



OPEN

## A denoising framework for 3D and 2D imaging techniques based on photon detection statistics

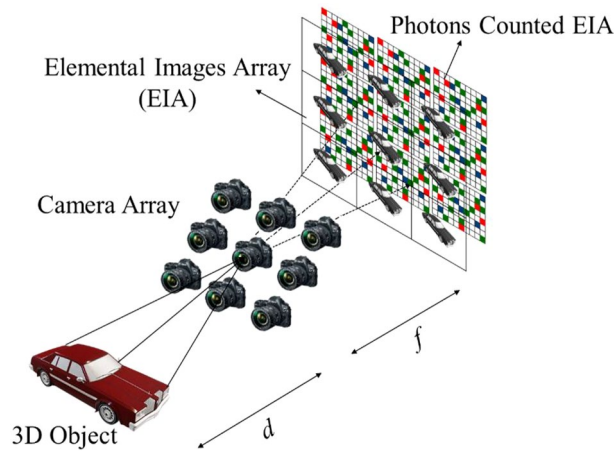
Vineela Chandra Dodda<sup>1,4</sup>, Lakshmi Kuruguntla<sup>1,4</sup>, Karthikeyan Elumalai<sup>1</sup>, Sunil Chinnadurai<sup>1</sup>, John T Sheridan<sup>2</sup> & Inbarasan Muniraj<sup>1,3</sup>✉

A method to capture three-dimensional (3D) objects image data under extremely low light level conditions, also known as Photon Counting Imaging (PCI), was reported. It is demonstrated that by combining a PCI system with computational integral imaging algorithms, a 3D scene reconstruction and recognition is possible. The resulting reconstructed 3D images often look degraded (due to the limited number of photons detected in a scene) and they, therefore, require the application of superior image restoration techniques to improve object recognition. Recently, Deep Learning (DL) frameworks have been shown to perform well when used for denoising processes. In this paper, for the first time, a fully unsupervised network (i.e., U-Net) is proposed to denoise the photon counted 3D sectional images. In conjunction with classical U-Net architecture, a skip block is used to extract meaningful patterns from the photons counted 3D images. The encoder and decoder blocks in the U-Net are connected with skip blocks in a symmetric manner. It is demonstrated that the proposed DL network performs better, in terms of peak signal-to-noise ratio, in comparison with the classical TV denoising algorithm.

Auto stereoscopic (i.e., glasses-free) 3D imaging and display techniques have numerous applications in numerous research fields e.g., biomedical, remote sensing, manufacturing, autonomous driving, and augmented reality (AR), to name a few<sup>1–3</sup>. Integral Imaging (II) is one of the techniques that captures a 3D object under incoherent light source and displays a reconstructed 3D scene which can be viewed without the use of special eye wear i.e., 3D glasses<sup>4,5</sup>. In principle, a multiple perspective of a 3D object must be recorded to reconstruct and display a 3D scene. For this purpose, various approaches have been demonstrated in the literature<sup>6</sup>. Owing to the simplified nature of the image capturing process, II has been widely applied<sup>1–3</sup>. To note, II was either combined with existing optical imaging systems or applied directly for auto-stereoscopic 3D imaging applications<sup>6</sup>. For instance, in<sup>7</sup>, an AR based navigation system was demonstrated for an *in-vivo* bio-imaging application. II system was combined with a conventional microscopy for a novel Light Field Microscope<sup>8</sup>. A method to synthesize a digital hologram using II dataset was demonstrated in<sup>9,10</sup>. Furthermore, imaging 3D objects under the turbid water was also proposed in<sup>11</sup>, to mention a few.

In addition to these, a method of photons detection under extremely dark conditions was combined with II systems, known as Photons Counted Integral Imaging (PCII), for low light 3D object imaging and visualization<sup>12</sup>. Thereafter, such system was examined for various applications such as biological imaging<sup>13</sup>, remote sensing<sup>14</sup>, night vision<sup>15</sup>, object detection<sup>16</sup>, autonomous driving<sup>17</sup> and data encryption<sup>18</sup>, to cite a few. We note that most of these systems were proposed to demonstrate the feasibility of capturing and displaying 3D images under low light. Therefore, these analyses were typically limited to the single channel or monochromatic imaging. Nevertheless, intuitively, colour perception of a 3D object in such a degraded environment should enable better scene interpretation. For this reason, we have developed a simplified single-channel based colour 3D imaging system<sup>19,20</sup>. Our system consists of a DSLR camera which translates both in horizontal (*x*) and vertical (*y*) directions to capture multiple two-dimensional (2D) images (often known as Elemental Images (EIs)) with different perspective. In one of our previous works<sup>19</sup>, we demonstrated that a 3D scene reconstruction is possible from just ~10 photons/scene. However, a 3D object (visual) recognition was possible only with >1000 photons/scene.

