End-to-end Learned, Optically Coded Super-resolution SPAD Camera

QILIN SUN, King Abdullah University of Science and Technology
JIAN ZHANG, Peking University
XIONG DUN, Tongji University
BERNARD GHANEM, King Abdullah University of Science and Technology
YIFAN PENG, Stanford University
WOLFGANG HEIDRICH, King Abdullah University of Science and Technology

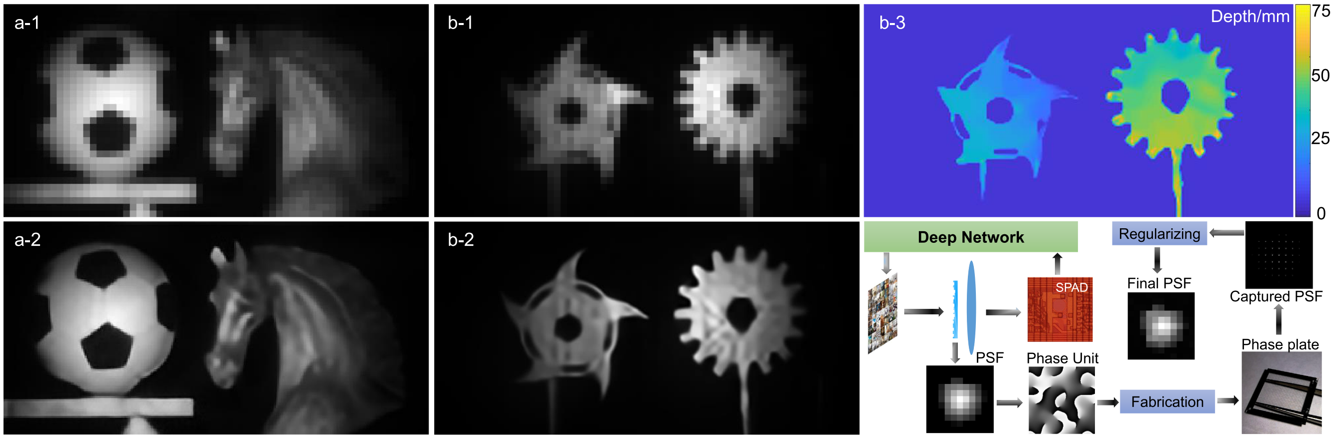

Fig. 1. Overview of our optically coded computational super-resolution SPAD camera. We computationally design phase plates that can suppress aliasing while preserving as much information as possible for super-resolution image reconstruction (right bottom). Fabricated using photolithography technique, this optimized phase plate produces the target PSF at the image plane. In this figure, we demonstrate two representative applications of our optically coded super-resolution SPAD camera: regular intensity imaging, as well as depth estimation, where we obtain high-quality super-resolved (4×) images (a-2) from raw data (a-1) modulated by our phase mask, and super-resolved (4×) intensity (b-2) and depth images (b-3) from the noisy raw data (b-1).

Single Photon Avalanche Photodiodes (SPADs) have recently received a lot of attention in imaging and vision applications due to their excellent performance in low-light conditions, as well as their ultra-high temporal resolution. Unfortunately, like many evolving sensor technologies, image sensors built around SPAD technology currently suffer from a low pixel count.

In this work, we investigate a simple, low-cost, and compact optical coding camera design that supports high-resolution image reconstructions from raw measurements with low pixel counts. We demonstrate this approach for regular intensity imaging, depth imaging, as well transient imaging.

Our method uses an end-to-end framework to simultaneously optimize the optical design and a reconstruction network for obtaining super-resolved images from raw measurements. The optical design space is that of an engineered point spread function (implemented with diffractive optics), which can be considered an optimized anti-aliasing filter to preserve as much high-resolution information as possible despite imaging with a low pixel count, low fill-factor SPAD array. We further investigate a deep network for reconstruction. The effectiveness of this joint design and reconstruction approach is demonstrated for a range of different applications, including high-speed imaging, and time of flight depth imaging, as well as transient imaging. While our work specifically focuses on low-resolution SPAD sensors, similar approaches should prove effective for other emerging image sensor technologies with low pixel counts and low fill-factors.

CCS Concepts: • **Computing methodologies** → **3D imaging**; **Computational photography**; **Antialiasing**;

Additional Key Words and Phrases: SPAD, diffractive optics, super-resolution, depth/transient imaging

ACM Reference format:
Qilin Sun, Jian Zhang, Xiong Dun, Bernard Ghanem, Yifan Peng, and Wolfgang Heidrich. 2020. End-to-end Learned, Optically Coded Super-resolution SPAD Camera. *ACM Trans. Graph.* 39, 2, Article 9 (January 2020), 14 pages.
https://doi.org/10.1145/3372261

Authors' addresses: Q. Sun, B. Ghanem, and W. Heidrich, King Abudullah University of Science and Technology, Al Khwarizmi Bldg 1, KAUST, Thuwal, 23955-6900, Saudi Arabia; emails: {qilin.sun, bernard.ghanem, wolfgang.heidrich}@kaust.edu.sa; J. Zhang, Peking University, Shenzhen Graduate School, Peking University, Nanshan District, Shenzhen, 518055, China; email: zhangjian.sz@pku.edu.cn; X. Dun, Tongji University, Lianqin Building 126, Huxi Campus, Tongji University, No. 500, Putuo District, Shanghai, 200331, China; email: dunx@tongji.edu.cn; Y. Peng, Stanford University, Packard Bldg, 350 Serra Mall, Stanford, CA, 94305-9510, USA; email: evanpeng@stanford.edu.

0730-0301/2020/01-ART9 $15.00
https://doi.org/10.1145/3372261

1 INTRODUCTION

Arrays of Single Photon Avalanche Diode (SPAD) have recently emerged as an alternative hardware solution to photomultiplier tubes (PMT) and streak cameras [Velten et al. 2012, 2013]. Features such as single photon light sensitivity and sub-nanosecond time resolution make this technology promising for many photon-starved applications such as time-of-flight [Shin et al. 2016], transient imaging [Gariepy et al. 2015; O'Toole et al. 2017], fluorescence lifetime imaging [Li et al. 2010; Schwartz et al. 2008] and positron emission tomography [Nemallapudi et al. 2015].

Unfortunately, image sensors built upon SPAD technologies still suffer from low spatial resolution (e.g., 64 × 32) and low fill-factor, i.e., the fact that the light-sensitive area of a pixel is only a small fraction of the pixel's total area (e.g., 3.14% in the MPD-SPC3 SPAD camera used in our experiments). Although recent research prototypes of SPAD arrays have substantially higher pixel counts (e.g., up to 512 × 512 pixels [Ulku et al. 2018]), they still fall short of the resolution of conventional image sensors. Therefore, our method is relevant to the latest generation of prototype SPAD image sensors as well as all commercially available SPAD arrays. Both the limited pixel count (e.g., Shin et al. [2016] and Sun et al. [2018]) and the limited fill-factor and its associated loss in light efficiency [Intermite et al. 2015; Pavia et al. 2014] have been targeted by recent research. However, no definitive solution is available at this time.

To date, computational imaging has achieved tremendous success in the fields of spatial resolution enhancement [Chen et al. 2015; Sun et al. 2018] and defocus deblurring super-resolution [Xiao et al. 2015]. Via point spread function (PSF) engineering [Pavani et al. 2009; Shechtman et al. 2014], researchers have succeeded in localizing microscopic point emitters in a 3D volume by inserting either a spatial light modulator (SLM) or a physical phase plate.

Although optimizing the parameters of diffractive optical elements (DOEs) for a computational camera has been studied intensively, state-of-the-art PSF engineering methods still for the most part do not consider the optical design together with the sensor performance and the reconstruction algorithm in a full end-to-end fashion. A notable exception is a recent work by Sitzmann et al. [2018], which employed an end-to-end optimization that jointly considers optics and image processing to extract optimal PSFs for the purposes of super-resolution and depth of field extension. Although this work takes a significant step towards full end-to-end design of cameras, the reconstruction method used is quite simple and with only fixed blocks; for example, the Wiener deconvolution. In our work, we extend this concept by *jointly* optimizing both the PSF design for the sampling model and the reconstruction algorithm, particularly in the context of a deep neural network.

Putting these pieces together, we aim to overcome the essential spatial resolution limit of SPAD sensors by developing an optically encoded super-resolution SPAD camera with only a single-shot capture procedure. This is achieved by a combination of an optical system that encodes the incident light and a deep neural network that faithfully decodes the high-resolution image. The optical encoding is interpreted as an engineered PSF, acting as an anti-aliasing filter that helps preserve as much information as possible, given the specific sampling pattern of SPAD sensors. We demonstrate significant improvements gained by our prototype when imaging natural scenes. While our method can in principle be applied in any imaging system that employs SPAD array sensors, we focus in particular on three applications: regular intensity imaging (including high-speed imaging), depth imaging, and transient (i.e., light-in-flight) imaging.

Our main technical contributions are as follows:

- We exploit an end-to-end design paradigm for computational super-resolution camera systems, incorporating both PSF design, imaging model, and deep network reconstruction. The system finds optimized compromises between sharpness and anti-aliasing for a given pixel fill-factor.
- We develop a novel single-shot optically coded SPAD camera that achieves an aggressive spatial resolution enhancement of 4×. By simply applying an ultra-thin phase plate that can be easily fabricated and assembled, we achieve an almost zero budget enhancement of hardware configuration.
- We build a prototype with a general phase plate being easily assembled in front of a regular lens. We validate our claims of resolving high-resolution images through simulations and real experiments in normal imaging, high-speed imaging, and time-of-flight (TOF)/transient imaging.

2 RELATED WORK

Computational imaging has been applied in both low-level vision tasks like artifact removal [Peng et al. 2019], and higher-level imaging applications like depth estimation [Levin et al. 2007, 2009]. Particularly, a large amount of work has studied image enhancement using the end-to-end method for applications such as haze removal [Cai et al. 2016], motion deblur [Gong et al. 2017], and time-of-flight imaging [Su et al. 2018]. In the following, we focus on a few more narrow categories of research that are most relevant to our work.

Image Super-resolution (SR). For target applications such as high-speed imaging, fluorescent lifetime imaging, time-of-flight depth or transient imaging, achieving an aggressive resolution enhancement is highly desirable. A large body of work is based on learning the mapping from low-resolution (LR) to high-resolution (HR) images, using techniques such as dictionary learning [Yang et al. 2008, 2010], local linear regression [Timofte et al. 2014; Yang and Yang 2013], random forests [Schulter et al. 2015], and CNNs [Dong et al. 2016a, 2016b; Shi et al. 2016]. Alternatively, one can employ a sparse coding–based network to fully explore the sparsity of natural images [Wang et al. 2015].

Ongoing research efforts have attempted to improve the SR quality using deeper networks [Kim et al. 2016a, 2016b]. Alternative work includes a Laplacian Pyramid SR network [Lai et al. 2017] and an enhanced deep SR network [Lim et al. 2017] that removes unnecessary modules in conventional residual networks [He et al. 2016]. More recently, Haris et al. [2018] proposed a deep back-projection network, exploiting iterative up and down sampling layers and providing an error feedback mechanism for projection errors at each stage.

