{-1}$ (and the nonlinear function $\frac{K}{\alpha}\Psi_q^{-1}(\cdot)$ which is less important here). The inverse transform $\mathcal{T}^{-1}$ is the algebraic inverse, which is a biased inverse transformation. As L grows, the bias of $\mathcal{T}$ will cause the estimate to deviate from its ideal value. Therefore, as one may expect, QISNet-TD performs worse than QISNet in general. We will demonstrate this behavior in the experiment section.

3.3. Training and Parameters

We implement both QISNet-TD and QISNet using 15 convolutional and 15 deconvolutional layers. Each layer uses 3×3 kernels, and 64 feature maps. The network nonlinearity is obtained using ReLu. The training dataset consists of 2000 images selected from the Pascal VOC 2008 dataset [23]. 128 patches of size 50×50 are randomly extracted from each image. The inputs used to train the networks are $1 - \frac{S_n}{L}$, which are images with Binomial "'noise". The ground truths are the clean images. The loss function is L_2-loss, which is optimized using Adam optimizer with a learning rate of 0.0001. The training converges to a local minimum [21] and it takes 8 hours using NVidia Geforce GTX TITAN GPU. For parameters, we set $q = 1$, $\alpha = 2K^2$, and $T = 16$.

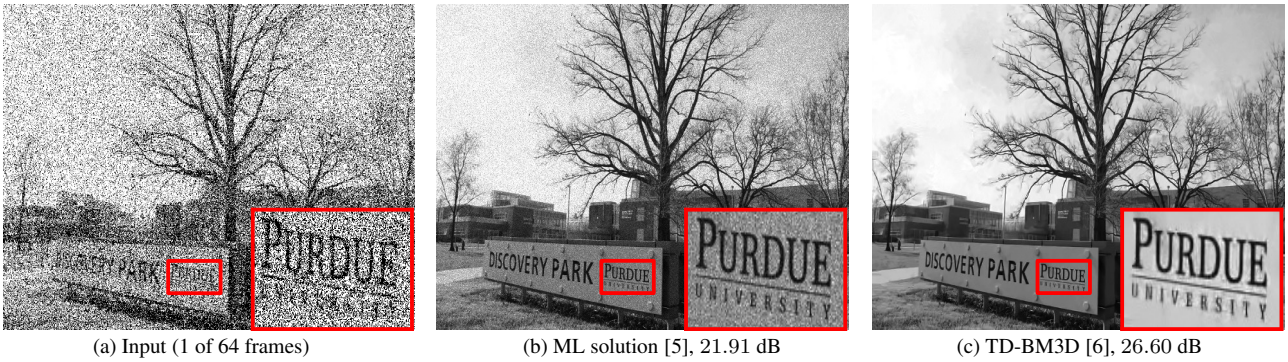
4. EXPERIMENTS

We synthesize QIS data from 77 images captured using a Canon EOS Rebel T6i camera. The images are captured on Purdue campus, which are guaranteed to be different from the Pascal VOC 2008 dataset used for training.

4.1. Reconstruction Quality

We compare the proposed networks with the Transform-Denoise using BM3D [6] and the classical MLE approach [5]. We study two cases: $K = 1$ and $K = 2$. Since $T = 16$, these correspond to $L = K^2T = 16$ and $L = 64$, respectively.

The results of the experiments are shown in Table 1. In this table, we divide the study into two parts. The first part is



(a) Input (1 of 64 frames) (b) ML solution [5], 21.91 dB (c) TD-BM3D [6], 26.60 dB

(d) QISNet-TD, 27.35 dB (e) QISNet, 27.36 dB (f) Ground Truth

Fig. 7. Reconstructed Images and their PSNR for $L = 64$.

the "Match" experiment, where during the QIS data synthesis we assume that the lowpass filter g_k is box-car. It is called "Match" because the variable S_n also assumes a box-car filter.

We observe that while TD-BM3D [6] offers almost 10dB improvement over MLE [5], the proposed networks give additional improvements. QISNet performs as good as than QISNet-TD for small L (27.41dB). For large L, QISNet is better (30.62dB with 30.51dB). This suggests that QISNet is indeed able to learn the transforms $(\mathcal{T}, \mathcal{T}^{-1})$ with sufficient amount of data. Visually, the results in Figure 7 show that the neural networks reconstruct more details.

4.2. Model Mismatch in G

The second part of the experiment is the "Mismatch" case. Here, by mismatch we meant that the box-car filter used in calculating S_n does not match with the lowpass filter used for generating the QIS data. Note that if the lowpass filter g_k is not box-car, one has to use an iterative algorithm such as gradient descent [5] or ADMM [8] to do the reconstruction. Iterative algorithms are not preferred as they are practically slow. Thus it is important to see how well the neural networks can tolerate the model mismatch.

The results of this part of the experiment are shown in Table 1. Our proposed QISNet-TD and QISNet are trained assuming box-car functions. As we can see from the table, as the mismatch becomes worse (from linear to cubic splines), the reconstruction PSNR also drops. However, the PSNR drop in the neural network approaches are not worse than Transform-Denoise. In fact, for all the mismatch filters, the networks still produce better reconstruction quality. One thing to note, however, is that if we know the lowpass filter, we can easily re-train the network to adapt to the filter. Transform-Denoise does not have this flexibility.

