



OPEN

## Deep learning approach for denoising low-SNR correlation plenoptic images

Francesco Scattarella<sup>1,2</sup>, Domenico Diacono<sup>2</sup>, Alfonso Monaco<sup>1,2</sup>✉, Nicola Amoroso<sup>2,3</sup>, Loredana Bellantuono<sup>2,4</sup>, Gianlorenzo Massaro<sup>1,2</sup>, Francesco V. Pepe<sup>1,2</sup>, Sabina Tangaro<sup>2,5</sup>, Roberto Bellotti<sup>1,2,6</sup> & Milena D'Angelo<sup>1,2,6</sup>

Correlation Plenoptic Imaging (CPI) is a novel volumetric imaging technique that uses two sensors and the spatio-temporal correlations of light to detect both the spatial distribution and the direction of light. This novel approach to plenoptic imaging enables refocusing and 3D imaging with significant enhancement of both resolution and depth of field. However, CPI is generally slower than conventional approaches due to the need to acquire sufficient statistics for measuring correlations with an acceptable signal-to-noise ratio (SNR). We address this issue by implementing a Deep Learning application to improve image quality with undersampled frame statistics. We employ a set of experimental images reconstructed by a standard CPI architecture, at three different sampling ratios, and use it to feed a CNN model pre-trained through the transfer learning paradigm U-Net architecture with VGG-19 net for the encoding part. We find that our model reaches a Structural Similarity (SSIM) index value close to 1 both for the test sample (SSIM =  $0.87 \pm 0.02$ ) and in 5-fold cross validation (SSIM =  $0.92 \pm 0.07$ ); the results are also shown to outperform classic denoising methods, in particular for images with lower SNR. The proposed work represents the first application of Artificial Intelligence in the field of CPI and demonstrates its high potential: speeding-up the acquisition by a factor 20 over the fastest CPI so far demonstrated, enabling recording potentially 200 volumetric images per second. The presented results open the way to scanning-free real-time volumetric imaging at video rate, which is expected to achieve a substantial influence in various applications scenarios, from monitoring neuronal activity to machine vision and security.

Correlation Plenoptic Imaging (CPI) is a recently established three-dimensional imaging modality that exploits the spatio-temporal correlations of light for enabling plenoptic imaging (PI) at the diffraction limit<sup>1–6</sup>. While in standard plenoptic imaging the required position and direction information are encoded in the intensity registered by a single sensor<sup>7–9</sup>, thus sacrificing image resolution, the volumetric information is retrieved in CPI by measuring spatio-temporal correlations between two disjoint sensors. As a result of the different physical mechanism regulating the two approaches, CPI considerably enlarges the maximum achievable depth of field, at a given resolution, with respect to conventional PI, and significantly improves the volumetric resolution<sup>5,10</sup>. Several alternative configurations of CPI have so far been proposed<sup>4,11–14</sup>, based on the correlation properties of either chaotic light<sup>1</sup> or entangled photons<sup>3,5</sup>. In all cases, the advantages connected with the use of the spatio-temporal correlation properties of light are counterbalanced by the main open challenge of correlation imaging: the low acquisition speed related with the need for collecting a statistically relevant quantity of samples (i.e., pairs of frames simultaneously acquired by the two sensors) to reconstruct the intensity correlation function. In general, the number of collected frames cannot be reduced too much without negatively affecting the image quality, i.e., its signal-to-noise ratio (SNR)<sup>15</sup>. This trade-off crucially affects temporal performance of CPI, thus limiting its range of effective applicability and its competitiveness with state-of-the-art volumetric imaging techniques, especially when dealing with moving objects. Recently, a first attempt was made by Massaro et al.<sup>16</sup> to operate CPI at a frame rate approaching video rate: here, correlated photon imaging was demonstrated at a rate of 10 *volumetric* images

