Robust and Guided Super-resolution for Single-Photon Depth Imaging via a Deep Network

Alice Ruget[1], Stephen McLaughlin[1], Robert K. Henderson[2], Istvan Gyongy[2], Abderrahim Halimi[1], Jonathan Leach[1,*]

[1]School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK

[2]School of Engineering, The University of Edinburgh, Edinburgh, EH9 3FF, UK

***Abstract*—The number of applications that use depth imaging is rapidly increasing, e.g. self-driving autonomous vehicles and auto-focus assist on smartphone cameras. Light detection and ranging (LiDAR) via single-photon sensitive detector (SPAD) arrays is an emerging technology that enables the acquisition of depth images at high frame rates. However, the spatial resolution of this technology is typically low in comparison to the intensity images recorded by conventional cameras. To increase the native resolution of depth images from a SPAD camera, we develop a deep network built to take advantage of the multiple features that can be extracted from a camera's histogram data. The network then uses the intensity images and multiple features extracted from down-sampled histograms to guide the up-sampling of the depth. Our network provides significant image resolution enhancement and image denoising across a wide range of signal-to-noise ratios and photon levels.**

LiDAR waveform, Guided Super-resolution, Deep network, Unet, robust reconstruction

I. Introduction

Light detection and ranging (LiDAR) is a leading technology for depth imaging. For example, LiDAR is one of the key systems for future connected and autonomous vehicles, and it is used in the latest smartphones and tablet to aid auto-focus assist and enhance virtual reality. Single-photon avalanche detector (SPAD) arrays are an emerging technology for depth estimation via LiDAR. These devices can capture depth and intensity information of a scene. To achieve this, a short laser pulse is used to illuminate a target, and the detector records the arrival time of photons reflected back by the scene with respect to a laser trigger. This data, known as time tagged data, can be used to generate a temporal histogram of counts with respect to the time of flights, where the peak in the histogram can be used to calculate the distance to the target.

Although SPAD arrays are becoming well established in LiDAR systems, there are several key challenges to overcome to fully exploit their potential. The single-photon sensitivity that the SPAD array provides, promises depth imaging at long ranges [1]–[3] and in degraded visual environments [4]–[6], but improving the performance in these scenarios can dramatically increase the use of the detectors in different applications. In addition, the native resolution is typically very low in comparison to conventional image sensors. Ultimately, it is desirable to operate the SPAD arrays at high frame rates, cover a large field-of-view at large distances, produce images at high resolutions, and perform well in a wide range of environmental conditions.

This work was supported by the EPSRC grants (EP/T00097X/1, EP/S001638/1 and EP/L016753/1); the UK Royal Academy of Engineering through the Research Fellowship Scheme RF/201718/17128; DSTL Dasa project DSTLX1000147844

Due to the nature of the challenges in single-photon depth imaging, computational post-processing techniques are known to be a very powerful method to improve the overall image quality, both in terms of signal-to-noise and resolution. Statistical methods reported in [7]–[9] take advantage of available prior information to do robust reconstruction in the low-photon regime. Machine learning approaches have also shown good performance for super-resolution applied to all kind of depth images [10] or specialized on single-photon data [11], [12].

This paper proposes and implements a machine learning network that performs robust depth estimation by exploiting multi-scale information. The algorithm is insensitive to the number of time bins of the raw data, and achieves guided up-sampling based on corresponding intensity. We design and apply the network so that it is suitable for the data provided by the Quantic 4x4 sensor [13]–[15]. After processing, the final resolution of our up-sampled depth images from the Quantic 4x4 sensor is increased by a factor of four, i.e., from 64x32 to 256x128 pixels.

This paper is organized as follows. Section II introduces the observation model and the classical depth estimation procedure. Section III introduces the proposed HistNet in details. Results and comparisons on simulated and real data are reported in Sections IV and V, respectively. Section VI presents our conclusions and future work.