<sup>1</sup>Department of Electronics and Communication Engineering, School of Engineering and Applied Sciences, SRM University AP, Mangalagiri, Andhra Pradesh 522240, India. <sup>2</sup>School of Electrical and Electronic Engineering, College of Architecture and Engineering, University College Dublin, Belfield, Dublin 4, Ireland. <sup>3</sup>LiFE Laboratory, Department of Electronics and Communication Engineering, Alliance College of Engineering and Design, Alliance University, Bengaluru, Karnataka 562106, India. <sup>4</sup>These authors contributed equally: Vineela Chandra Dodda and Lakshmi Kuruguntla. ✉email: inbarasan.m@gmail.com



**Figure 1.** Pickup process of Photons Counted 3D Integral Imaging. Here, camera array depicts the rectangular translation of a single camera on the pickup plane to capture multiple 2D elemental images.

It is worth mentioning that the photons collection, in a given time interval by a photo sensor, is purely a random process. As a result of this spatial and temporal randomization, PCIs that are recorded, in general, represents a binary image i.e., either the presence (1) or absence (0) of photons. Accordingly, a 3D scene that are generated using the PCII dominated by impulse-like noises<sup>20</sup>. In such cases, it is ideal to use a denoising filter to remove the excessive noises that are prevalent in an image to preserve the scene as much as possible.

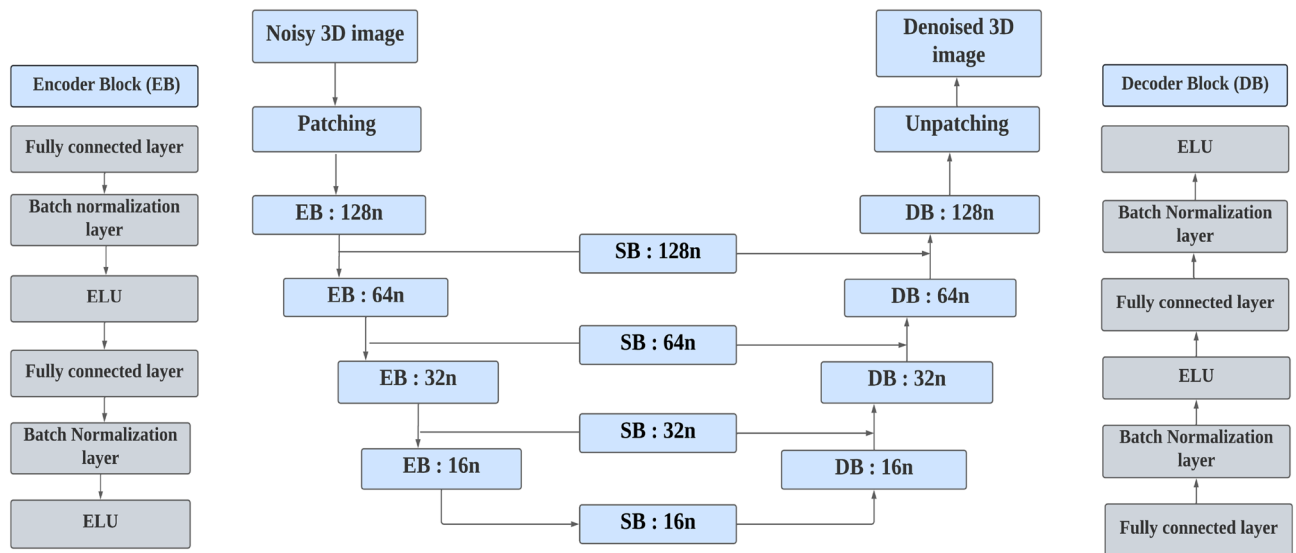
Since first proposed, Deep Learning (DL) frameworks<sup>21,22</sup> have received considerable attention across all disciplines and also among optical engineers and scientists. For instance, DL was applied for 3D object recognition and classification in very low light illumination conditions<sup>23</sup>. DL was also been shown to be suppressing noises that occur due to the misalignment of optics in diffraction tomography<sup>24</sup>. Similarly, DL network was proposed to remove the speckle noise from Digital Holographic (DH) imaging dataset. In<sup>25</sup>, authors developed a Denoising Convolutional Neural Network (DNCNN) to remove the speckle noise that occurs during phase measurements in DH imaging system. Whereas in<sup>26</sup>, multi-scale U-net architecture was used together with a customized cost function i.e., weighted combination of mean absolute error (MAE) and edge loss to minimize the noise from DH system. Further in<sup>27</sup>, an attention based CNN was proposed in which a customized cost function was developed using polarization loss to denoise the polarimetric images. In addition to this, in<sup>28</sup>, the authors proposed a combined KullBack Liebler (KL) divergence and Total Variation (TV) regularization as a cost function to enhance the denoising performance of the conventional CNN. Recently, a modified DNCNN was developed to enhance the quality of the reconstructed image of polarimetric 3D imaging under a degraded environment<sup>29</sup>. It is evident from these studies that supervised learning (SL) method was primarily used. It is known that, in general, SL requires a larger labeled clean data to train and test the network. However, such a clean labeled data may not be available and generating (synthetically) larger dataset is a time-consuming task. This process is applicable for PCII systems and therefore sufficiently larger (training) dataset is not available. To overcome such limitations, in this paper, for the first time, we propose to use a method which does not require a labeled dataset. Such a network is known as an end-to-end unsupervised network. In this work, we propose to use U-Net architecture with skip blocks<sup>30</sup> to denoise the photon counted 3D integral imaging dataset.