The mentioned approaches take a traditional image processing approach, whereby the imaging hardware is given and not part of the design decision. Computational imaging approaches, where the imaging hardware and the reconstruction method are *co-designed*, promise improved system performance. This is the approach we take in this work, specifically with the design of an optimal sampling strategy for low-pixel-count, small fill-factor SPAD image sensors.

PSF Engineering for Computational Imaging. The optics and computational imaging communities have widely investigated the deliberate design of (non-Dirac) point spread functions (PSFs) with favorable properties for specific applications. One of the earliest approaches was *wavefront coding*, a method to make the PSF depth-invariant in an attempt to extend the depth of field [Dowski and Cathey 1995; George and Chi 2003]. Recently, the utility of PSF engineering was expanded to 3D to realize a 3D super-resolution effect [Yeh and Waller 2016]. Encoding the aperture of the optical system not only enables recovery of depth information with great fidelity but also generates a high-resolution image image [Levin et al. 2007; Zhou et al. 2011]. Furthermore, coded aperture techniques have been intensively incorporated into compressive sensing [Arce et al. 2014; Llull et al. 2013; Marcia et al. 2009].

Instead of inserting a (usually binary) coded aperture, we investigate the link between the aperture and the image plane in the domain of diffractive optics. By introducing a phase modulation diffractive optical element into the aperture, one has greater flexibility to design the desired PSF in the image plane. There has been a wide range of optimization-based algorithms capable of generating desirable phase or amplitude distributions in both the spatial and the spectral domains. To this end, iterative methods based on Gerchberg-Saxton search, simulated annealing, or direct binary search have been applied to design both monochromatic and broadband DOEs [Kim et al. 2012; Qu et al. 2015].

Another related avenue of investigation is the design of diffractive optical elements to serve as replacements for refractive lenses in imaging systems. Peng et al.'s work on achromatic DOE lenses [2016] started a sequence of DOE design works with similar methodology [Heide et al. 2016; Peng et al. 2018; Petrov et al. 2017]. Instead of automated end-to-end design, the PSF design and reconstruction method are developed separately with a human in the loop. Some recent works [Datta et al. 2018; Zhao et al. 2018] have explored the role of anti-aliasing filters in image super-resolution; however, they use analytical filters (Butterworth and Gaussian, respectively), instead of end-to-end learned ones.

Imaging with SPAD Sensors. Time-correlated single photon counting (TCSPC) [O'Connor 2012] is a common technique for pico-second rate recording of photon events using SPAD arrays. It has been widely applied, for example, in fluorescence lifetime imaging [Li et al. 2010, 2012]. By repeatedly measuring the time duration between a laser pulse and the corresponding transient photon arrival, one can achieve typically sub-nanosecond resolution. Starting with first photon imaging [Kirmani et al. 2014], several approaches have been proposed to abstract the correct temporal information such as temporal deconvolution [Sun et al. 2018], pile-up compensation [Heide et al. 2018; Pediredla et al. 2018] and non-line-of-sight imaging [Heide et al. 2019; Lindell et al. 2019].

To overcome the limitations of low fill-factor and low spatial resolution, researchers have used 2D translation setups to shift a 2D SPAD array with a fixed lens [Shin et al. 2016], or used a galvo mirror setup to scan a 1D line SPAD camera [Lindell et al. 2018; O'Toole et al. 2017]. An alternative approach is the use of DMD-based focal plane spatial modulation to enable a compressive sensing design with SPAD arrays [Sun et al. 2018]. This method requires high precision mechanics and additional imaging optics. Other works have focused primarily on improving the fill-factor of SPAD arrays [Intermite et al. 2015; Pavia et al. 2014].

Although state-of-the-art methods have yielded a reasonable spatial resolution, they are significantly complicating the camera design, and/or require multi-shot image acquisitions, which makes it impossible to image non-repeatable phenomena. We seek a computational super-resolution imaging solution that can maintain all the advantages of SPAD sensors including the snapshot capability, i.e., super-resolution reconstruction from a single image capture.

End-to-end Computational Cameras. Motivated by recent advances in hardware as well as optimization methods, researchers have started to investigate joint optimization over optics like binary masks [Iliadis et al. 2016] for compressive sensing and even sensor structure like a color filter array [Chakrabarti 2016]. More recently, an end-to-end optimization [Sitzmann et al. 2018] over more complicated phase modulation elements was reported. In work parallel to ours, full end-to-end pipelines have been shown recently for the design of depth-encoding PSFs in shape-from-defocus applications [Chang and Wetzstein 2019; Wu et al. 2019].

In addition to conventional imaging applications, diffractive optical elements can also be used as convolutional layers in neural networks [Chang et al. 2018] to speed up the process. Instead, we are inspired to simulate our imaging model for SPAD sensor using a convolutional layer. Taking the convolutional layer into a physical world, we are able to realize the difficult super-resolution task for low fill-factor and low-resolution SPAD sensor by incorporating both optics and deep reconstruction networks.

3 JOINT LEARNING OF OPTICS AND DEEP NETWORK RECONSTRUCTION

We aim to realize super-resolution imaging over a SPAD sensor that suffers from both low resolution and low fill-factor. These two problems will result in significant spatial aliasing and the associated reconstruction artifacts [Parker 2017]. To address this issue, we introduce an optical low-pass filter (OLPF) into the optical system of the camera. The OLPF acts as an *anti-aliasing filter*, which is specially designed to suppress aliasing while preserving as much information as possible for super-resolution image reconstruction.

In our framework, this filter and the matching reconstruction network are *jointly* learned in an end-to-end sense, as illustrated in Figure 2. Specifically, we first synthesize the low-resolution input using a convolutional layer $conv(11, 1)$,[1] representing the PSF and the sensor sampling model, followed by a feature extraction step to generate LR feature maps. Then, at the projection stages a mapping between the LR feature maps and the HR feature maps is built. Finally, a reconstruction step is added to convert the HR feature maps into high-resolution images.

[1] For convenience, we denote a convolutional layer as $conv(f, n)[.]$ and a transposed convolutional layer as $conv^T(f, n)[.]$, where f is the filter size and n is the number of filters.

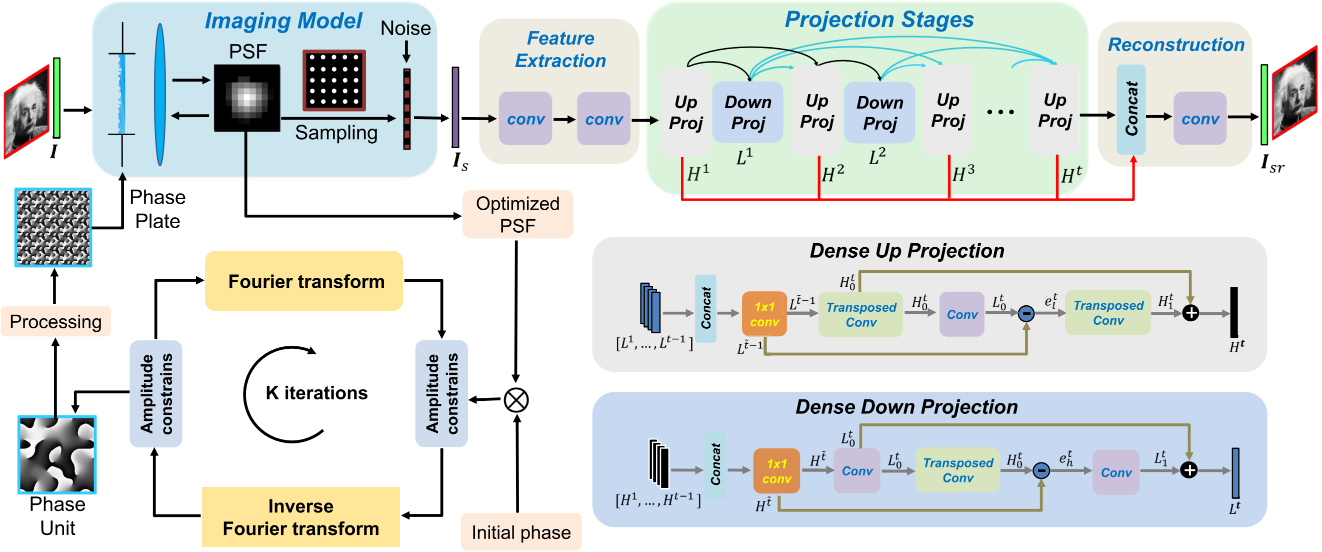

Fig. 2. Framework for joint learning of imaging model and reconstruction. The anti-aliasing filter (PSF) for the low fill-factor SPAD array is learned using our design paradigm. In each forward pass, the synthetic PSF is convolved with a batch of images, and Poisson noise is added to account for sensor's counting noise after the interval sampling process. After obtaining the optimized PSF, we apply a Gerchberg-Saxton–based phase-retrieval algorithm to derive the phase mask. The reconstruction network is composed of three main parts: initial feature extraction, back projection stages, and reconstruction step. The back-projection stage (bottom right), alternating between reconstruction of H^t and L^t, consists of T up projection stages and $T-1$ down projection stages. Each unit is connected with the outputs of all previous units.

After training, we extract the optimal PSF from the weights of $conv(11, 1)$ and then apply a Gerchberg-Saxton (GS)-based phase retrieval algorithm to derive the phase mask (see bottom left of Figure 2), which acts as an optical coder installed at the front focal plane of a regular lens to generate the optimal PSF for later implementations. To account for differences between the design and the fabrication of the phase mask, the real-world PSF of the mask can be calibrated, and the reconstruction network can be fine-tuned through re-training.

In the following, we first detail the image formation model, incorporating the anti-aliasing filter applied to the sampling model of SPAD array and the phase mask optimization to generate the learned PSF combined with a regular imaging lens. Next, we present the deep neural network reconstruction and the time profile sharpening strategy.

3.1 Image Formation

3.1.1 Anti-aliasing Filtering and Image Sampling. As mentioned, the fill-factor of most current SPAD imaging sensors is very low; that is, the light-sensitive area of the pixel is much smaller than the total area occupied by the pixel structure. For example, the SPAD array used in our experiments (MPD-SPC3) has a pixel pitch of 150 μm horizontally and vertically; however, the active area is only 30 μm in each dimension. The physical low pixel count and small fill-factor severely degrade the image quality, creating the desire for super-resolved image reconstruction. To avoid aliasing, the image signal should be pre-filtered with a low-pass filter of the appropriate cut-off frequency, followed by a down-sampling process [Parker 2017]. Again, the goal is to trade-off sharpness and aliasing to find a good compromise that preserves most details of interest.

Due to the low resolution of the sensor array, we can reasonably neglect off-axis aberrations like coma. Image formation becomes a shift-invariant convolution of a latent image with a kernel. To this end, we jointly learn the optimal anti-aliasing filter (e.g., the convolved kernel) and the reconstruction network to eventually preserve the finest details of natural images to realize a super-resolution enhancement. The quantitative evaluation of applying this desired OLPF is detailed in Section 4.