II. Observation model

A. Observed Data

In this work, we develop a network to process the data of the Quantic 4x4 SPAD array sensor, which generates the histograms of counts on-chip and operates in a hybrid acquisition mode [14], [15]. The Quantic 4x4 camera alternates between two measurements modes at a temporal rate exceeding 1000 frames per second : a high-resolution (HR) intensity measurement with a spatial resolution of 256x128, and a low-resolution (LR) histogram of photon counts containing depth information at a resolution of 64x32x16 (16 being the number of time bins of each of the 64x32 histograms). When the camera operates in depth mode, the histogram gathers the photon counts of 4x4 pixels, hence a lower spatial resolution. In the following, we assume the high resolution intensities are measured at even frames (denoted by $\boldsymbol{h}_{2k} \in \mathscr{R}^M$) and histograms at odd frames (denoted by $\boldsymbol{h}_{2k+1}$). For each pixel (i, j), the number of photons hitting the detector at time bin t follows a Poisson distribution [8], [9] and is expressed as function of the intensity $r_{i,j,t,2k+1}$ and the depth $d_{i,j,t,2k+1}$ as follows

$$h_{i,j,t,2k+1} = \mathscr{P}\left[r_{i,j,t,2k+1} * \mathscr{G}(d_{i,j,t,2k+1}, \sigma) + b_{i,j,2k+1}\right], \tag{1}$$

with $b_{i,j,2k+1}$ the background level which is assumed constant for all time bins of a given pixel, $\mathscr{G}$ is the impulse response of the SPAD camera approximated by a Gaussian function $\mathscr{G}(m, \sigma)$ with average m and standard deviation $\sigma \approx 0.5$ histogram bins [14], [15], $\mathscr{P}$ stands for the Poisson distribution. For each super-pixel (I,J), the LR histogram $H_{I,J,t,2k+1}$ given by the camera is the sum of the histogram over four by four pixels in the spatial dimensions:

$$H_{I,J,t,2k+1} = \sum_{i=I-2}^{i=I+2} \sum_{j=J-2}^{i=J+2} h_{i,j,t,2k+1}. \tag{2}$$

The resulting histograms contain LR depth information which can be estimated as indicated in the next section.

B. Depth estimation

Assuming known background level and a Gaussian system impulse response, the maximum likelihood estimator (MLE) of the depth is obtained as the central mass of the received signal photon time-of-flights (assuming depths are far from the observation window edges). This estimator is approximated for each depth pixel (I, J) as (in the following, we omitted the frame index k for clarity)

$$d_{I,J} = \frac{\sum_{t=t_1}^{t_2} t * \max(0, H_{I,J,t} - b_{I,J})}{\sum_{t=t_1}^{t_2} \max(0, H_{I,J,t} - b_{I,J})}, \tag{3}$$

with $H_{I,J,t}$ being the photon counts acquired in pixel (I, J) for time bin $t \in [1, T]$, T =16 is the number of time bins of the histogram , $b_{I,J}$ is the background level of pixel (I, J) estimated as the median of each pixel, $t_1 = \max(1, d^{max} - 2\sigma)$, $t_2 = \min(T, d^{max} + 2\sigma)$ and d^{max} represents the location of the signal peak estimated as the location of the maximum of the histogram of counts. This estimator provides poor performance when imaging under extreme conditions due to high-background level or low illumination imaging (i.e., photon sparse regime). This highlights the need for an advanced algorithm to estimate robust and high resolution depth maps. This paper proposes a tailored deep neural network which estimate a HR depth map by exploiting the available high resolution intensity image (to guide spatial upsampling), and the multi-scale information of the observed histograms of counts (to improve robustness).

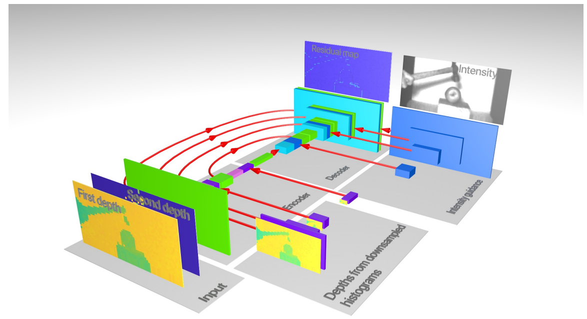

Fig. 1. **Representation of the HistNet.** Input to the network consists of the first and the second depth maps. Multi-resolution depths features are integrated along the contracting path of the U-Net. The intensity image is processed at multiple resolution and integrated along the expansive part of the U-Net. Skip connections between the contracting and the expansive paths are displayed as red arrows.