## Results

Figure 1 shows the Photons Counted Integral Imaging (PCII) setup. In principle, PCII can be implemented in two steps. *Step 1 (Pickup)*: In this process, a 3D object is being captured (in multiple different perspectives) by moving the camera both in vertical and horizontal directions. This process results in capturing four-dimensional (4D) light-field data i.e., two spatial dimensions ( $x, y$ ) and two angles in which light rays are measured ( $\theta_x, \theta_y$ )<sup>10</sup>. The captured images are known as Elemental Images (EIs) or an Elemental Image Array (EIA). *Step 2 (Reconstruction)*: The recorded 2D EIs are combined to produce a 3D scene. Reconstruction is in fact a reverse process of step 1, therefore by using a ray back propagation technique, a 3D scene can be reconstructed. It is worth to mention the fact that during reconstruction process the objects that were positioned at the same (object) plane combined flawlessly, thus they appear clearly in focus. The objects that were positioned in other planes appear blurry i.e., out of focus. A detailed Bayer patterned based EIA capturing and reconstructing can be found at<sup>19,20</sup>.

In some applications, image reconstruction is shown to be possible with only low scattered photons<sup>31</sup>. This can be done either by employing a physical camera, e.g. EMCCD and sCMOS, to capture a scene at low light conditions<sup>16,31</sup>, or by using the computational approaches<sup>12,32</sup>. In this work, we have used a computational approach, in such cases, a Poisson distribution is used to estimate the photons at any given image<sup>12,29</sup>. Let the total number of photons detected in a normalised elemental image ( $\hat{E}I$ ) is  $n_p$ , then by using a Poisson parameter  $\lambda$ , we can estimate the photons counted image as given below<sup>32</sup>:

$$P(C_x|I_x) \sim \text{PoissonDistribution}(\lambda = n_p \times \hat{E}I(x, y)) \quad (1)$$



**Figure 2.** Architecture of the unsupervised denoising network; EB-encoder block, DB-decoder block, and SB-skip block.

We then apply the parametric maximum likelihood estimator (MLE) to the photons counted elemental images to reconstruct the photons counted 3D sectional images<sup>19</sup>:

$$MLE\{I_p^Z\} = \frac{1}{n_p RS} \sum_{r=1}^R \sum_{s=1}^S C_{rs} \left( x + r \left( \frac{shf_x}{MF} \right), y + s \left( \frac{shf_x}{MF} \right) \right) \quad (2)$$

where  $MF$  denotes the magnification factor of the imaging system which is given as  $MF = z/d$  in which  $d$  is the distance between pick up grid and image plane, see Fig. 1. Subscripts  $r, s$  indicate the pick-up location of the elemental image and  $p$  denotes the photon counted images.  $C_{rs}(\cdot)$  is the photon counted pixel value of the  $(r, s)$ th elemental image, corresponding to the voxel value  $I_p^z$ <sup>19</sup>.

**Denoising network.** In this section, we describe the opted denoising deep learning architecture i.e., U-Net, see Fig. 2. As aforementioned, this is an end-to-end fully unsupervised denoising approach where the noisy photons counted 3D sectional image is fed as an input to the network. This network uses multiple encoder/decoder layers in a symmetric manner to retrieve denoised image with very few training data. Let  $x$  denote the clean 3D sectional image,  $n$  be the noise added by the photon counting process to the II system and  $y$  represent the resulted noisy photons counted 3D sectional images. Mathematically, this is given as:

$$y = x + n \quad (3)$$

The objective of image denoising problem is to restore  $x$  from  $y$  by attenuating the noise  $n$ . This process can be given as follows:

$$\hat{x} = \mathcal{H}(y; \Theta) \quad (4)$$

where  $\hat{x}$  be the estimated denoised photons counted image,  $\mathcal{H}(\cdot; \Theta)$  is a parametric function and  $\Theta$  are the trainable parameters. The major components in the U-Net are encoder and decoder blocks with skip connection layers<sup>33–35</sup>. In addition to this, skip blocks (SB) are added to the skip connection strategy in U-Net architecture to avoid vanishing gradients problem. These skip blocks and skip connections are designed in encoder and decoder blocks according to the features of the photons counted 3D images. Moreover, this strategy has the advantage of retaining useful image information. The intuitive reason behind adding the skip connections is that the low-level encoder extracts the abstract features which can be lost during the training process of the neural network. To avoid such loss in the features, we add skip connection from the encoder layers to the corresponding decoder layers. In the training process, 3D input image is given in the form of patches to the network. The advantage of applying such a patching technique is to increase the number of training samples such that the image features are learned by the network accurately. In the patching process, the patch window is moved horizontally and vertically to cover the whole image<sup>36</sup>. The patched input image is converted to 1D vector and fed as an input to the network. After removing the noises, we unpatch the 1D vector and convert back to the size of input data.