At the position (x, y) on the sensor, the detected signal $I_s(x, y)$ is expressed as:

$$I_s(x, y) = \mathcal{P}(\mathcal{S}(\boldsymbol{p}_\lambda * \boldsymbol{I})), \qquad (1)$$

where $\mathcal{S}$ is a 2D sampling operator corresponding to the physical structure of SPAD sensor, $\boldsymbol{I}$ is the latent image formed on the sensor, $\boldsymbol{p}_\lambda$ is the kernel (or PSF) realized by the optical system, and $\mathcal{P}$ represents a generator of the Poisson noise, which is the appropriate noise model for low-light scenarios that are typical for SPAD imaging.

3.1.2 Learning Optimal PSF Using End-to-end Design. To obtain the optimal PSF $\boldsymbol{p}_{\lambda\text{opt}}$ using our end-to-end framework, we model our PSF as well as the low-resolution sampling process of the SPAD array as a convolutional layer $conv(11, 1)$. In each forward pass, the synthetic PSF (convolutional layer) is convolved with a batch of images, and Poisson noise is added to account for photon shot noise after the interval sampling process. In other words, we represent both the PSF and the sampling process as layers in our neural network during training, and then physically realize the learned result as a custom DOE for our SPAD camera (see Section 3.3).

To determine the size of the kernel, we take a large kernel 21 × 21 at the beginning, and then we found only an 11 × 11 region of the filter had non-zero values. Therefore, we take 11 × 11 as the kernel size of the PSF whose physical dimension is $412.5 \times 412.5 \mu m^2$.

3.2 Image Reconstruction

Image reconstruction is the final stage for applications such as regular intensity imaging or high-speed imaging, and the second last stage for applications such as depth and transient imaging. For our camera the reconstruction is formulated as an optimization problem of a data fitting term with an additional regularization term:

$$\min_{\boldsymbol{I}} \frac{1}{2} \|\mathcal{S}(\boldsymbol{p}_{\lambda\text{opt}} * \boldsymbol{I}) - \boldsymbol{I}_s\|_2^2 + \beta \|\Phi(\boldsymbol{I})\|_1, \tag{2}$$

where $\Phi(\cdot)$ denotes the transform coefficients of $\boldsymbol{I}$ with respect to some transform Φ that can be either linear or optimized non-linear. Sparsity in the transform space $\Phi(\boldsymbol{I})$ is encouraged by the ℓ_1 norm with β being a regularization parameter.

Usually, natural images are non-stationary in classic domains such as DCT, gradients, and wavelets, which may result in an ill-posed problem under such an imaging model. Although an optimized PSF model can preserve a large amount of spatial information, conventional optimization-based methods fail to faithfully reconstruct good quality results when the sampling ratio is very low (e.g., in our case with a sampling ratio only 3.14%). To this end, a trainable architecture for super-resolution with powerful learning ability for features meets our strict requirements as our learned PSF itself encodes features. We choose the state-of-the-art method, dense deep back-projection networks (D-DBPN) [Haris et al. 2018], as our reconstruction network, as shown in Figure 2. The D-DBPN framework introduces an iterative error correcting feedback mechanism to characterize the features in previous layers. More importantly, it addresses the mutual dependency by taking the back-projection from HR domain to LR domain.

3.2.1 *Framework Architecture.*

As shown in Figure 2, the end-to-end framework to obtain the optimal filter and reconstruction network can be divided into four parts:

(a) Imaging model. As we have already discussed in Section 3.1.1, we take the physical imaging model as the first part of our end-to-end framework. The joint framework is used to learn the optimal anti-aliasing filter. After fabricating the filter, we then refine the learning process of the reconstruction network with additional training to account for fabrication errors. For more details, please refer to Section 5.

(b) Initial feature extraction. The initial feature maps L^0 are constructed using a $conv(3, n_0)$ layer to extract features and a $conv(1, n_R)$ layer to pool the features and reduce the dimension from n_0 to n_R. In the experiments, n_0 is set as 256; and n_R, which is the number of filters used in each projection unit, is set as 64.

(c) Back-projection. As illustrated in Figure 2, at tth stage (T = 7 stages in total), the LR feature maps $[L^1, L^2, \ldots, L^{t-1}]$ and HR feature maps $[H^1, H^2, \ldots, H^t]$ are concatenated to be used as input for up- and down-projection units, respectively. In each projection unit, we use a $conv(1, n_R)$ to merge all previous outputs from each unit after the shown concatenation process.

The up-projection is defined as follows:

$$\begin{aligned}
&\text{scale up} && H_0^t = conv^T(f_p, n_R)[L^{t-1}],\\
&\text{scale down} && L_0^t = conv(f_p, n_R)[H_0^t],\\
&\text{residual:} && e_t^l = L_0^t - L^{t-1},\\
&\text{scale residual up:} && H_1^t = conv^T(f_p, n_R)[e_t^l],\\
&\text{output feature map:} && H^t = H_0^t + H_1^t.
\end{aligned} \tag{3}$$

The down-projection is defined as follows:

$$\begin{aligned}
&\text{scale down} && L_0^t = conv(f_p, n_R)[H^t],\\
&\text{scale up} && H_0^t = conv^T(f_p, n_R)[L_0^t],\\
&\text{residual:} && e_t^h = H_0^t - H^t,\\
&\text{scale residual down:} && L_1^t = conv(f_p, n_R)[e_H^l],\\
&\text{output feature map:} && L^t = L_0^t + L_1^t.
\end{aligned} \tag{4}$$

(d) Reconstruction. Finally, we take the concatenated HR feature maps $[H^1, H^2, \ldots, H^t]$ as input and use a $conv(3, 1)$ layer to reconstruct the target HR image.

3.2.2 *Training Details.*

To train the network, we use the mean square error (MSE) loss function. In the stated framework, we use an 8×8 convolutional layer with a stride of four and a padding of two. All convolutional and transposed convolutional layers are followed by a parametric rectified linear unit. We trained our network using the high-resolution images from the DIV2K dataset, using a batch size of 64. For convenience, the LR image resolution was 32×32 (half the size of our SPAD array), and the HR image size was 128×128. We take a convolution layer $conv(11, 1)$ as our PSF following the sampling model of the SPAD sensor to simulate the LR images from HR images. We use ADAM as the optimizer with momentum set to 0.9 and weight decay set to 10^{-4}. The learning rate is initialized to 10^{-4} for all layers and decayed by a factor of 10 for every half of total epochs. All experiments were conducted using Pytorch on a single NVIDIA TITAN Xp GPU. For learning the optimal PSF, we trained the whole framework with 50 epochs taking around 40 hours. After calibrating the PSF generated by the fabricated phase mask, we take the weights of the network trained above as initialization and continue to train the reconstruction network with 11 epochs taking around 8 hours.

3.3 Phase Mask Generation

After obtaining the optimal PSF with our framework, we establish the relationship between the PSF and the phase mask. We first analyze the propagation of light from the phase mask to the image plane, and then present the details of phase mask design.

3.3.1 *Optical Model.*

As shown in Figure 3, the mask is placed at the front focal plane of the lens and acts as the pupil of the whole system. For modeling the light propagation, we apply scalar diffraction theory [Goodman 2005] to approximate the paraxial incident wave. The phase of a complex-valued incident wave is delayed by a phase profile $\phi(x', y')$ proportionally to the height map of a diffractive optical element $h(x', y')$:

$$\phi(x', y') = \Delta n \frac{2\pi}{\lambda} h(x', y'), \tag{5}$$

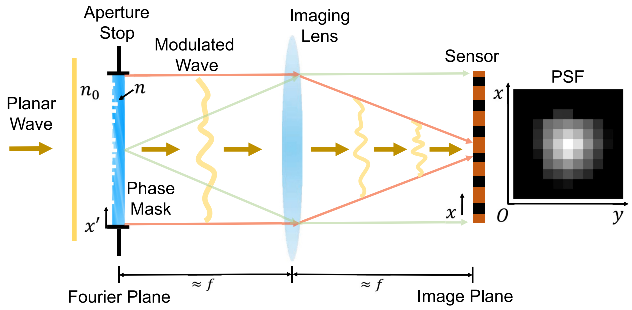

Fig. 3. Illustration of light propagation and desired PSF. The phase mask (i.e., DOE) is set at the equivalent Fourier plane of imaging lens to modulate the incident light and produces the desired PSF on the sensor.

where λ is the wavelength, (x', y') is the location on the phase mask plane, and $\Delta n = n - n_0$ represents the refractive index difference between air (n_0) and the substrate material (n). Placed at the front focal plane of a lens together with our customized limited stop, the phase mask acts as the complex pupil function.

The incident wave field $U_\lambda(x', y', z = 0_-) = A(x', y')\phi_d(x', y')$ is modulated by the phase mask, shown as:

$$U_\lambda(x', y', z = 0_+) = U_\lambda(x', y', z = 0_-) \cdot e^{\mathbf{i}\phi(x', y')}, \tag{6}$$

where we use the notation $z = 0_-$ and $z = 0_+$ to denote positions just before and just after the mask, respectively.

Using the Fresnel approximation, the light propagates through a lens with a focal length f to the image plane is then formulated as:

$$\begin{aligned} U_\lambda(x, y) = {} & \frac{e^{\mathbf{i}kf}}{\mathbf{i}\lambda f} \int\int_\Sigma U_\lambda(x', y', z = f) e^{-\frac{\mathbf{i}k}{2f}(x'^2 + y'^2)} \\ & e^{\frac{\mathbf{i}k}{2f}[(x - x')^2 + (y - y')^2]} dx'dy' \\ = {} & \int\int_\Sigma \phi(x', y') e^{-i2\pi \frac{x'x + y'y'}{\lambda f}} dx'dy', \end{aligned} \tag{7}$$

where $k = 2\pi/\lambda$ is the wave number, (x, y) is the location on the image plane, and $e^{-\frac{\mathbf{i}k}{2f}(x'^2 + y'^2)}$ represents the optical transfer function of the lens. Note that Equation (7) represents essentially a Fourier transform (FT).

For an imaging system, the diffractive PSF on the image plane is eventually obtained as:

$$p_\lambda(x, y) \propto \|(\mathcal{F}\{\phi(x', y')\}\|^2. \tag{8}$$

3.3.2 Phase Retrieval. After deriving the relationship between PSF and the phase mask, we can design a physical height profile $h(x', y')$ on a substrate of refractive index n to implement an image-plane PSF $\boldsymbol{p}_\lambda$ using the Gerchberg-Saxton (GS) [Gerchberg and Saxton 1972] phase retrieval algorithm based on Equation (5).