III. Guided super-resolution for single-photon LiDAR data

A. Network Architecture

This paper proposes a new learning-based algorithm, denoted HistNet, for depth denoising and up-sampling. We consider a U-net architecture [16] and incorporate guidance information as in [10]. Instead of processing the high dimensional observed histograms (which would be time and memory consuming), our solution is based on the extraction of informative features from the SPAD array data (LR histogram data and HR intensity image) and using these to retrieve the HR depth. Figures 1 and 2 show the considered architecture and input features. The network considers as input of an encoder the concatenation of a first depth map (III-A1) and a second depth map as detailed in following sections (III-A2). Multi-scale information of the histogram has been exploited in several state-of-the art 3D LiDAR denoising algorithms [7]–[12]. This information is incorporated to the encoder using a guidance branch containing multi-scale depth features (III-A3). In addition, the high resolution intensity image is exploited by considering a guidance branch in the decoder. The processing time to calculate each of these features is minimal, adding very little computational overhead to our overall procedure. The network produces a residual map $\mathscr{R}$ [10], [17] that can be added to the first depth map to render the final high-resolution depth map. These features are described in more details in the following sections.

1) First Depth Map: The estimated depth map in Section II-B is 4x up-scaled using a nearest neighbour interpolation.

2) Second Depth Map: When the Quantic 4x4 sensor operates in the depth mode, each super-pixel in the histogram gathers the photon counts of 4x4 pixels. Therefore, some histograms might present multiple peaks when observing multiple surfaces located at different depths. While the first depth map is calculated by identifying the strongest peak, we compute a second depth map based on the second strongest peak. We set the following criterion on the minimum number of photon

	type	*output shape*
Encoder *(25 108 992 parameters)*		
Input	cat(first depth, second depth)	256x128x2
L0	cv+cv	256x128x64
L1	mp+cat(LD1)+cv+cv	128x64x128
L2	mp+cat(LD2)+cv+cv	64x32x256
L3	mp+cat(LD3)+cv+cv	32x16x512
L4	mp+cat(LD4)+cv+cv	16x8x1024
Decoder *(31 058 368 parameters)*		
L5	dcv+cat(L4, I4)+cv+cv	32x16x512
L6	dcv+cat(L3, I3)+cv+cv	64x32x256
L7	dcv+cat(L2, I2)+cv+cv	128x64x128
L8	dcv+cat(L1, I1)+cv+cv	256x128x64
L9	cv	256x128

	type	*output shape*
Intensity Guidance*(1 549 824 parameters)*		
Input	Intensity Image	256x128
I1	cv	256x128x64
I2	mp + cv	128x64x128
I3	mp + cv	64x32x256
I4	mp + cv	32x16x512
Depth Guidance *(9600 parameters)*		
LD1	Input D1 + cv	128x64x64
LD2	Input D2 + cv	64x32x128
LD3	Input D3 + cv	32x16x256
LD4	Input D4 + cv	16x8x512

Fig. 2. **Details of network architecture and associated parameters**. The "output shapes" of the layers specified in the fourth column correspond to the case of processing our real data (i.e. histograms of spatial resolution 64x32 and 128x64 intensity images).cv stands for convolutional layer; mp for max-pooling layer; cat() for concatenation with the layers specified in the brackets and dcv for deconvolutional layer.

counts a relevant peak should contain. For each pixel (I, J), we consider a peak at bin t to be relevant if

$$H_{I,J,t} > b_{I,J} + \alpha\sqrt{b_{I,J}} \qquad (4)$$

with $H_{I,J,t}$ being the number photon counts of pixel (I, J) at time bin t; $b_{I,J}$ being the background level at pixel (I, J); $\sqrt{b_{I,J}}$ represents the standard-deviation of the Poisson distributed background counts; and α is a user adjusted parameter to ensure the second depth does not come from the noise but mostly represent real signal. We set $\alpha = 12$ in what follows, and the second depth is set to zero if this criterion is not met.

3) Multi-scale depth features: Multiple resolution scales have been shown to help depth estimation, in particular in high noise scenarios [8], [10], [15]. This information is included to the network by using four depth features D1, D2, D3 and D4 of different resolution scales. D1 is obtained by down-sampling the previously obtained 256×128 first depth map by two in both spatial dimensions using nearest neighbour interpolation. D2 is obtained by computing center of mass on the $64 \times 32 \times 16$ LR histogram. D3 and D4 are obtained by down-sampling this histogram by a factor of two and four by summing the neighbouring pixels in the spatial dimensions, hence obtaining histograms of size $32 \times 16 \times 16$ and $16 \times 8 \times 16$ respectively. Thanks to this process of down-sampling at the level of histogram, the resultant D3 and D4 have higher signal-to-noise than the first depth map, albeit with a lower resolution.