Table 1. PSNR in dB for $L = 16$ and $L = 64$

	Method	Mismatch			Match
		Linear	Quad	Cubic	Box-Car
$L = 16$	MLE	15.74	15.69	15.64	15.84
	TD-BM3D	25.67	25.44	25.23	26.40
	QISNet-TD	26.38	26.04	25.74	**27.41**
	QISNet	**26.39**	**26.05**	**25.76**	27.40
$L = 64$	MLE	19.94	19.93	19.92	21.12
	TD-BM3D	25.45	25.40	25.33	29.90
	QISNet-TD	25.51	25.47	25.39	30.51
	QISNet	**25.57**	**25.52**	**25.45**	**30.62**

5. CONCLUSION

We proposed deep neural networks for reconstructing images for Quanta Image Sensors. Our networks can replace the existing Transform-Denoise pipeline, while offering better image reconstruction results. Practically, we anticipate that the networks can eventually be put on neuromorphic chips for better speed and performance.

Acknowledgement: We thank Eric Fossum for offering suggestions to this work, Renan Rojas-Gomez for discussing [17], and NVidia for donating GeForce GTX Titan.

6. REFERENCES

[1] 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.

[2] N. A. W. Dutton, L. Parmesan, A. J. Holmes, L. A. Grant, and R. K. Henderson, "320 $\times$ 240 oversampled digital single photon counting image sensor," in *Proc. Symp VLSI Circuits*, Jun. 2014, pp. 1–2.

[3] I. M. Antolovic, S. Burri, C. Bruschini, R. Hoebe, and E. Charbon, "Nonuniformity analysis of a 65k pixel CMOS SPAD imager," *IEEE Trans. Electron Devices*, vol. 63, no. 1, pp. 57–64, Jan. 2016.

[4] S. Masoodian, J. Ma D. Starkey, Y. Yamashita, and E. R. Fossum, "A 1mjot 1040fps 0.22e-rms stacked bsi quanta image sensor with cluster-parallel readout," in *Proc. Int. Image Sensor Workshop, Hiroshima, Japan.*, 2017, pp. 230–233.

[5] 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, Apr. 2012.

[6] 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, pp. 1961, 2016.

[7] F. Yang, L. Sbaiz, E. Charbon, S. Ssstrunk, and M. Vetterli, "Image reconstruction in the gigavision camera," in *Proc. IEEE 12th Int. Conf. on Computer Vision Workshops (ICCV Workshops), 2009*, Sep. 2009, pp. 2212–2219.

[8] S. H. Chan and Y. M. Lu, "Efficient image reconstruction for gigapixel quantum image sensors," in *Proc. IEEE Global Conf. Signal and Information Processing (GlobalSIP'14)*, Dec. 2014, pp. 312–316.

[9] U. S. Kamilov and H. Mansour, "Learning optimal nonlinearities for iterative thresholding algorithms," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 747–751, May 2016.

[10] A. Kappeler, S. Yoo, Q. Dai, and A. K. Katsaggelos, "Super-resolution of compressed videos using convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP'16)*, Sept 2016, pp. 1150–1154.

[11] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans Image Process*, vol. 26, no. 7, pp. 3142–3155, July 2017.

[12] T. Meinhardt, M. Moeller, C. Hazirbas, and D. Cremers, "Learning proximal operators: Using denoising networks for regularizing inverse imaging problems," Available online at: https://arxiv.org/abs/1704.03488, Apr. 2017.

[13] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Tran Image Process*, vol. 26, no. 9, pp. 4509–4522, Sept 2017.

[14] C. A. Metzler, A. Mousavi, and R. G. Baraniuk, "Learned D-AMP: Principled neural network based compressive image recovery," Available online at: https://arxiv.org/abs/1704.06625, Nov. 2017.

[15] M. Iliadis, L. Spinoulas, and A. K. Katsaggelos, "Deep fully-connected networks for video compressive sensing," *Digit Signal Process*, vol. 72, pp. 9 – 18, 2018.

[16] T. Remez, O. Litany, and A. Bronstein, "A picture is worth a billion bits: Real-time image reconstruction from dense binary threshold pixels," in *Proc. 2016 IEEE Int. Conf. Comp. Photography (ICCP)*, May 2016, pp. 1–9.

[17] R. A. Rojas, W. Luo, V. Murray, and Y. M. Lu, "Learning optimal parameters for binary sensing image reconstruction algorithms," in *Proc. IEEE Int. Conf. Image Process. (ICIP'17)*, Sep. 2017, pp. 2791–2795.

[18] O. A. Elgendy and S. H. Chan, "Optimal threshold design for quanta image sensor," *IEEE Trans. Comput. Imaging*, vol. 4, no. 1, pp. 99–111, March 2018.

[19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications: New York, USA, 1965.

[20] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3-4, pp. 246–254, 1948.

[21] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. Advances in Neural Inf. Process. Syst.*, 2016.

[22] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE CS Conf Comp Vis Pattern Recognition*, Jun 2010, pp. 2528–2535.

[23] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2008 (VOC2008) Results," http://www.pascal-network.org/challenges/VOC/voc2008/workshop/index.html.