The core of the phase retrieval is shown on the bottom left of Figure 2. In the beginning, a random phase distribution serves as the initial estimate subject to the amplitude of the PSF. Then, using the initial phase and the amplitude constraint (between 0 and 1) of learned PSF, we apply an inverse Fourier transform on this synthesized complex field function. The resulting phase part of the discrete complex field is preserved while the amplitude part is discarded. In the next round, this preserved phase is plugged into the forward propagation procedure of applying a Fourier transform to update the amplitude estimate of the complex field on the image plane. Eventually, the process is repeated a finite number of times to converge to an optimal phase profile. For more details, please refer to the work by Morgan et al. [2004]. Since we optimize the phase plate for only one wavelength (that of our picosecond laser), we are guaranteed to obtain a phase plate that can generate the optimal PSF we desire. As shown in Figure 4, the correlation coefficient between the PSF generated by phase plate and the learned PSF is 0.9996, and the RMSE between them is 0.0061. This all means the optimal PSF is accurately realized by the phase mask.

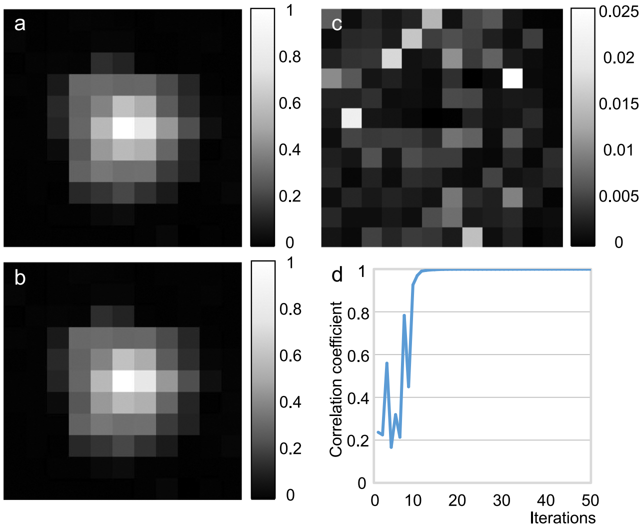

Fig. 4. Efficiency illustration of GS phase retrieval method for our design. (a) Learned PSF; (b) Simulated PSF using the phase profile optimized by GS method; (c) The absolute error between (a) and (b), and the RMSE is 0.0061; (d) The correlation coefficient of the learned PSF and the PSF generated by phase plate, and finally it converges to 0.9996.

3.3.3 Phase Mask Tiling. As shown in Figure 5, a subpixel on the learned PSF has a size of $l_p = 37.5$ μm. Accordingly, the size of phase profile obtained using Equation (7) is $l_u = \lambda f / l_p = 0.8733$ mm, which would make for a very small, square aperture. To design optical systems with larger apertures, one could over-parameterize the design space to optimize the phase profile over a defined larger aperture. This would require a re-design of the pattern for each aperture size and rule out the use of the aperture stop diaphragm in the main camera lens.

A simple alternative that overcomes these issues is to side-by-side replicate the small optimized phase pattern described above to tile the aperture. In our prototype, we tile a square area of edge length $L = 14$ mm, which defines a maximum aperture that can be further stopped down using the lens diaphragm. The tiling has the effect of creating a discrete dot pattern instead of a continuous PSF in the image plane. At a size of $l_p l_u / L = 2.34$ μm, the individual dots are significantly smaller than a sub-pixel, and their center-to-center spacing is exactly the sub-pixel pitch, which also matches the edge length of the light-sensitive area of a SPAD pixel.

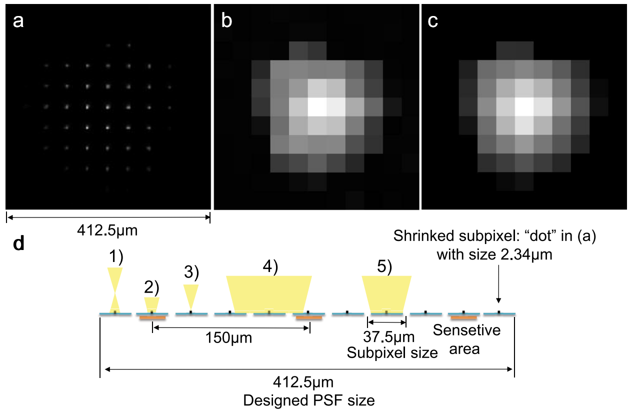

Fig. 5. Calibrating the PSF generated by our fabricated phase plate. (a) Captured PSF; (b) Synthetic learned PSF; (c) Effective PSF as a result of combining PSF (a) with the SPAD pixel sampling pattern; (d) Illustrating the effect of focus on the dot pattern from (a)—see text for details.

Therefore, as the SPAD sensor integrates spatially over the light-sensitive area, it integrates over exactly one of the dots in the dot pattern, which is equivalent to implementing the continuous version of the PSF designed above.

As an added benefit, the dot pattern simplifies the alignment process in the assembly of the optical system. As illustrated in Figure 5, (d1)–(d3), slight defocus does not spread the energy out of the subpixel block. If we were to instead employ a large, non-repeating mask, then a slight defocus would spread energy to neighboring subpixels, equivalent to an additional low-pass filter, as illustrated in Figure 5, (d4)–(d5).

3.4 Temporal Sharpening for Depth and Transient Imaging

To extract temporal information from our reconstructed images, we use a recent reported temporal PSF model [Sun et al. 2018] for SPAD sensors to sharpen our reconstructed 3D data. For depth and transient imaging, our SPAD sensor works in time-correlated single photon counting (TCSPC) mode.

This model is useful for precise temporal localization of Gaussian laser pulses from an observed time profile at each pixel $\boldsymbol{I_i}$, using a model of the temporal response of the SPAD pixel, $\boldsymbol{\Pi}(t)$. The gate signal $\boldsymbol{\Pi}(t)$ is not a simple rectangular pulse, but is distorted according to a resistor-capacitor (RC) circuit response (also compare Figure 6, bottom left). We band limit this RC model with a small Gaussian filter ($\sigma_f = 100$ ps in the experiments)—see Figure 6, bottom center.

The observed time profile at each pixel $\boldsymbol{I_i}$ is then modeled as a convolution of this gate model $\boldsymbol{\Pi}(t)$ with the Gaussian laser pulse $\boldsymbol{G}(t; A, \mu) = A\boldsymbol{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}$, where the parameters A and μ of the Gaussian are initially unknown. They can be determined by solving the following minimization problem for each pixel:

$$\min_{A,\mu} \|\mathrm{G}(t; A, \mu) * \boldsymbol{\Pi}(t) - \boldsymbol{I_i}\|_2^2, \tag{9}$$

where $*$ denotes the convolution. Please refer to the original paper of Sun et al. [2018] for technical details. Instead of using a Gaussian model for the laser pulse, we note that it would be straightforward to substitute other models, such as an exponentially modified Gaussian [Heide et al. 2014] to estimate parameters for inter-reflection, subsurface scattering, or fluorescent lifetime imaging (FLIM).

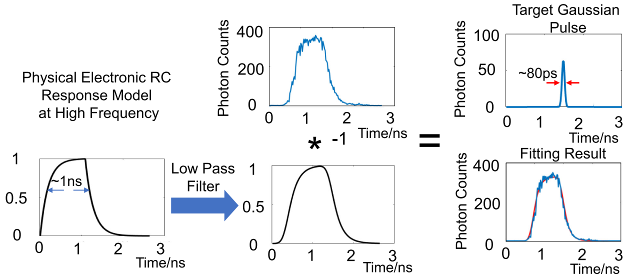

Fig. 6. Modeling the temporal PSF of the system as the convolution of distorted SPAD gate signal and a Gaussian laser pulse profile [Sun et al. 2018]. The data of the histogram is selected from location (45, 167) in Figure 1 (b-3).

Table 1. Quantitative Assessment of Current SR Methods over the Low Fill-factor Sampling Model in PSNR and SSIM (grayscale)

Methods	Set5	Set14	BSDS100
Bicubic	24.26/0.8336	21.51/0.7589	20.83/0.7175
SRCNN	25.27/0.8620	22.34/0.7812	21.58/0.7397
VSDR	25.45/0.8717	22.57/0.7915	21.74/0.7481
Ours	**27.17/0.9019**	**23.97/0.8066**	**23.82/0.7691**

4 EVALUATION IN SIMULATION

We first present a quantitative comparison of some of state-of-the-art SR methods such as VSDR [Kim et al. 2016b] and SRCNN [Dong et al. 2016b]. Table 1 shows that although these kinds of methods perform well on conventional super-resolution problems, they fail in the low fill-factor case. In this table, each of the methods, including our own reconstruction network, was trained using the low fill-factor model (i.e., without an anti-aliasing filter) on the same DIV2K dataset. We also tried VSDR and SRCNN on the optical design obtained with our method, but the resulting SNR and SSIM results are slightly worse than in the low fill-factor case shown in the table.

Next, we present a quantitative comparison of applying our reconstruction network to *four* different sampling models: (1) Low fill-factor sampling model, which considers the SPAD sensor model without the phase mask; (2) Full fill-factor sampling model, which is common for other imaging sensors; (3) Low fill-factor sampling model, which considers the SPAD sensor with a Gaussian PSF of standard deviation $\sigma_N = \sqrt{3\log 2/\pi} \approx 0.459$, corresponding to a least-squares fit of the sinc function that corresponds to the ideal low-pass filter; (4) Our sampling model, which considers the SPAD sensor model with setting the phase mask at the front focal plane of imaging lens.

To make a fair comparison, we use the same training dataset and parameters to retrain the network for the low fill-factor model, full fill-factor model, and a low fill-factor model with a Gaussian PSF. We then assess on three well-known datasets: Set5

Table 2. Quantitative Comparison of 4× Super-resolution under Different Sampling Models in PSNR and SSIM (Grayscale)

Model	Set5	Set14	BSDS100
Low fill-factor	27.17/0.9019	23.97/0.8066	23.82/0.7691
Full fill-factor	29.77/0.9317	26.13/0.8442	25.59/0.8069
Gaussian (optimal)	30.41/0.9360	26.68/0.8498	26.05/0.8157
Gaussian (w./o. re-training)	20.46/0.8087	19.64/0.7268	20.07/0.7016
Ours	**30.76/0.9399**	**26.91/0.8557**	**26.23/0.8198**

σ_N is chosen to best approximate the ideal low-pass filter with a Gaussian (see text).

[Bevilacqua et al. 2012], Set14 [Zeyde et al. 2010], and BSDS100 [Arbelaez et al. 2011]. Table 2 summarizes the averaged PSNR and SSIM scores. We observe that, without the aid of our phase mask, the original low fill-factor model exhibits significantly worse performance than ours both in terms of PSNR and SSIM. Concerning the Gaussian PSF, even in comparison to a perfectly shaped Gaussian diffuser (which would need to be carefully designed, manufactured, and aligned for a specific sensor geometry, and would certainly not be an off-the-shelf part), the scores and recovered image detail (see Figure 8) are still worse than that of our end-to-end system. In addition, we also evaluate a hypothetical full fill-factor model that might be feasible with alternative sensor designs. The results show a clear advantage of our end-to-end design over all alternatives sampling patterns on all datasets.