4) High resolution intensity: Akin to [10]–[12], this paper uses a HR intensity map to guide the reconstruction of the depth. The intensity image is directly obtained from the SPAD detector Quantic 4x4. This intensity image has a spatial resolution of 256×128, which is four times larger than the 64×32 histogram spatial resolution.

B. Implementation details

The considered loss function is based on the ℓ_1-norm that is known to promote sparsity [18], [19] allowing the reconstruction of separate depth surfaces

$$\mathscr{L}(\theta) = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left|\mathscr{R}_{m,n}(\theta) + d_{m,n} - d_{m,n}^{ref}\right|, \qquad (5)$$

with M the number of images within one batch, N the number of pixels of each image, θ the trainable parameters of the network, the $\mathscr{R}$ the residual map predicted by HistNet, d the first depth map, and d^{ref} the ground truth depth. During the training, a batch-mode learning method with a batch size of M = 64 was used. We implemented HistNet within the Tensorflow framework and use the ProximalAdagradOptimizer optimizer [20], as this enables the minimization of the ℓ_1-loss function. The learning rate was set to 1e-1. The training was performed on a NVIDIA RTX 6000 GPU for 2000 epochs.

C. Datasets

Using the observation model described in Section II-A, we simulate realistic SPAD array measurements (LR histogram and HR intensity) from 23 scenes of the MPI Sintel Depth Dataset [21], [22] for the training and validation dataset, and from scenes of the Middlebury dataset [23], [24] for the test dataset. We increase the number of training images by a factor of eight by using all possible combinations of 90° image rotations and flips. Furthermore, the images are split into overlapping patches of size 96x96 with a stride of 48.

IV. Results on synthetic data

A. Quantitative criteria

We consider two metrics to assess the level of noise affecting the data, the average number of signal photon counts per pixel (*ppp*), given by

$$ppp = \frac{1}{N}\sum_{i,j}\sum_{t=t_1}^{t_2}(h_{i,j,t} - b_{i,j}), \qquad (6)$$

and the signal to background ratio (SBR) defined as

$$SBR = \frac{1}{N}\sum_{i,j}\frac{\sum_{t=t_1}^{t_2}(h_{i,j,t} - b_{i,j})}{b_{i,j}\,(t_2 - t_1)}. \qquad (7)$$

The quality of the reconstruction is measured using the root mean squared error (RMSE) $= \sqrt{\frac{1}{N}\|\mathscr{R} + d - d^{ref}\|^2}$ and the absolute depth error (ADE) $= \left|\mathscr{R} + d - d^{ref}\right|$, $\mathscr{R}$ being the residual map predicted by HistNet after the training, d the up-scaled version of the low resolution depth map, i.e the first depth map, and d^{ref} the ground truth.

B. Comparison algorithms

We compare the results of HistNet with the following methods:

- *Nearest-neighbour interpolation* this is the first depth map (the input of the network)
- *Guided Image Filtering (GIF) of He et al. 2013* [25] the GIF is applied to the first depth map with the HR intensity image as a guide.
- *DepthSR-Net of Guo et al. 2019* [10] We retrained this network using the same training datasets as our network. It performs 4x upscaling guided by the HR intensity map
- *Algorithm of Gyongy et al. 2020* [15] designed to process the Quantic 4 x4 SPAD array and consists of various steps of guided filtering and up-sampling with low computational cost.