<sup>1</sup>Dipartimento Interateneo di Fisica M. Merlin, Università degli Studi di Bari Aldo Moro, 70125 Bari, Italy. <sup>2</sup>Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Bari, 70125 Bari, Italy. <sup>3</sup>Dipartimento di Farmacia - Scienze del Farmaco, Università degli Studi di Bari Aldo Moro, 70125 Bari, Italy. <sup>4</sup>Dipartimento di Biomedicina Traslazionale e Neuroscienze (DiBrain), Università degli Studi di Bari Aldo Moro, 70124 Bari, Italy. <sup>5</sup>Dipartimento di Scienze del Suolo, della Pianta e degli Alimenti, Università degli Studi di Bari Aldo Moro, 70125 Bari, Italy. <sup>6</sup>These authors contributed equally: Roberto Bellotti and Milena D'Angelo. ✉email: alfonso.monaco@ba.infn.it

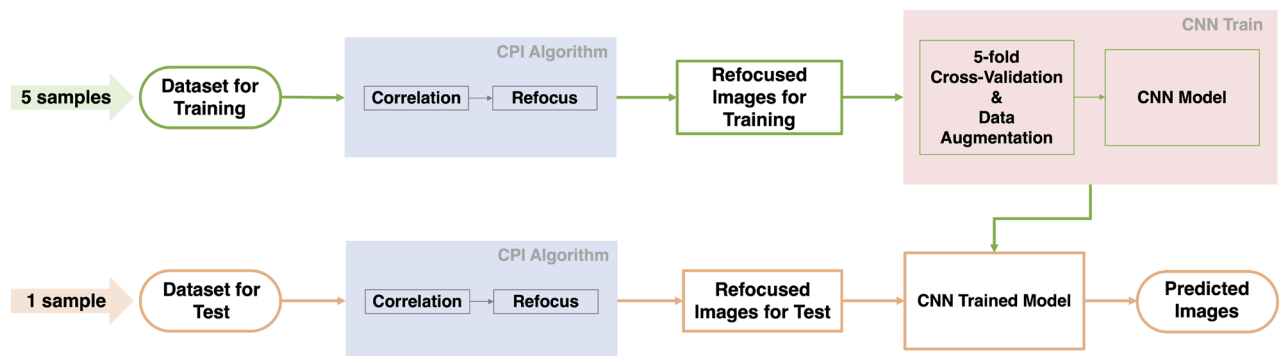
per second using SwissSPAD2, an array of single photon avalanche photodiodes (SPAD) capable of capturing up to  $10^5$  frames per second<sup>14,17–19</sup>. However, to achieve such a frame rate, a trade-off has been made in terms of image quality. The present work addresses this open challenge by developing Artificial Intelligence methods to reduce the number of required frames for extracting effective signals from the typically noisy background of CPI, thus speeding up acquisition while still achieving an established target in terms of image quality.

Artificial Intelligence has led to the widespread use of deep learning (DL) techniques in various fields<sup>20,21</sup>, including image denoising, which has acquired considerable attention. In 2015, Liang et al.<sup>22</sup> and Xu et al.<sup>23</sup>, used deep networks for image denoising tasks, employing the first Convolutional Neural Network (CNN) architecture for this goal. Later, Mao et al.<sup>24</sup> utilized multiple convolutions and deconvolutions to suppress noise and restore high-resolution images. Additionally, Zhang et al.<sup>25</sup> employed the denoising CNN (DnCNN) for image denoising, super-resolution, and JPEG image blocking, through a framework consisted of convolutions, back-normalization, rectified linear unit (ReLU), and residual learning. Considering the tradeoff between denoising performance and speed, Lefkimiatis<sup>26</sup> proposed the color non-local network (CNLNet), which combined non-local self-similarity (NLSS) and CNN to efficiently suppress color-image noise. Also in the context of correlated-photon imaging, the use of deep learning techniques has been explored for addressing the problem of noise; a mutual beneficial effect for both imaging speed and image quality has been demonstrated in two scenarios<sup>27–29</sup>: ghost imaging (GI) and computational ghost imaging (CGI). In these contexts, several DL applications have been implemented to increase the quality of the retrieved images with a reduced number of realizations<sup>30–35</sup>, as well as to extend the use of these imaging techniques for tracking moving objects<sup>36</sup>. It is also worth noticing that the exploratory use of Artificial Intelligence techniques is increasingly spreading to other applications of optics and quantum photonics<sup>37,38</sup>.