In the following, we will discuss the details of network blocks used: The encoder block consists of two fully connected layers, two batch normalization layers and two activation layers. The principle behind encoding operation is dimension-reduction thereby extracting the useful image content from the noisy image  $y$ . The encoding operation is expressed as follows:

$$d_{e1} = W_{e1}P_m + b_{e1} \quad (5)$$

where  $d_{e1}$  is the output from the encoder and  $P_m$  is the input to the encoder which is generated by the patching process.  $W_{e1}$  and  $b_{e1}$  are the weight and bias matrices of the  $m$ th encoding layer in the encoding process. Batch normalization layer is placed to prevent internal covariate shifts<sup>34</sup>. The low-level parameters in the network change the high-level data distributions during training process. This leads to the reduction of network accuracy due to the accumulation of error. However, the batch normalization layer speeds up the network and prevents gradient vanishing problem. The output of the batch normalization layer is passed to Exponential Linear Activation (ELU) function layer<sup>37</sup>, which is given as:

$$ELU(x) = \begin{cases} x & \text{if } x \geq 0 \\ a(e^x - 1) & \text{otherwise,} \end{cases} \quad (6)$$

where  $a$  is hyper-parameter and  $a \geq 0$ . The advantage of ELU is that it tries to make the mean activations (average activations of neurons in the layer for the given input) close to zero, thus speeding up the network. The output of each encoder block is given as:

$$\hat{d}_{e1} = ELU(W_{e1}P_m + b_{e1}) \quad (7)$$

The role of decoder is to reconstruct the photons counted 3D images from the abstract features extracted from the encoder block. The encoder and decoder blocks are symmetrical in structure. The decoder also consists of two dense layers, two batch normalization layers and two activation function layers. The output of decoder block is as follows:

$$\hat{r}_{dm} = ELU(W_{dm}K_m + b_{dm}) \quad (8)$$

where  $W_{dm}$  and  $b_{dm}$  be the weights and biases of the  $m$ th decoder layer and  $K_m$  is the input for the  $m$ th decoder layer. The number of neurons in the encoder block are 128, 64, 32, 16, 8 and vice-versa for the decoder block<sup>34</sup>. The last layer of the network is fully connected layer and reconstructs the output patches into the size of input photons counted image. In the training process, the selection of cost function plays a vital role to obtain optimum parameters i.e., weights and biases. To minimize the cost function, various optimization algorithms were proposed in the literature. For example, gradient descent, stochastic gradient descent, Adaptive Gradient Algorithm (ADAGRAD), Adaptive Moment Estimation (ADAM), to name a few<sup>35,38,39</sup>. In our work, we use the ADAM optimizer to update the parameters  $\Theta$ . The merits of ADAM include: easy implementation, computationally inexpensive and requires less memory<sup>35</sup>. The ADAM optimizer updates the parameters ( $\rho$ ) as shown below:

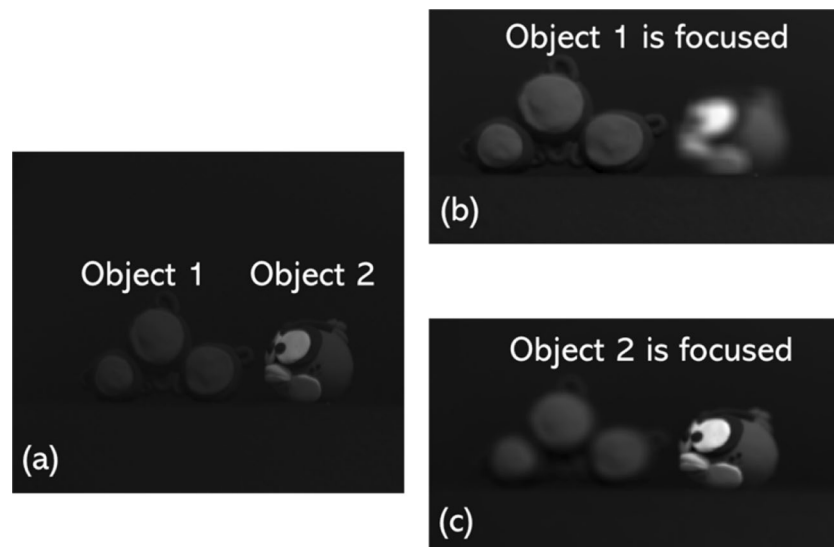
$$\rho_{(t+1)} = \rho_t - \frac{\eta}{\sqrt{\hat{v}(t)} + \epsilon} \hat{n}(t), \quad (9)$$

where  $\hat{v}(t)$  and  $\hat{n}(t)$  are bias corrected first and second moments defined as  $v_t/1 - \beta_2$  and  $n_t/1 - \beta_1$ , respectively. Terms  $n_t$  and  $v_t$  are exponentially moving averages obtained by using  $n_t = \beta_1 n_{t-1} + (1 - \beta_1)g_t$  and  $v_t = \beta_2 v_{t-1} + (1 - \beta_2)g_t^2$ , respectively. We note,  $\beta_1$  and  $\beta_2$  represents exponential decay rates for the first and second moments with the value of 0.90 and 0.999, respectively. The  $g_t$  is gradient of cost function with respect to time and  $\eta$  is learning rate which is generally set as 0.001<sup>35</sup>. The Mean Squared Error (MSE) is used as the cost function in our training process. The loss is calculated between the input noisy patched photons counted 3D images  $P$  and output denoised patches obtained from the network as follows:

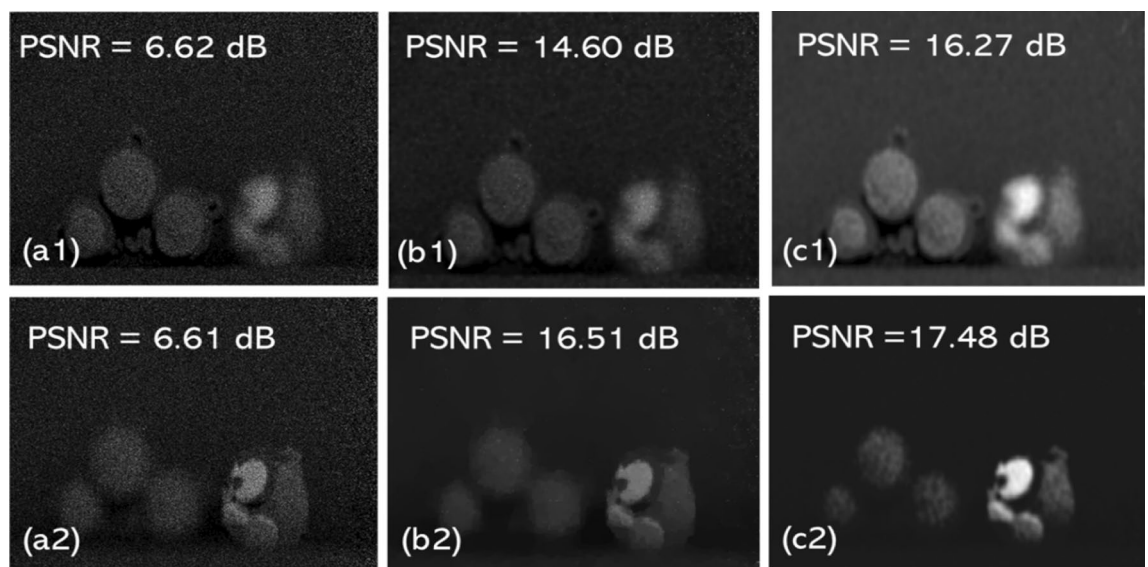
$$C(\Theta) = \min \| \psi(P; \Theta) - P \|^2 \quad (10)$$

where  $\psi$  denotes the denoising approach,  $\psi(P; \Theta)$  are the output patches. During the training process, we adapt an optimization strategy i.e., early stopping. When the cost function of the validation set does not decrease for four consecutive epochs, the denoising network will stop training and save the best parameters<sup>34</sup>. The SB block consists of one fully connected layer, one batch normalization layer, and one ELU layer. The output of  $m$ th decoder layer is connected to the output of SB block  $\hat{s}_{dm}$ . The final output after each decoder block is  $\hat{r}_{edl} = \{\hat{d}_{e1}, \hat{s}_{dm}\}$ .

**Experimental results.** To test the performance of our proposed denoising network, we used  $10 \times 10$  elemental images that were captured by shifting our CCD camera with equal separations of 5 mm in both horizontal (H) and vertical (V) directions (see Fig. 1). In our experiments, we used two 3D objects: one is a tri-colored ball known as Object 1 in Fig. 3a and second is a toy bird referred as Object 2 in Fig. 3a. From the pick up grid (i.e., imaging sensor) Object 1 and Object 2 were kept at 370 mm and 500 mm away, respectively. Further, the focal length of our imaging system is 50 mm and the pixel size is  $7.4\mu\text{m} \times 7.4\mu\text{m}$ . To note, the elemental images were initially recorded at the size of  $1668(H) \times 1668(V)$  but later resized into  $422(H) \times 539(V)$  before fed into our proposed DL network. Figure 3a depicts the two 3D objects used in our experiments and Fig. 3b, c shows the clean sectional images i.e., reconstructed 3D depth images via computational approach as described in<sup>19</sup> without using the photon counting technique. As can be seen from Fig. 3b, c in the reconstructed sectional images only one of the objects is clearly focused (depends on the corresponding depth location) while the other object is off-focused or defocused. Further, to note, the images shown in Fig. 3 were captured as Bayer patterned image (i.e., GRBG format) which has a potential to be converted as a colour (RGB) image using interpolation techniques. Conversion of the Bayer image into a RGB image is out of the scope of this article, therefore not described here. However, readers are recommended to explore<sup>19,20</sup> where a detailed 3D imaging setup, reconstruction technique



**Figure 3.** Three-dimensional (3D) objects used in our experiments: (a) EI of the 3D scene with Bayer format, (b) 3D reconstructed sectional image in which Tri-coloured balls is focused and (c) 3D reconstructed sectional image in which Toy bird is focused.