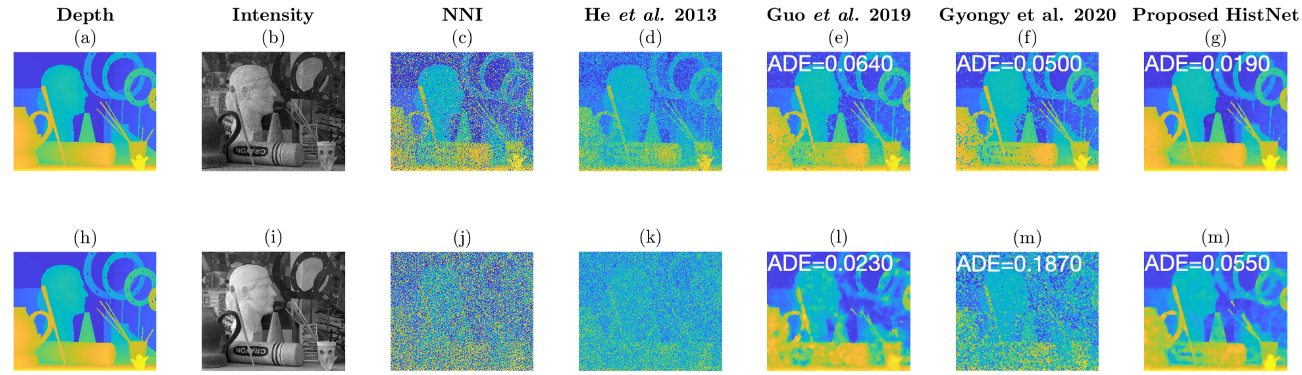

Fig. 3. **Comparison of reconstruction techniques for measurements simulated for the "medium signal-to-noise scenario" (a-g) and for the "low signal-to-noise scenario" (h-n).** ADE is the absolute depth error calculated with normalized data between 0 and 1. (a/h) displays the ground truth depth image; (b/i) displays the ground truth of the intensity. From left to right, the reconstructed depths using (c/j) the nearest neighbours interpolation; (d/k) the GIF [25]; (e/l) the DepthSR-Net algorithm [10]; (f/m) Gyongy's algorithm [15]; and (g/n) the proposed HistNet algorithm.

TABLE I
QUANTITATIVE COMPARISON OF THE DIFFERENT RECONSTRUCTION METHODS FOR 4X UP-SAMPLING ON SIMULATED MEASUREMENTS WITH A HIGH, MEDIUM AND LOW SIGNAL-TO-NOISE. RMSE IS THE ROOT-MEAN-SQUARE ERROR; ADE IS THE ABSOLUTE DEPTH ERROR.

	NNI		**GIF**		**Guo *et al.* [10]**		**Gyongy *et al.* [15]**		**Proposed HistNet**	
Rec time per scene	1ms		0.4s		7s (on GPU)		4s		7s (on GPU)	
Training on **high** signal-to-noise data with second depth; ppp=1200 counts and SBR=2										
Scene	RMSE	ADE	RMSE	ADE	RMSE	ADE	RMSE	ADE	RMSE	ADE
Art	0.053	0.038	0.046	0.039	0.026	0.0080	0.043	0.0076	**0.023**	**0.0027**
Reindeer	0.040	0.035	0.037	0.035	0.015	0.0051	0.023	0.0040	**0.012**	**0.0018**
Training on **medium** signal-to-noise data without second depth; ppp=4 counts and SBR=0.02										
Art	0.32	0.22	0.22	0.17	0.054	0.023	0.11	0.050	**0.053**	**0.019**
Reindeer	0.31	0.21	0.21	0.16	0.047	0.024	0.12	0.060	**0.040**	**0.019**
Training on **low** signal-to-noise data without second depth; ppp=4 counts and SBR=0.006										
Art	0.363	0.276	0.27	0.22	0.102	0.064	0.248	0.187	**0.082**	**0.055**
Reindeer	0.357	0.272	0.259	0.206	0.083	0.053	0.234	0.168	**0.075**	**0.050**

C. *Results*

We simulate realistic SPAD measurements from 1104x1376 HR scenes of the Middlebury dataset [23], [24]. Three different noise scenarios were considered: a scenario mimicking the lighting conditions of [15] with ppp = 1200 counts and SBR = 2 denoted "high signal-to-noise scenario", a scenario corresponding to a lower photon count and lower signal to noise with ppp = 4 counts and SBR = 0.02 denoted "medium signal-to-noise scenario", and a scenario corresponding to a lower photon count and much lower signal to noise with ppp = 4 counts and SBR = 0.006 denoted "low signal-to-noise scenario" . We trained a separate network for the three different noise scenarios. Quantitative comparison in Table I for two scenes show HistNet performs better in terms of the RMSE and ADE for all noise scenarios. Figure 3 shows that the proposed method provides best results for low and medium signal-to-noise. The processing time of the different methods is reported in Table I. The reconstructions of HistNet and DepthSR-Net were performed on a NVIDIA RTX 6000 GPU. Results on more validation data, a study on the robustness to different SBR and ppp levels, and results for 8x upsampling can be found in [26].