In this work, we apply DL techniques to address the noise reduction problem in CPI. Despite its similarities with other correlation-based imaging techniques such as GI and CGI, the effect of noise on the measured four-dimensional correlation function is very specific to CPI and its various alternative architectures<sup>15,39,40</sup>. Thus, models developed previously in other contexts cannot be applied directly and a dedicated approach must be developed. We feed our deep model with a sample of refocused images obtained within the experiment performed in Massaro et al.<sup>16</sup>. We employ a model based on the U-Net architecture, where we use, for the encoding part, the convolution section of a pretrained VGG-19 net, thus realizing a transfer learning model to improve the denoising power. To demonstrate the effectiveness of our deep model, we compare our results with those obtained by using a combination of two well-known image noise-reduction filters, namely, bilateral<sup>41</sup> and Gaussian filters. This work represents the first application of an artificial intelligence method to the field of correlated photon-based plenoptic imaging, and paves the way for extending CPI to scan-free real-time volumetric imaging at video rate.

## Results

We adopted a DL strategy to mitigate noise in CPI. To increase the de-noising capacity of our system, we used a transfer learning approach with a model in which we combined a U-Net architecture and a popular deep CNN named VGG-19 (see “Materials and methods” for details) that were previously trained through the ImageNet database. A scheme of the workflow implemented in the present research is displayed in Fig. 1. We exploited captures of 6 different planar transmissive targets to obtain as many 128x128 pixels refocused images. Then, for each target, we produced three sets of undersampled refocused images (100 images for each set) using three different sampling ratios ( $S$ ) (0.025%, 0.25% and 5%). So, after the image generation procedure, we obtained 6 data sets containing one hundred 128x128 pixels images for each  $S$ . At a fixed  $S$ , we used 5 generated data sets to train the network within a 5-fold cross validation (CV) procedure and the sixth data set to test the model. Inside the 5-fold CV framework, repeated 100 times, we applied a data augmentation procedure on 4 of 5 data sets for a total of 4400 training images after the data augmentation procedure and 100 images used to validate the model. Figure S2 of the Supplementary Materials section shows loss and learning rate as a function of epochs for a single network implementation. To estimate the quality of the output images, we evaluated their Structural Similarity (SSIM) with their respective labels. Further details are given in the “Material and methods” section. The result of 5-fold procedure are reported in Table 1 where SSIM ranging from 0.63 for  $S = 0.025\%$  to 0.92 for  $S = 5\%$ . We reported computational network parameters in Table S1 of the Supplementary Materials section.