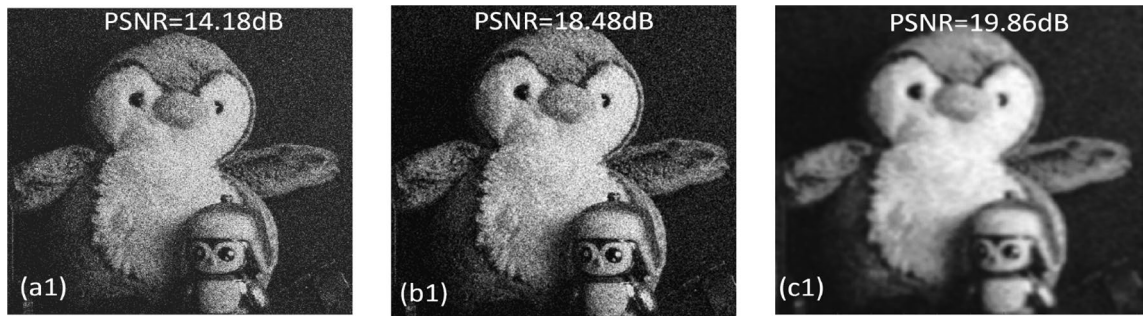


**Figure 4.** Denoised results: (a1, b1, c1) represents noisy Photon counted 3D sectional image, TV denoised image and result of our proposed denoising method when object 1 is in focus, respectively and (a2, b2, c2) represents the noisy Photon counted 3D sectional image, TV denoised image and result of our proposed denoising method when object 2 is in focus, respectively.

and a comparison of various interpolation techniques can be found (Fig. 4). In addition to this 3D dataset, we have also tested the proposed denoising network on a single-photon detector based 2D dataset i.e., Quanta Image sensor (QIS)<sup>40</sup>, for instance see Fig. 5. To reiterate, owing to the stochastic nature of the photons arrival to an imaging sensor, photon shot noise prevails at almost every standard imaging (i.e., CCD and CMOS) system. In such scenarios, denoising becomes a non-trivial task as the captured images are not only dark (see for instance, Figs. 3 and 5) but the noise is camouflaged with the recorded scene which makes it hard to distinguish from the object of interest. In the following, we present our denoising results.

The results are produced by performing simulations on an Intel® Xeon® Silver 4216 CPU @2.10 GHz (2 processors) with 256 GB RAM, 64-bit operating system. The software used is Spyder integrated development environment from Anaconda Navigator. It took around 68 s to run the python code and obtain the results. To note,  $n_p = 5000$  photons/scene is used to synthesize the photon counted 3D sectional image (PCSI), as described in Eq. (1). Our network is then trained using a single PCSI, which is fed into our DL network in the form of multiple patches. This patching process reduces the time required to create a dataset either by a physical or a synthetic process<sup>41</sup>. Thus, the demand for larger dataset is obviated as the network learns the features only





**Figure 5.** Denoised results: (a1), (b1), (c1) represents noisy QIS image, TV denoised image and result of our proposed denoising method with corresponding PSNR values.

through these patches. We note, our network was tested with various patch sizes to achieve better denoising accuracy. It was estimated, based on our simulations, an  $8 \times 8$  patch size provides a superior result in terms of peak signal-to-noise ratio (PSNR). It is worth to mention the fact that the selection of smaller patch size gives the finer details while the larger patch size may lose the finer details from an inputted image<sup>42</sup>. To note, we used 20% of the PCSI patches for validation and 60% of patches are allotted for training purpose. The validation loss (val\_loss) is continuously monitored, via the model checkpoint callback function, to estimate the optimum model i.e., the loss is minimized. In addition to this, we also used an early stopping criterion that stops the training process when the val\_loss is not converging even after at least 5 epochs, in the interest of computational time. In this work, 15 epochs were used with a learning rate of 0.001 to train the network. Further, we also calculated the computational complexity for the proposed denoising architecture and it is estimated the time taken for the classical TV denoising algorithm was 15.09018 s while it takes 2110 s (for both training and testing) to execute the proposed DL network.