V. RESULTS ON MEASURED REAL DATA

We test the performance of HistNet on measurements captured by the Quantic 4x4 camera [15]. The spatial resolution of the histogram data is of 32x64 and the number of time bins is of 16. The resolution of the intensity image is of 128x256. This data shows ppp=1200 and SBR>2, and we used HistNet trained on the corresponding scenario. Figure 4 shows the reconstruction of Quantic 4x4 data via HistNet together with a comparison with different reconstruction algorithms. We see that HistNet leads to more accurate image with sharper edges.

VI. CONCLUSION

This paper presented a deep network for up-scaling and denoising of depth images obtained using a single-photon SPAD array detector. The network operates on informative multi-scale depth features to improve robustness and use the HR intensity image as a guide. The network showed best performance when compared to other algorithms using simulated and real data, under different noise scenarios. Future work will focus on the high frame rate of the SPAD array sensor and use information in the temporal domain to achieve better spatial resolutions for depth images. We also propose to tackle the misalignment between the histogram and the

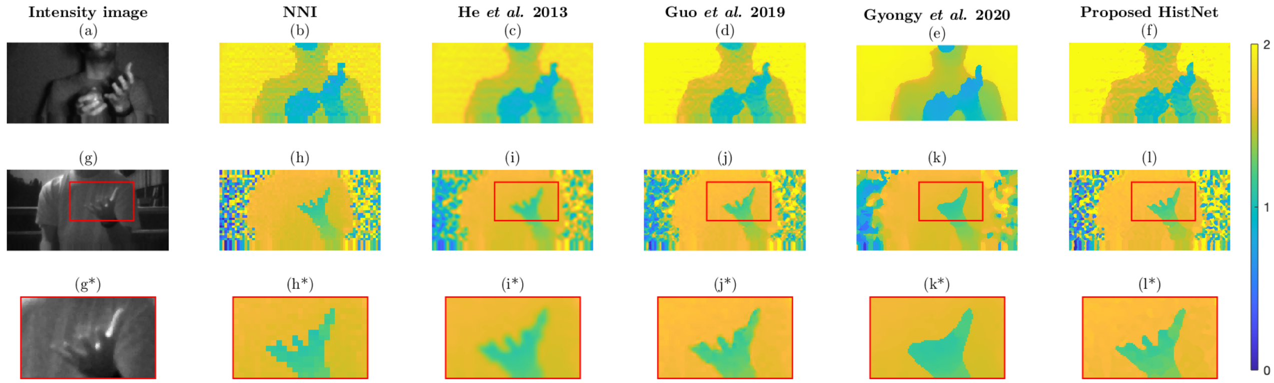

Fig. 4. **Reconstruction of Quantic 4x4 data.** (a-f) correspond to the indoors juggling scene. (g-l) correspond to the outdoors juggling scene. (g*-l*) correspond to closeup views of (g-l). (a/g) displays the reflectivity image from the SPAD [15]; From left to right, reconstructed depths using (b/h) NNI; (c/i) the GIF algorithm [25]; (d/j) the DepthSR-Net algorithm [10]; (e/k) Gyongy's algorithm in [15] ; and (f/l) the proposed HistNet algorithm.

intensity image, which is inherent to the operating mode of our SPAD detector that acquires them alternately.

Acknowledgements

The authors are grateful to STMicroelectronics and the ENIAC-POLIS project for chip fabrication for the Quantic 4x4 sensor. We are grateful to the authors of the DepthSRNet paper for sharing their code [10]. The code for this work can be found at https://github.com/HWQuantum/HistNet.

References

[1] S. Chan, A. Halimi, F. Zhu, I. Gyongy, R. K. Henderson, R. Bowman, S. McLaughlin, G. S. Buller, and J. Leach. Long-range depth imaging using a single-photon detector array and non-local data fusion. *Scientific Reports*, 9(1):8075, 2019.

[2] A. McCarthy, X. Ren, A. Della Frera, N. R. Gemmell, N. J. Krichel, C. Scarcella, A. Ruggeri, A. Tosi, and G. S. Buller. Kilometer-range depth imaging at 1550 nm wavelength using an InGaAs/InP single-photon avalanche diode detector. *Opt. Express*, 21(19):22098–22113, Sep 2013.

[3] A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller. Single-photon three-dimensional imaging at up to 10 kilometers range. *Opt. Express*, 25(10):11919–11931, May 2017.

[4] A. Maccarone, A. McCarthy, X. Ren, R. E. Warburton, A. M. Wallace, J. Moffat, Y. Petillot, and G. S. Buller. Underwater depth imaging using time-correlated single-photon counting. *Opt. Express*, 23(26):33911–33926, Dec 2015.