Figure 7 visualizes several examples selected from the test dataset. Sampling of a low fill-factor sensor destroys most information, thereby the reconstructed results suffer from noticeable artifacts and distortions. These artifacts are alleviated by our proposed method. For instance, the texture on the butterfly is well preserved but is in comparison corrupted by artifacts in the low fill-factor case without the phase mask. The full fill-factor sensor shows slightly better performance than that of the low fill-factor sensor, since it averages the information at all frequencies across the full pixel block. Instead, our sampling model preserves the most desired information, showing reconstruction results closer to ground truth (GT). To this end, we believe our anti-aliasing filtering design contributes to preserving interesting details while suppressing other artifacts.

5 PROTOTYPE AND ASSESSMENTS

In this section, we assess the modulation transfer function (MTF) of our imaging system and present the prototype results of three application scenarios. Before detailing the experimental assessments, we briefly summarize the fabrication of the phase masks and the calibration of the PSFs.

5.1 Prototype

Fabrication. The phase mask is discretized into eight levels that can then be realized by repeatedly applying photo-lithography and reactive ion etching (RIE) three times [Morgan et al. 2004; Peng et al. 2016] on a 0.5 mm Fused Silica substrate. The principal wavelength is 655 nm and a 2π phase modulation is used to wrap the height map. Refer to the supplemental document for fabrication details.

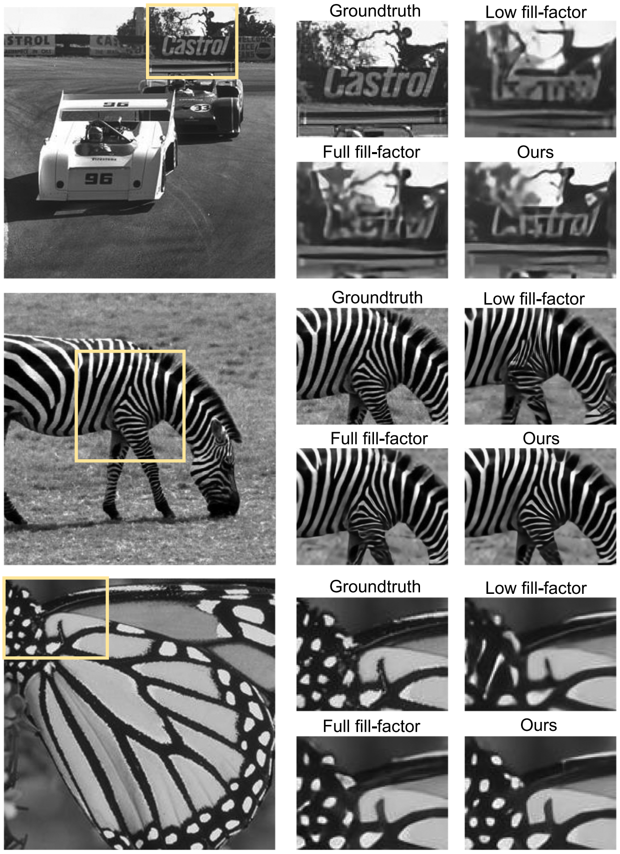

Fig. 7. Selected examples of 4× super-resolution under different sampling models. For the low fill-factor case, we directly apply the low fill-factor model of SPAD to sample the high-resolution images to obtain 1/4 resolution images. For the full fill-factor case, we average the 4×4 pixel area to obtain 1/4 resolution images. For our method, we apply the low fill-factor sampling model of SPAD with pre-filtering using our learned PSF kernel.

We use a FLIR mono sensor GS3-U3-50S5M with a pixel pitch of 3.45 μm to calibrate the PSF of the fabricated phase plate. The phase plate is placed at the front focal plane of a Canon 50 mm lens. A point light source with a 655 nm/10 nm bandpass filter is set 1.35 m away from the sensor. Figure 5(a) shows the calibrated PSF of our fabricated phase mask (see Section 3.3). The sparse dot pattern structure is due to the tiling of the phase plate as described in Section 3.3.3.

5.2 MTF Analysis

We use the slanted edge method [Burns and Williams 2002] to assess the modulation transfer functions (MTFs) of our results and that of the low-resolution reference, as shown in Figure 9. We observe outliers larger than 1 in the plot of the SR image without phase mask (orange plot). In contrast, the MTF of our super-resolution camera is closer to the desired MTF in optical systems: smoothly and monotonously decreasing from an amplitude of 100% for the DC term to ~10% at the Nyquist limit of the SR image, with no erroneous maxima for higher frequencies. This result is enabled by better preservation of super-resolution information

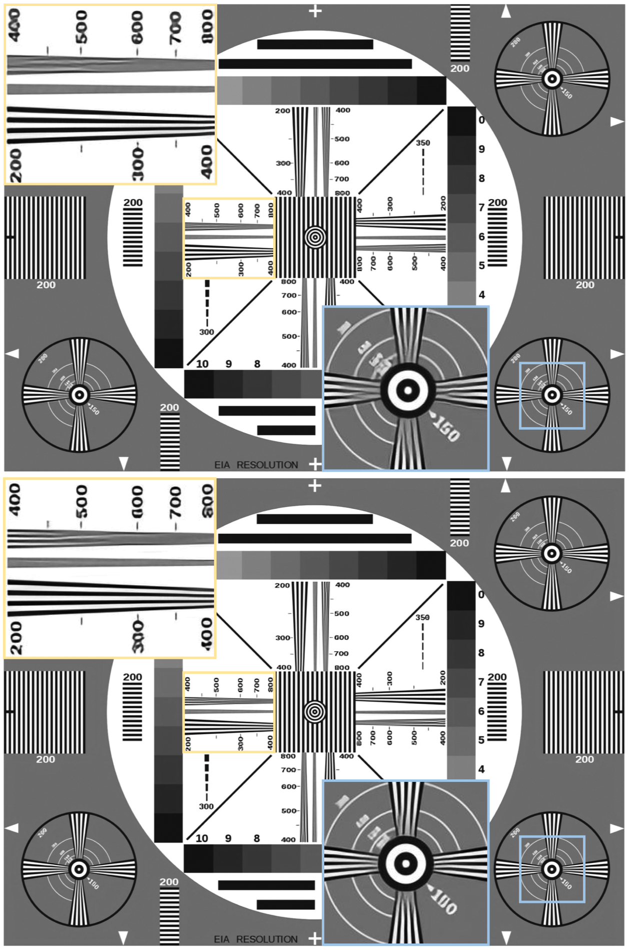

Fig. 8. Imaging performance of the best Gaussian PSF (top) and our end-to-end learned PSF (bottom). Our end-to end learned approach does show significantly better preservation of details above the Nyquist limit (see insets).

in our learned PSFs. Here, we remind the reader that MTFs are intended to characterize linear systems and may not be the best metric of assessing non-linear computational imaging systems such as ours.

5.3 Intensity Imaging

Experimental setup. The prototype of normal intensity imaging is illustrated in Figure 10. We use an MPD-SPC3 SPAD array as the detector. The phase mask is optimized for imaging daily scenes and human activity. The SPAD array is operated in snapshot mode with the integration time set as 52 μs. We sum up 100 frames before read-out, corresponding to a total integration time of around 5.2 ms.

Results of intensity imaging. To validate the practicability of the proposed optically coded single-shot super-resolution design, we employ the fabricated phase mask on a normal imaging setup that acts as the basis of alternative applications; for example, depth and transient imaging, as well as low-light imaging. A sequence of raw images (upsampled to the size of the reconstructed images for ease of comparison) is shown in Figure 11(1). The advantages of generating the optimal PSF specifically designed for the SPAD sensor's low fill-factor structure are significant. The reconstructed super-resolution results (i.e., Figure 11(2) faithfully preserves many details without introducing artifacts. Therefore, for such a kind of low fill-factor sensor structure, our method succeeds in preserving the spatial information.

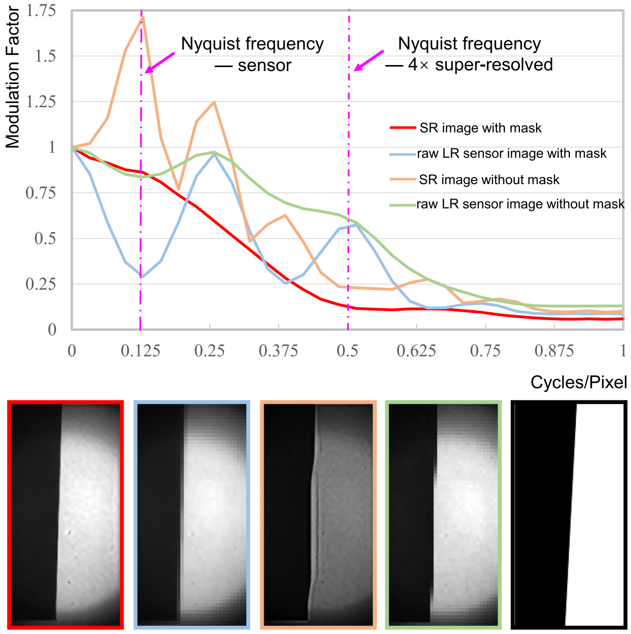

Fig. 9. MTFs derived from experimental results, including raw LR sensor image and 4× super-resolved SR image with and without phase mask, respectively. The corresponding images are revealed with different color plots, and the ideal 4× SR image is marked by black color.

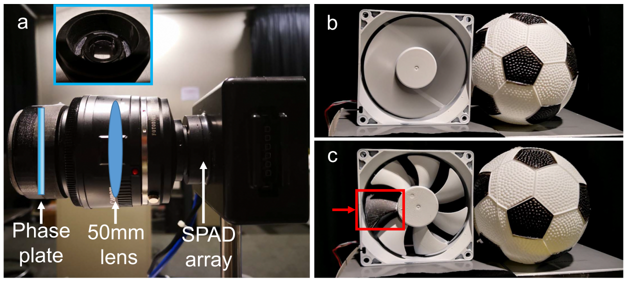

Fig. 10. Prototype for normal/high-speed imaging and the scene: (a) The prototype of normal imaging and high speed imaging. (b) The scene of running fans captured with a regular RGB sensor. (c) Static states of the scene shown in (b) and the red marked area are manually set as black to mark the rotating position.