In addition to this, to quantitatively evaluate the performance of the opted denoising DL network against the photon counted 3D sectional images and the 2D QIS dataset, we have used two standard image quality metrics: first is the Peak signal-to-noise ratio (PSNR) which is defined as follows:

$$PSNR = 10 \cdot \log_{10} \frac{I_{max}^2}{MSE} \quad (11)$$

where  $I_{max}$  is the maximum possible pixel intensity value of an image. MSE refers to the mean squared error between the clean image i.e., (Fig. 3b, or c) and the corresponding noisy PCSI (i.e., Fig. 4)<sup>19</sup>. For instance, the PSNR value given in Fig. 4b2 is an estimation between Figs. 3c and 4b2. The second metric, that we used in our simulations to test the performance of the opted denoising DL network, is structural similarity index measure (SSIM), which is given as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (12)$$

where  $\mu_x$ ,  $\mu_y$ ,  $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_{xy}$  represents mean, standard deviation, cross-covariance and  $C_1$  and  $C_2$  denotes the constant values, respectively<sup>43</sup>. For our proposed DL network, SSIM is calculated as 0.8540 when the Object 1 is in focus (i.e., Fig. 4a1), but when the same 3D sectional image was denoised using the classical TV denoising technique we obtained SSIM of 0.7218. Similarly, when the Object 2 is in focus (i.e., Fig. 4a2) the SSIM value of 0.6445 is achieved with the proposed DL method and 0.5913 with the classical TV denoising method. Furthermore, for the QIS dataset, we estimated SSIM of 0.3222 when our proposed DL method was applied and 0.3205 for the classical TV denoising method.

It is therefore evident from these results that the proposed DL method outperforms the classical TV denoising by maximum of 1.91 dB (for 3D dataset), 1.38 dB (for QIS dataset) in terms of PSNR and by maximum of 0.1322 (for 3D dataset) and 0.0017 (for QIS dataset) in terms of SSIM.

## Conclusion

In summary, we have proposed a deep learning network for denoising the 3D (photons counted three-dimensional integral imaging) and 2D (Quanta Image Sensor) dataset. We demonstrated that it is possible to denoise the low light level imaging dataset using a fully unsupervised network. In this work, encoder and decoder with skip blocks were opted to learn object features from the noisy photon counted 3D sectional images and QIS based images. The patches are selected randomly, covering the whole data, to obtain satisfactory results. As the denoising network does not require clean labels, the method is feasible for use in a wide variety of scenarios. It is therefore planned to extend this investigation by more closely identifying the patching process and parameter tuning in the architecture to achieve better denoised results<sup>42</sup>. This includes examining such network on some classical optical imaging systems that suffer from inevitable noises.

## Data availability

Data for this paper are not publicly available but shall be provided upon reasonable request to the corresponding author.