[5] A. Halimi, A. Maccarone, A. McCarthy, S. McLaughlin, and G. S. Buller. Object depth profile and reflectivity restoration from sparse single-photon data acquired in underwater environments. *IEEE Transactions on Computational Imaging*, 3(3):472–484, 2017.

[6] A. Wallace, A. Halimi, and G. Buller. Full waveform LiDAR for adverse weather conditions. *IEEE Transactions on Vehicular Technology*, PP:1–1, 04 2020.

[7] J. Tachella, Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, S. McLaughlin, and J. Tourneret. Bayesian 3D reconstruction of complex scenes from single-photon LiDAR data. *SIAM Journal on Imaging Sciences*, 12:521–550, 2019.

[8] J. Rapp and V. K Goyal. A few photons among many: Unmixing signal and noise for photon-efficient active imaging. *IEEE Transactions on Computational Imaging*, 3:445–459, 2017.

[9] A. Halimi, R. Tobin, A. McCarthy, J. M. Bioucas-Dias, S. McLaughlin, and G. S. Buller. Robust restoration of sparse multidimensional single-photon LiDAR images. *IEEE Transactions on Computational Imaging*, 6:138 – 152, July 2019.

[10] C. Guo, C. Li, Huazhu Fu J. Guo, and P. Han. Hierarchical features driven residual learning for depth map super-resolution. *IEEE Transactions on Image Processing*, 28(5):2545, MAY 2019.

[11] Z. Sun, D. B. Lindell, O. Solgaard, and G. Wetzstein. SPADnet: deep RGB-SPAD sensor fusion assisted by monocular depth estimation. *Opt. Express*, 28(10):14948–14962, May 2020.

[12] D. B. Lindell, M. O'Toole, and G. Wetzstein. Single-photon 3D imaging with deep sensor fusion. *ACM Trans. Graph.*, August 2018.

[13] S. W. Hutchings, N. Johnston, I. Gyongy, T. Al Abbas, N. A. W. Dutton, M. Tyler, S. Chan, J. Leach, and R. K. Henderson. A reconfigurable 3-d-stacked SPAD imager with in-pixel histogramming for flash LiDAR or high-speed time-of-flight imaging. *IEEE Journal of Solid-State Circuits*, 54(11):2947–2956, 2019.

[14] R. K. Henderson, N. Johnston, S. W. Hutchings, I. Gyongy, T. A. Abbas, N. Dutton, M. Tyler, S. Chan, and J. Leach. 5.7 a 256×256 40nm/90nm cmos 3D-stacked 120db dynamic-range reconfigurable time-resolved SPAD imager. In *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, pages 106–108, 2019.

[15] I. Gyongy, Sam W. Hutchings, A. Halimi, M. Tyler, S. Chan, F. Zhu, S. McLaughlin, R. K. Henderson, and J. Leach. High-speed 3D sensing via hybrid-mode imaging and guided upsampling. *Optica*, 2020.

[16] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[17] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

[18] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[19] D. L. Donoho and B. F. Logan. Signal recovery and the large sieve. *SIAM Journal on Applied Mathematics*, 52(2):577–591, 1992.

[20] Y. Singer and J C Duchi. Efficient learning using forward-backward splitting. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 495–503. Curran Associates, Inc., 2009.

[21] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *European Conf. on Computer Vision (ECCV)*, Part IV, LNCS 7577, pages 611–625. Springer-Verlag, October 2012.

[22] J. Wulff, D. J. Butler, G. B. Stanley, and M. J. Black. Lessons and insights from creating a synthetic optical flow benchmark. In Andrea Fusiello, Vittorio Murino, and Rita Cucchiara, editors, *ECCV Workshop on Unsolved Problems in Optical Flow and Stereo Estimation*, Part II, LNCS 7584, pages 168–177. Springer-Verlag, October 2012.

[23] D. Scharstein and C. Pal. Learning conditional random fields for stereo. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, June 2007.

[24] H. Hirschmüller and D. Scharstein. Evaluation of cost functions for stereo matching. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, June 2007.

[25] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, 2013.

[26] A. Ruget, S. McLaughlin, R. K. Henderson, I. Gyongy, A. Halimi, and J. Leach. Robust super-resolution depth imaging via a multi-feature fusion deep network. *Opt. Express*, 29(8), Apr 2021.