Results of reference experiments. To further demonstrate that our phase mask works as designed, we performed a reference experiment for the same scenes without phase mask. Figure 11(3) presents the raw images without phase mask. The visualization of the raw images contains more high frequencies compared with those with phase mask. These undesirable high frequencies only

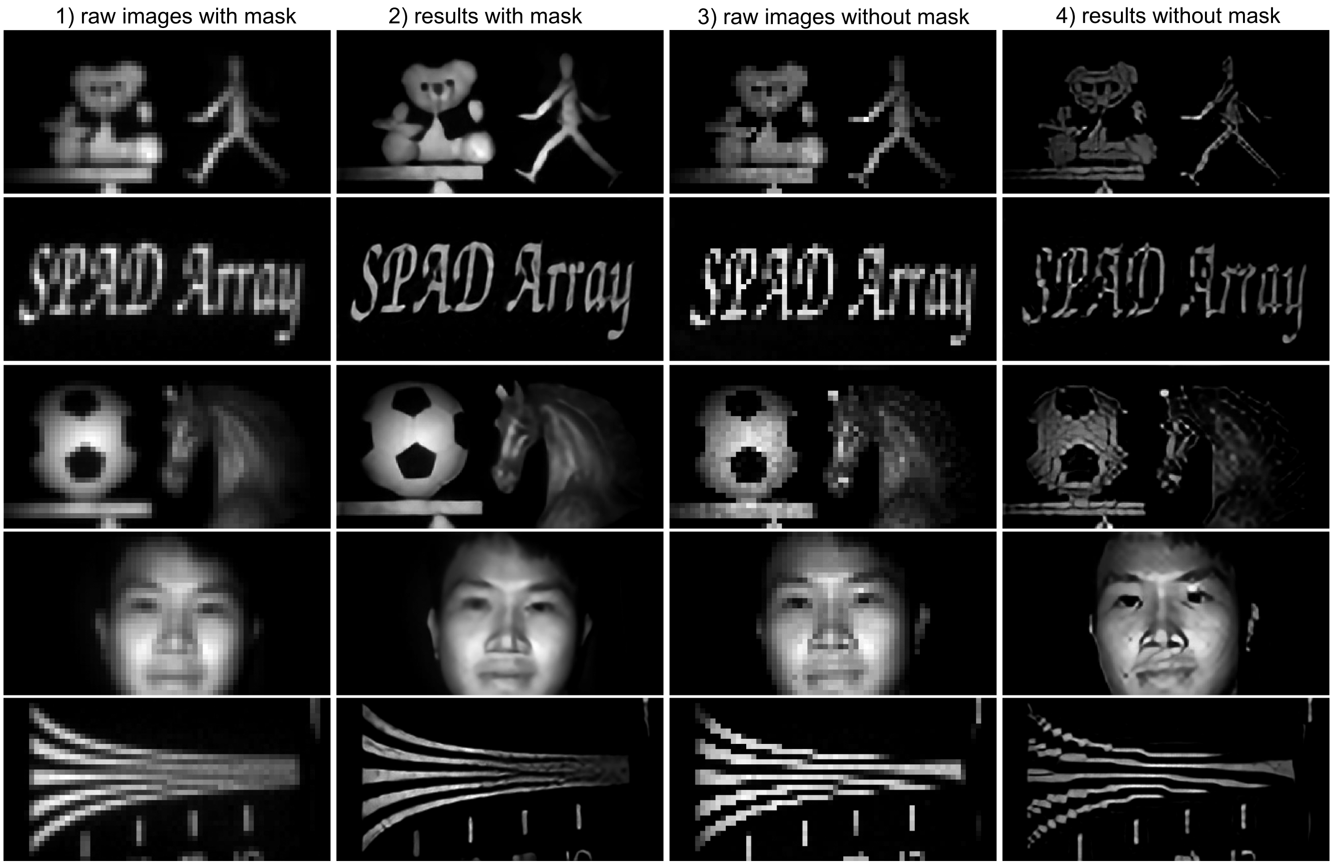

Fig. 11. Results of normal imaging. (1) Captured raw images with phase mask and with dark counts and background noise removed. (2) Results with phase mask. (3) Captured raw images without phase mask and with dark counts and background noise removed. (4) Results without phase mask.

result in the loss of fine details we want to preserve, but also introduce strong artifacts, as illustrated in Figure 11(4).

In comparison, our phase mask can preserve the most useful information while suppressing aliasing, consistent with the simulation results as described in Section 4.

5.4 High-speed Imaging

Experimental setup. We use the same camera setup described above. The SPAD array is operated in snapshot mode at a frame rate of 1,250 fps with the integration time set as 80 μs. In this example, we sum up 10 frames before read-out. As illustrated in Figure 10(b), we use a CPU fan as a high-speed spinning object. One of the blades is marked black as a position tracker, as shown in Figure 10(c).

Results of high-speed imaging. The optically coded single-shot super-solution camera fits well with unsynchronized and non-repeatable conditions, where time-sequential spatial resolution enhancement methods such as compressive sensing with a DMD, 2D mechanical scanning, or 1D line scanning are not applicable. As illustrated in Figure 12, we successfully capture and reconstruct the frames of a high-speed rotating fan (roughly calculated at 3,750 rpm from the shown frames). Figure 12(a) presents the captured raw data with darkcounts and background noise removed. Figure 12(b) presents the reconstructed 4× super-resolved frames. We can distinguish the fine details of fan and football. For more details, please refer to the supplemental video.

5.5 Depth and Transient Imaging

Experimental setup. Figure 13 illustrates the experimental setup for depth and transient imaging and the corresponding scenes. We use a 655 nm picosecond laser (PicoQuant LDH P-650) with an average power of around 1 mW as the illumination source. The Full width at half maximum of the laser pulses is around 80 ps, and the repetition rate is 50 MHz. To illuminate the scene smoothly, we scatter the laser beam using a diffuser and use an 80 mm plano-convex lens to re-concentrate the overly scattered beam.

We operate the SPAD camera in TCSPC mode with a 200 ps gate width and a 20 ps phase shift per cycle. The integration time is set to 52 μs, and 1,500 frames are summed up before read-out. In total, the capture process lasts around 9.8 s.

During the capture, the SPAD array sends the synchronizing signal to trigger the laser driver and then counts the arrival photons with a fixed phase offset of the gate. After sufficient integration, the SPAD camera shifts the gate window (i.e., 20 ps delay) and captures another frame until covering all designed phase offsets.

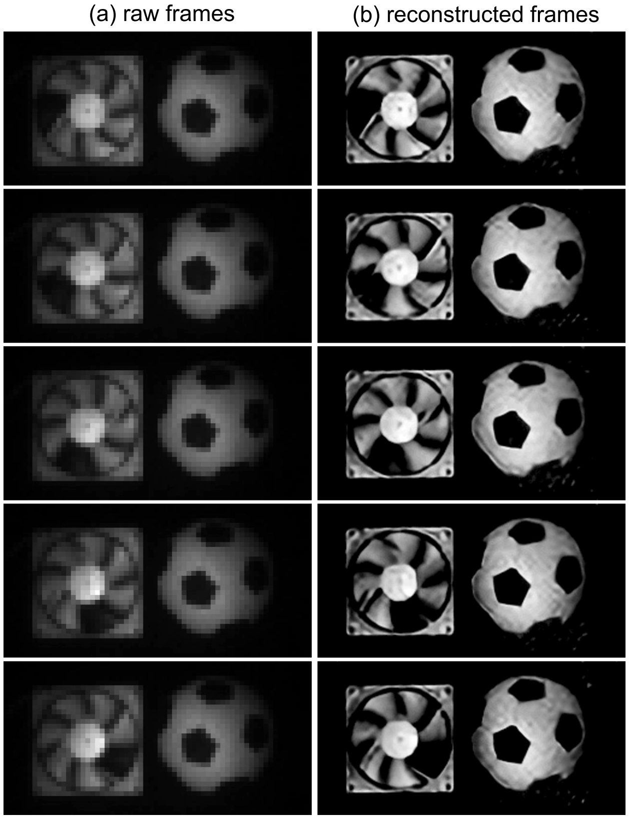

Fig. 12. Results of high-speed imaging. The displayed data are selected for every five frames, and we set the frame rates around 1,250 fps. (a) Selected raw frames with darkcounts and background noise removed. (b) Reconstructed high-resolution frames.

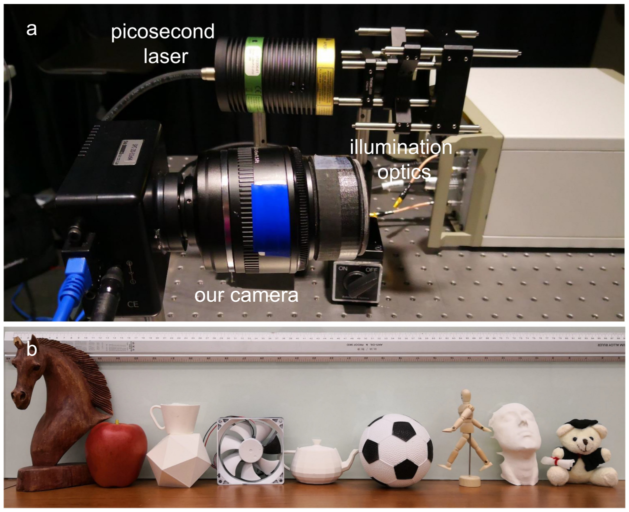

Fig. 13. Photograph of hardware setup of depth and transient imaging (a); and the scenes used in experiments (b).

Results of depth imaging. In this experiment, we demonstrate the ability to resolve the geometric details of several objects (fans, horse, wooden toys, etc.) in the scene depicted in Figure 14. As shown, the reconstructed intensity (Figure 14(b)) and depth images (Figure 14(c)) exhibit details that are hardly distinguishable in raw data; for instance, the edges of fans and wooden toys. From Figure 14(a), we observe that the raw images obtained by summing over the time axis remains very noisy, although the dark counts and background noise have been mostly removed. Compared to the raw data of intensity imaging, i.e., Figure 11(a), the summed pixel values show a considerably larger uncertainty, which makes it challenging to reconstruct good quality results. This is because the output power of our laser is very low with an average output only around 1 mW. Furthermore, the light is scattered to illuminate the entire scene. Consequently, only a few photons can be collected by our camera after bouncing back.

Results of transient imaging. Figure 15 presents the selected results of reconstructed transient frames. A mirror is placed near the objects to reflect the light. In Figure 15(a), the light pulse starts hitting the objects, resulting in a gradual increase and then a gradual decrease of the illumination. Later, the reflected light from the objects propagates to the mirror. Similarly, the reflected image (left part) shows the same phenomenon as the objects that the illumination gradually increases and then gradually decreases. The results in Figure 15(b) show a similar process. Thus, we have successfully captured and reconstructed high-resolution transient phenomena from the low-resolution raw data. Please refer to the supplementary video for a better visualization.

6 DISCUSSION

Fabrication feasibility and generalization. Our optimized PSFs are relatively small, which means that the phase plate only needs to diffract the light slightly, which can be achieved with relatively large feature sizes (5 μm in our experiments). This easily fits within the fabrication capability of inexpensive mass-production methods like micro-imprinting. In practice, the assembling accuracy (rotation ±4°, displacement ±2 mm) shows a minimal impact on reconstruction results. It is viable to design systems where the phase plate can be easily switched by end-users—as simply as switching a regular lens—to maximize the performance for different application scenarios. We believe the proposed design paradigm can be generalized to alternative low fill-factor and low-resolution sensors, such as onboard pixel processing circuits [Donati et al. 2007], 3D cameras, fluorescent analyzers, thermal cameras, and so on.