Received: 16 August 2022; Accepted: 9 January 2023

Published online: 24 January 2023

## References

1. Stern, A. & Javidi, B. Three-dimensional image sensing, visualization, and processing using integral imaging. *Proc. IEEE* **94**, 591–607 (2006).
2. Park, J.-H., Hong, K. & Lee, B. Recent progress in three-dimensional information processing based on integral imaging. *Appl. Opt.* **48**, H77–H94 (2009).
3. Xiao, X., Javidi, B., Martinez-Corral, M. & Stern, A. Advances in three-dimensional integral imaging: sensing, display, and applications. *Appl. Opt.* **52**, 546–560 (2013).
4. Lippmann, M. La photographie intégrale. *Compt. Rend.* **146**, 446–451 (1908).
5. Ives, H. E. Optical properties of a lippmann lenticulated sheet. *JOSA* **21**, 171–176 (1931).
6. Javidi, B. *et al.* Roadmap on 3d integral imaging: sensing, processing, and display. *Opt. Express* **28**, 32266–32293 (2020).
7. Liao, H., Iwahara, M., Hata, N. & Dohi, T. High-quality integral videography using a multiprojector. *Opt. Express* **12**, 1067–1076 (2004).
8. Levoy, M., Ng, R., Adams, A., Footer, M. & Horowitz, M. Light field microscopy. In *booktitleACM SIGGRAPH 2006 Papers*, 924–934 (2006).
9. Shaked, N. T., Rosen, J. & Stern, A. Integral holography: white-light single-shot hologram acquisition. *Opt. Express* **15**, 5754–5760 (2007).
10. Wan, M. *et al.* Orthographic projection images-based photon-counted integral fourier holography. *Appl. Opt.* **58**, 2656–2661 (2019).
11. Cho, M. & Javidi, B. Three-dimensional visualization of objects in turbid water using integral imaging. *J. Display Technol.* **6**, 544–547 (2010).
12. Tavakoli, B., Javidi, B. & Watson, E. Three dimensional visualization by photon counting computational integral imaging. *Opt. Express* **16**, 4426–4436 (2008).
13. Taguchi, K. & Iwaczyk, J. S. Vision 20/20: single photon counting x-ray detectors in medical imaging. *Med. Phys.* **40**, 100901 (2013).
14. McCarthy, A. *et al.* Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection. *Opt. Express* **21**, 8904–8915 (2013).
15. Laurenzis, M., Christnacher, F., Klein, J., Hullin, M. B. & Velten, A. Study of single photon counting for non-line-of-sight vision. In *booktitleAdvanced Photon Counting Techniques IX*, vol. **9492**, 31–38 (organizationSPIE, 2015).
16. Markman, A., Shen, X. & Javidi, B. Three-dimensional object visualization and detection in low light illumination using integral imaging. *Opt. Lett.* **42**, 3068–3071 (2017).
17. Rapp, J., Tachella, J., Altmann, Y., McLaughlin, S. & Goyal, V. K. Advances in single-photon lidar for autonomous vehicles: working principles, challenges, and recent advances. *IEEE Signal Process. Mag.* **37**, 62–71 (2020).
18. Muniraj, I. *et al.* Low photon count based digital holography for quadratic phase cryptography. *Opt. Lett.* **42**, 2774–2777 (2017).
19. Moon, I., Muniraj, I. & Javidi, B. 3d visualization at low light levels using multispectral photon counting integral imaging. *J. Display Technol.* **9**, 51–55 (2013).
20. Muniraj, I., Guo, C., Lee, B.-G. & Sheridan, J. T. Interferometry based multispectral photon-limited 2d and 3d integral image encryption employing the Hartley transform. *Opt. Express* **23**, 15907–15920 (2015).
21. Vincent, P. *et al.* Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **11** (2010).
22. Altmann, Y., Aspden, R., Padgett, M. & McLaughlin, S. A Bayesian approach to denoising of single-photon binary images. *IEEE Trans. Comput. Imaging* **3**, 460–471 (2017).
23. Markman, A. & Javidi, B. Learning in the dark: 3d integral imaging object recognition in very low illumination conditions using convolutional neural networks. *OSA Continuum* **1**, 373–383 (2018).
24. Choi, G. *et al.* Cycle-consistent deep learning approach to coherent noise reduction in optical diffraction tomography. *Opt. Express* **27**, 4927–4943 (2019).
25. Montresor, S., Tahon, M., Laurent, A. & Picart, P. Computational de-noising based on deep learning for phase data in digital holographic interferometry. *APL Photonics* **5**, 030802 (2020).
26. Jeon, W., Jeong, W., Son, K. & Yang, H. Speckle noise reduction for digital holographic images using multi-scale convolutional neural networks. *Opt. Lett.* **43**, 4240–4243 (2018).
27. Liu, H., Zhang, Y., Cheng, Z., Zhai, J. & Hu, H. Attention-based neural network for polarimetric image denoising. *Opt. Lett.* **47**, 2726–2729 (2022).
28. Lindell, D. B., O’Toole, M. & Wetzstein, G. Single-photon 3d imaging with deep sensor fusion. *ACM Trans. Graph.* **37**, 113–1 (2018).
29. Usmani, K., O’Connor, T. & Javidi, B. Three-dimensional polarimetric image restoration in low light with deep residual learning and integral imaging. *Opt. Express* **29**, 29505–29517 (2021).
30. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, 234–241 (Springer, 2015).
31. Velten, A. *et al.* Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nat. Commun.* **3**, 1–8 (2012).
32. Goodman, J. W. *Statistical Optics* (Wiley, 2015).
33. Lempitsky, V., Vedaldi, A. & Ulyanov, D. Deep image prior. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9446–9454 (IEEE, 2018).
34. Yang, L. *et al.* Unsupervised 3-D random noise attenuation using deep skip autoencoder. In *IEEE Transactions on Geoscience and Remote Sensing* (2021).
35. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
36. Chen, Y., Zhang, M., Bai, M. & Chen, W. Improving the signal-to-noise ratio of seismological datasets by unsupervised machine learning. *Seismol. Res. Lett.* **90**, 1552–1564 (2019).
37. Clevert, D.-A., Unterthiner, T. & Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint [arXiv:1511.07289](https://arxiv.org/abs/1511.07289) (2015).
38. Sun, S., Cao, Z., Zhu, H. & Zhao, J. A survey of optimization methods from a machine learning perspective. *IEEE Trans. Cybern.* **50**, 3668–3681 (2019).
39. Saad, O. M. & Chen, Y. Deep denoising autoencoder for seismic random noise attenuation. *Geophysics* **85**, V367–V376 (2020).
40. Sanghvi, Yash, Gnanasambandam, Abhiram & Chan, Stanley H. Photon limited non-blind deblurring using algorithm unrolling. *IEEE Trans. Comput. Imaging* **8**, 851–864 (2022).
41. Alkinani, M. H. & El-Sakka, M. R. Patch-based models and algorithms for image denoising: a comparative review between patch-based images denoising methods for additive noise reduction. *EURASIP J. Image Video Process.* **2017**, 1–27 (2017).

42. Kuruguntla, L., Dodda, V. C. & Elumalai, K. Study of parameters in dictionary learning method for seismic denoising. *IEEE Trans. Geosci. Remote Sens.* (2021).
43. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).

### Acknowledgements

VC, LK, KE, and SC acknowledges the support of SRM University AP research fund. IM acknowledges the financial support from Science and Engineering Research Board (SERB) under SRG/2021/001464 scheme, Department of Science and Technology, Government of India. JTS acknowledged the support of University College Dublin. Authors thank Prof. Stanley H Chan of Purdue University, USA for providing the QIS dataset.

### Author contributions

V.C., I.M. planned the project; L.K., S.C., K.E contributed in simulations; I.M., J.T.S. mentored the project; all authors discussed the results and contributed equally for the manuscript preparation.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to I.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023