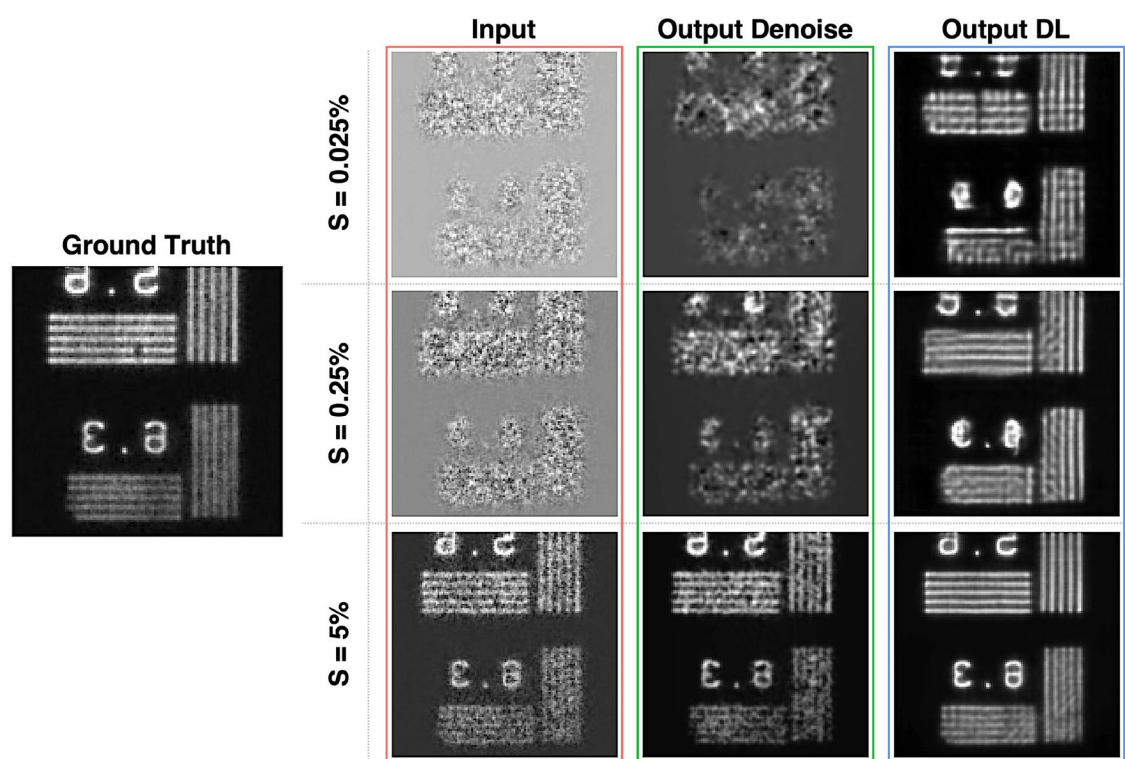


**Figure 1.** The flowchart of CPI using convolutional neural networks (CNNs). Green represents the training stage and orange the testing stage. The steps composing the analysis and the related findings are described in detail in the “Results” and “Materials and methods” sections.

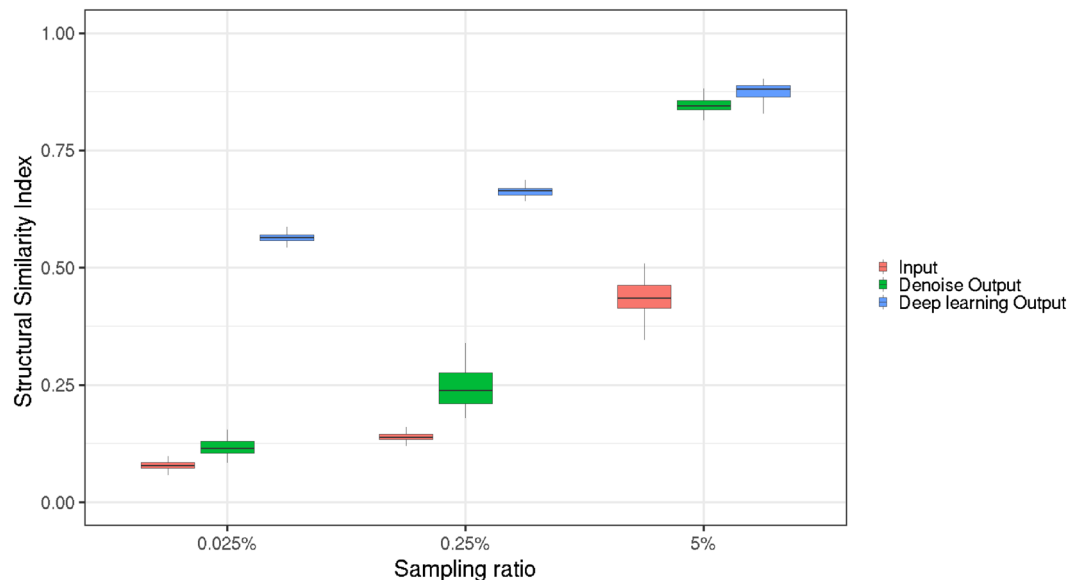
Then we tested the trained model through the sixth data set of generated images. Figure 2 shows a qualitative comparison between an image of the test sample obtained by the standard CPI refocusing algorithm in different  $S$  conditions (Input), the corresponding output images of the conventional denoising and the outputs of DL models. By visual inspection, starting from  $S = 5\%$ , we can see that the image reconstructed by the deep neural network, is almost identical to the ground truth image. At a lower sampling ratio ( $S = 0.25\%$ ), DL algorithm provides a reasonable reconstruction of the target. For an even lower sampling ratio of  $0.025\%$ , we get only a partial reconstruction of the target, but the result is excellent when compared with the initial CPI image. Our denoising strategy (Bilater + Gaussian filters) is able to reconstruct the image with a good yield only in the case of the highest sampling ratio. These findings are confirmed by applying another well-known denoise algorithm, named Block-matching and 3D (BM3D) filter<sup>42</sup>, as shown in Fig. S6 in the Supplementary Materials section. Furthermore, to state the robustness of our results, we show more reconstructions of the test dataset in Figs. S3, S4 and S5. The capability of CNNs to reconstruct much clearer images can also be observed in Fig. 3, where we report, for each sampling ratio, the distributions of SSIM (computed with respect to the ground truth) obtained by applying both our DL framework and the standard denoising strategy to the test sample. DL resulted the best performing method, providing SSIM ranging from 0.46 for  $S = 0.025\%$  to 0.87 for  $S = 5\%$ , and, according to a Kruskal–Wallis test<sup>43</sup>, the two methodologies of noise reduction are significantly different ( $p < 0.1$ ) for each considered  $S$ .