Limitations for depth and transient imaging. We reasonably ignore the multipath effect at the stage of proof-of-concept, since current illumination region is constrained within a level of a few decimeters. But there are several limitations that affect the reconstruction quality of depth and transient imaging. On the one hand, the picosecond laser used in our experiments has a power of only 1 mW. On the other hand, current photon detection efficiency (PDE) is only 12% at the wavelength of 655 nm. These two essential hardware constraints, in tandem with the need of diffusing laser beam into a 2D space to illuminate the whole scene, result in a fact that only a few reflected photons can be

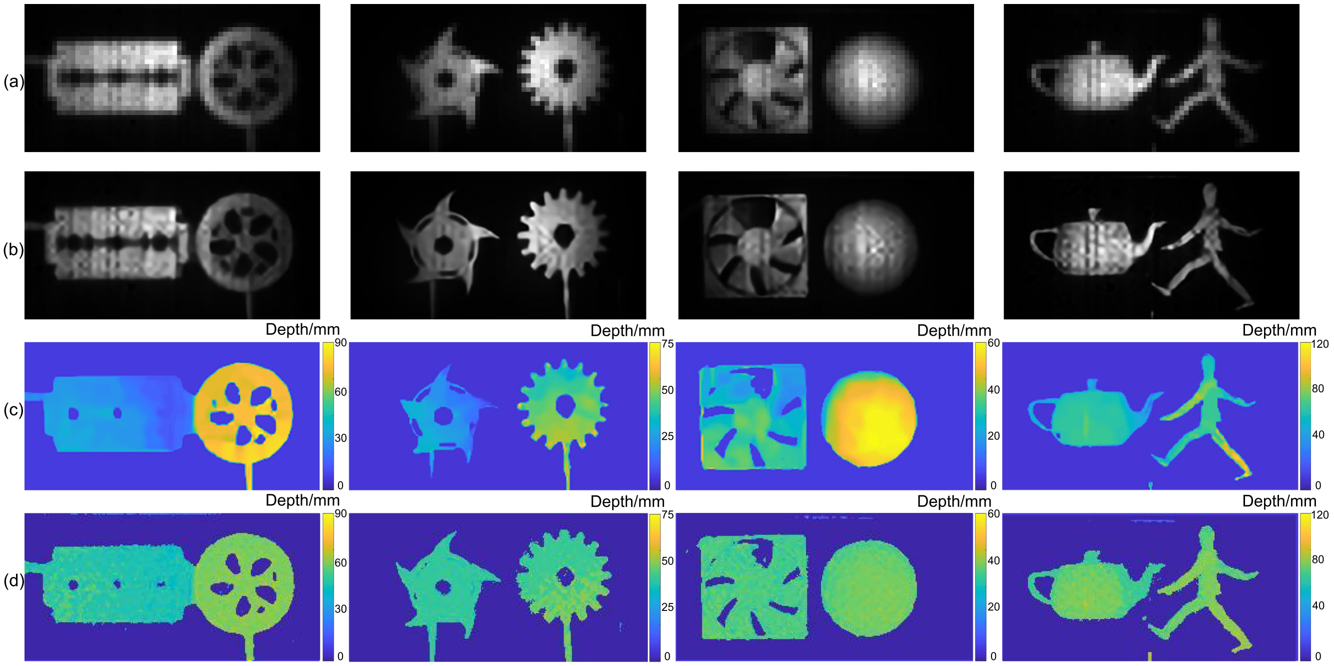

Fig. 14. Results of depth imaging: (a) raw image (with darkcounts and background noise removed) summed over the time dimension; (b) reconstructed intensity image according to (a); (c) reconstructed depth image; (d) reconstructed depth image without temporal deconvolution.

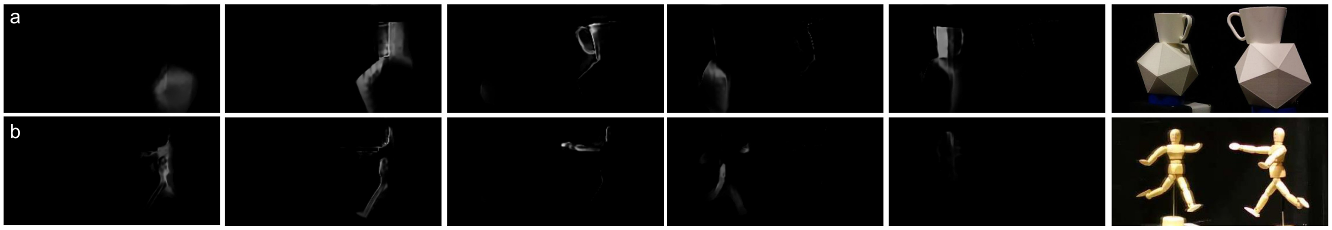

Fig. 15. Results of transient imaging. From left to right and from top to bottom are the selected frames of our reconstructed transient video stream. We here present one frame out of every nine frames with a visualized interval of 180 ps. The bottom right of (a) shows the captured scene containing a cup, a polyhedron, and a large mirror. The bottom right of (b) is the captured scene containing a wooden skeleton and a large mirror.

collected by the sensor. In contrast, line SPAD-based scanning methods [Lindell et al. 2018; O'Toole et al. 2017] scatter the laser beam only into a line and use the spectra of 450 nm, corresponding to a SPAD detection PDE around 50%. Therefore, currently the relatively lower light efficiency of our method adds difficulties to tackle the strong noise in the reconstruction.

Future Work. In-depth and transient imaging applications' increased illumination power always improves measurement range and robustness to ambient light. However, safety and cost concerns set tight limits to the laser power in many scenarios. To overcome this problem, using an intensity-modulated continuous laser, similar to amplitude modulated continuous wave (AMCW) time-of-flight sensors, can be a good alternative. A future direction of research would be to build a counting and digital version of AMCW TOF sensors using continuous wave illumination. This can be achieved by replacing the two capacitors that collect the charge of a photodiode with two counting units that count the photons of SPAD. In this way the SPAD-PMD device can lower the requirements on illumination while exhibiting more robustness to ambient light. SPAD arrays are a particularly promising technology for the field of fluorescent lifetime imaging, where state-of-the-art hardware solutions either suffer from low resolution or require complex and time-consuming mechanical scanning. To this end, optimizing a phase mask can enable a fast, high-resolution, and scanning-free fluorescent lifetime imaging system.

7 CONCLUSION

In conclusion, we present a general design paradigm to realize an optically coded single-shot super-resolution camera for low fill-factor sensors. This is achieved by incorporating optical design, sensor modeling, and deep network reconstruction. We build a high-resolution SPAD camera and demonstrate its viability in the application scenarios of intensity, high speed, and depth/transient imaging. Our approach for the first time overcomes the spatial resolution limit of existing SPAD sensor arrays with a single-shot capture, without the need of any mechanical scanning or repeatable measurement. The hardware improvement requires only a relatively inexpensive phase mask to the front focal plane of an existing optical system. We envision a wide range of applications across computer vision, sensing, and microscopic imaging.

ACKNOWLEDGEMENT

This work was fully supported by King Abdullah University of Science and Technology individual baseline funding and Visual Computing Center competitive funding.

REFERENCES

Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. 2011. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 5 (2011), 898–916.

Gonzalo R. Arce, David J. Brady, Lawrence Carin, Henry Arguello, and David S. Kittle. 2014. Compressive coded aperture spectral imaging: An introduction. *IEEE Sig. Proc. Mag.* 31, 1 (2014), 105–115.

Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. 2012. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *Proceedings of the British Machine Vision Conference*, Richard Bowden, John Collomosse, and Krystian Mikolajczyk (Eds.). BMVA Press, 135.1–135.10.

Peter D. Burns and Don Williams. 2002. Refined slanted-edge measurement for practical camera and scanner testing. In *Proceedings of the IS & T's PICS Conference.* Society for Imaging Science and Technology, 191–195.

Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. 2016. Dehazenet: An end-to-end system for single image haze removal. *IEEE Trans. Image Proc.* 25, 11 (2016), 5187–5198.

Ayan Chakrabarti. 2016. Learning sensor multiplexing design through back-propagation. In *Proceedings of the Conference on Advances in Neural Information Processing Systems.* 3081–3089.

Julie Chang, Vincent Sitzmann, Xiong Dun, Wolfgang Heidrich, and Gordon Wetzstein. 2018. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci. Rep.* 8, 12324 (2018).

Julie Chang and Gordon Wetzstein. 2019. Deep optics for monocular depth estimation and 3D object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'19).*

Huaijin Chen, M. Salman Asif, Aswin C. Sankaranarayanan, and Ashok Veeraraghavan. 2015. FPA-CS: Focal plane array-based compressive imaging in short-wave infrared. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15).* IEEE, 2358–2366.

Soma Datta, Nabendu Chaki, and Khalid Saeed. 2018. Minimizing aliasing effects using faster super resolution technique on text images. In *Transactions on Computational Science XXXI.* Springer, 136–153.

Silvano Donati, Giuseppe Martini, and Michele Norgia. 2007. Microconcentrators to recover fill-factor in image photodetectors with pixel on-board processing circuits. *Opt. Exp.* 15, 26 (2007), 18066–18075.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2016b. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 2 (2016), 295–307.

Chao Dong, Chen Change Loy, and Xiaoou Tang. 2016a. Accelerating the super-resolution convolutional neural network. In *Proceedings of the European Conference on Computer Vision.* Springer, 391–407.

Edward R. Dowski and W. Thomas Cathey. 1995. Extended depth of field through wave-front coding. *Appl. Opt.* 34, 11 (1995), 1859–1866.

Genevieve Gariepy, Nikola Krstajić, Robert Henderson, Chunyong Li, Robert R. Thomson, Gerald S. Buller, Barmak Heshmat, Ramesh Raskar, Jonathan Leach, and Daniele Faccio. 2015. Single-photon sensitive light-in-fight imaging. *Nat. Commun.* 6 (2015).

Nicholas George and Wanli Chi. 2003. Extended depth of field using a logarithmic asphere. *J. Opt. A: Pure Appl. Opt.* 5, 5 (2003), S157.

Ralph W. Gerchberg and W. O. Saxton. 1972. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* 35 (1972), 237.

Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian D. Reid, Chunhua Shen, Anton Van Den Hengel, and Qinfeng Shi. 2017. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17)*, Vol. 1. IEEE, 5.

Joseph W. Goodman. 2005. *Introduction to Fourier Optics.* Roberts and Company Publishers.

Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. 2018. Deep back-projection networks for super-resolution. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 1664–1673. DOI: 10.1109/CVPR.2018.00179

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16).* IEEE, 770–778.

Felix Heide, Steven Diamond, David B. Lindell, and Gordon Wetzstein. 2018. Sub-picosecond photon-efficient 3D imaging using single-photon sensors. *Scientific Reports* 8, 17726 (2018).

Felix Heide, Matthew O'Toole, Kai Zang, David Lindell, Steven Diamond, and Gordon Wetzstein. 2019. Non-line-of-sight imaging with partial occluders and surface normals. *ACM Trans. Graph.* 38, 3 (2019), 22:1–22:10. DOI: 10.1145/3269977

Felix Heide, Qiang Fu, Yifan Peng, and Wolfgang Heidrich. 2016. Encoded diffractive optics for full-spectrum computational imaging. *Sci. Rep.* 6 (2016).

Felix Heide, Lei Xiao, Andreas Kolb, Matthias B. Hullin, and Wolfgang Heidrich. 2014. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Opt. Exp.* 22, 21 (2014), 26338–26350.

Michael Iliadis, Leonidas Spinoulas, and Aggelos K. Katsaggelos. 2020. Deepbinarymask: Learning a binary mask for video compressive sensing. *Digital Signal Processing* 96 (2020), 102591.

Giuseppe Intermite, Aongus McCarthy, Ryan E. Warburton, Ximing Ren, Federica Villa, Rudi Lussana, Andrew J. Waddie, Mohammad R. Taghizadeh, Alberto Tosi, Franco Zappa, et al. 2015. Fill-factor improvement of Si CMOS single-photon avalanche diode detector arrays by integration of diffractive microlens arrays. *Opt. Exp.* 23, 26 (2015), 33777–33791.

Ganghun Kim, José A. Domínguez-Caballero, and Rajesh Menon. 2012. Design and analysis of multi-wavelength diffractive optics. *Opt. Exp.* 20, 3 (2012), 2814–2823.

Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. 2016a. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16).* IEEE, 1646–1654.

Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. 2016b. Deeply recursive convolutional network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16).* IEEE, 1637–1645.

Ahmed Kirmani, Dheera Venkatraman, Dongeek Shin, Andrea Colaço, Franco N. C. Wong, Jeffrey H. Shapiro, and Vivek K. Goyal. 2014. First-photon imaging. *Science* 343, 6166 (2014), 58–61.

Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. 2017. Deep Laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17).* IEEE, 624–632.

Anat Levin, Rob Fergus, Frédo Durand, and William T. Freeman. 2007. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.* 26, 3 (2007), 70.

Anat Levin, Samuel W. Hasinoff, Paul Green, Frédo Durand, and William T. Freeman. 2009. 4D frequency analysis of computational cameras for depth of field extension. *ACM Trans. Graph.* 28, 3 (July 2009). ACM, 97.

David Day-Uei Li, Simon Ameer-Beg, Jochen Arlt, David Tyndall, Richard Walker, Daniel R. Matthews, Viput Visitkul, Justin Richardson, and Robert K. Henderson. 2012. Time-domain fluorescence lifetime imaging techniques suitable for solid-state imaging sensor arrays. *Sensors* 12, 5 (2012), 5650–5669.

Day-Uei Li, Jochen Arlt, Justin Richardson, Richard Walker, Alex Buts, David Stoppa, Edoardo Charbon, and Robert Henderson. 2010. Real-time fluorescence lifetime imaging system with a 32× 32 0.13 μm CMOS low dark-count single-photon avalanche diode array. *Opt. Exp.* 18, 10 (2010), 10257–10269.

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 2017. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17) Workshops*, Vol. 1. IEEE, 3.

David B. Lindell, Matthew O'Toole, and Gordon Wetzstein. 2018. Single-photon 3D imaging with deep sensor fusion. *ACM Trans. Graph.* 37, 4 (July 2018), 1–12.

David B. Lindell, Gordon Wetzstein, and Matthew O'Toole. 2019. Wave-based non-line-of-sight Imaging using fast f–k migration. *ACM Trans. Graph.* 38, 4 (2019), 116. DOI: 10.1145/3306346.3322937

Patrick Llull, Xuejun Liao, Xin Yuan, Jianbo Yang, David Kittle, Lawrence Carin, Guillermo Sapiro, and David J. Brady. 2013. Coded aperture compressive temporal imaging. *Opt. Exp.* 21, 9 (2013), 10526–10545.

Roummel F. Marcia, Zachary T. Harmany, and Rebecca M. Willett. 2009. Compressive coded aperture imaging. In *Computational Imaging VII*, Vol. 7246. International Society for Optics and Photonics, 72460G.

Brian Morgan, Christopher M. Waits, John Krizmanic, and Reza Ghodssi. 2004. Development of a deep silicon phase Fresnel lens using gray-scale lithography and deep reactive ion etching. *J. Microelectromech. Syst.* 13, 1 (2004), 113–120.

Mythra Varun Nemallapudi, Stefan Gundacker, Paul Lecoq, Etiennette Auffray, Alessandro Ferri, Alberto Gola, and Claudio Piemonte. 2015. Sub-100 ps coincidence time resolution for positron emission tomography with LSO: Ce codoped with Ca. *Phys. Med. Biol.* 60, 12 (2015), 4635.

Desmond O'Connor. 2012. *Time-correlated Single Photon Counting.* Academic Press.

Matthew O'Toole, Felix Heide, David B. Lindell, Kai Zang, Steven Diamond, and Gordon Wetzstein. 2017. Reconstructing transient images from single-photon sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17)*. IEEE, 2289–2297.

Michael Parker. 2017. *Digital Signal Processing 101, Second Edition: Everything You Need to Know to Get Started*. Newnes, Newton, MA.

Sri Rama Prasanna Pavani, Michael A. Thompson, Julie S. Biteen, Samuel J. Lord, Na Liu, Robert J. Twieg, Rafael Piestun, and W. E. Moerner. 2009. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proc. Nat. Acad. Sci.* 106, 9 (2009), 2995–2999.

Juan Mata Pavia, Martin Wolf, and Edoardo Charbon. 2014. Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery. *Opt. Exp.* 22, 4 (2014), 4202–4213.

Adithya K. Pediredla, Aswin C. Sankaranarayanan, Mauro Buttafava, Alberto Tosi, and Ashok Veeraraghavan. 2018. Signal processing based pile-up compensation for gated single-photon avalanche diodes. *arXiv preprint arXiv:1806.07437* (2018).

Yifan Peng, Xiong Dun, Qilin Sun, Felix Heide, and Wolfgang Heidrich. 2018. Focal sweep imaging with multi-focal diffractive optics. In *Proceedings of the International Conference on Computational Photography (ICCP'18)*. IEEE, 1–8.

Yifan Peng, Qiang Fu, Felix Heide, and Wolfgang Heidrich. 2016. The diffractive achromat full spectrum computational imaging with diffractive optics. *ACM Trans. Graph.* 35, 4 (2016), 31.

Yifan Peng, Qilin Sun, Xiong Dun, Gordon Wetzstein, and Wolfgang Heidrich. 2019. Learned large field-of-view imaging with thin-plate optics. *ACM Trans. Graph.* 38, 6 (Nov. 2019), 1–14. ACM.

Maksim Petrov, Sergey Bibikov, Yuriy Yuzifovich, Roman Skidanov, and Artem Nikonorov. 2017. Color correction with 3D lookup tables in diffractive optical imaging systems. *Procedia Eng.* 201 (2017), 73–82.

Weidong Qu, Huarong Gu, Hao Zhang, and Qiaofeng Tan. 2015. Image magnification in lensless holographic projection using double-sampling Fresnel diffraction. *Appl. Opt.* 54, 34 (2015), 10018–10021.

Samuel Schulter, Christian Leistner, and Horst Bischof. 2015. Fast and accurate image upscaling with super-resolution forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*. IEEE, 3791–3799.

David Eric Schwartz, Edoardo Charbon, and Kenneth L. Shepard. 2008. A single-photon avalanche diode array for fluorescence lifetime imaging microscopy. *IEEE J. Solid-state Circ.* 43, 11 (2008), 2546–2557.

Yoav Shechtman, Steffen J. Sahl, Adam S. Backer, and W. E. Moerner. 2014. Optimal point spread function design for 3D imaging. *Phys. Rev. Lett.* 113, 13 (2014), 133902.

Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16)*. IEEE, 1874–1883.

Dongeek Shin, Feihu Xu, Dheera Venkatraman, Rudi Lussana, Federica Villa, Franco Zappa, Vivek K. Goyal, Franco N. C. Wong, and Jeffrey H. Shapiro. 2016. Photon-efficient imaging with a single-photon camera. *Nat. Commun.* 7 (2016).

Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gorden Wetzstein. 2018. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Trans. Graph.* 37, 4 (July 2018), 1–13.

Shuochen Su, Felix Heide, Gordon Wetzstein, and Wolfgang Heidrich. 2018. Deep end-to-end time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*. IEEE, 6383–6392.

Qilin Sun, Xiong Dun, Yifan Peng, and Wolfgang Heidrich. 2018. Depth and transient imaging with compressive SPAD array cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*. IEEE, 273–282.

Radu Timofte, Vincent De Smet, and Luc Van Gool. 2014. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Proceedings of the Asian Conference on Computer Vision*. 111–126.

Arin Can Ulku, Claudio Bruschini, Ivan Michel Antolović, Yung Kuo, Rinat Ankri, Shimon Weiss, Xavier Michalet, and Edoardo Charbon. 2018. A 512× 512 SPAD image sensor with integrated gating for widefield FLIM. *IEEE J. Select. Topics Quant. Electron.* 25, 1 (2018), 1–12.

Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G. Bawendi, and Ramesh Raskar. 2012. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nat. Commun.* 3 (2012), 745.

Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. 2013. Femto-photography: Capturing and visualizing the propagation of light. *ACM Trans. Graph.* 32, 4 (2013), 44.

Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han, and Thomas Huang. 2015. Deep networks for image super-resolution with sparse prior. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'15)*. 370–378.

Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. 2019. PhaseCam3D—learning phase masks for passive single view depth estimation. In *Proceedings of the IEEE International Conference on Computational Photography (ICCP'19)*. IEEE, 1–8.

Lei Xiao, Felix Heide, Matthew O'Toole, Andreas Kolb, Matthias B. Hullin, Kyros Kutulakos, and Wolfgang Heidrich. 2015. Defocus deblurring and superresolution for time-of-flight depth cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*. IEEE, 2376–2384.

Chih-Yuan Yang and Ming-Hsuan Yang. 2013. Fast direct super-resolution by simple functions. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'13)*. 561–568.

Jianchao Yang, John Wright, Thomas Huang, and Yi Ma. 2008. Image super-resolution as sparse representation of raw image patches. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08)*. IEEE, 1–8.

Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma. 2010. Image super-resolution via sparse representation. *IEEE Trans. Image Proc.* 19, 11 (2010), 2861–2873.

Li-Hao Yeh and Laura Waller. 2016. 3D super-resolution optical fluctuation imaging (3D-SOFI) with speckle illumination. In *Computational Optical Sensing and Imaging*. Optical Society of America, CW5D–2.

Roman Zeyde, Michael Elad, and Matan Protter. 2010. On single image scale-up using sparse-representations. In *Proceedings of the International Conference on Curves and Surfaces*. Springer, 711–730.

Can Zhao, Aaron Carass, Blake E. Dewey, Jonghye Woo, Jiwon Oh, Peter A. Calabresi, Daniel S. Reich, Pascal Sati, Dzung L. Pham, and Jerry L. Prince. 2018. A deep learning based anti-aliasing self super-resolution algorithm for MRI. In *Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 100–108.

Changyin Zhou, Stephen Lin, and Shree K. Nayar. 2011. Coded aperture pairs for depth from defocus and defocus deblurring. *Int. J. Comput. Vis.* 93, 1 (2011), 53–72.

Received November 2018; revised September 2019; accepted November 2019