Sampling ratio ( $S$ )	SSIM label-input	SSIM label-output
0.025%	0.050 $\pm$ 0.001	0.630 $\pm$ 0.150
0.25%	0.090 $\pm$ 0.010	0.680 $\pm$ 0.140
5%	0.280 $\pm$ 0.010	0.920 $\pm$ 0.070

**Table 1.** Summary of the performance measure (SSIM) of the proposed deep learning (DL) approach, as obtained through a 5-fold cross validation procedure repeated 100 times. The SSIM values obtained for the 3 considered sampling ratios are reported with the corresponding standard deviations. Input refers to images obtained by CPI; output to images obtained after implementing our DL model.



**Figure 2.** The panel shows a qualitative comparison of the output of our algorithms applied to the test dataset. Starting from the left we see: the Ground Truth image, the input images obtained by the standard undersampled CPI algorithm, the output of the denoising method and the output of DL.  $S$  stands for the sampling ratio.



**Figure 3.** SSIM distributions for the test sample and for the three different sampling ratios so far considered. We computed SSIM (with respect to the ground truth image) for the input CPI images (red), the images obtained with the standard denoise procedure (green) and those from DL (blue).

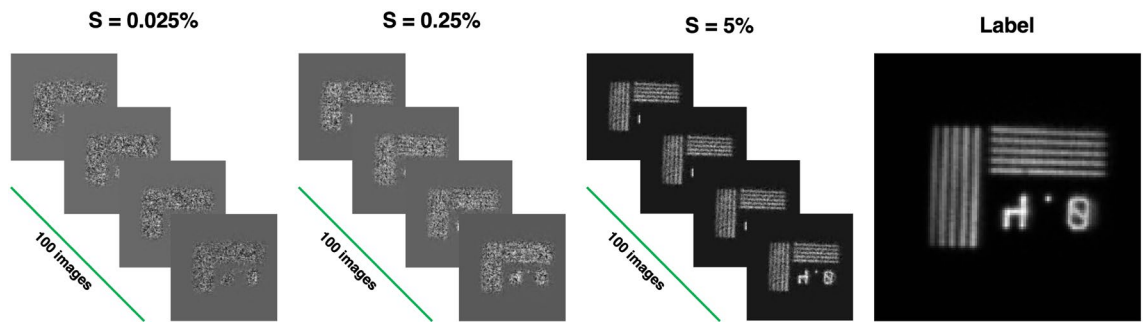
## Discussion

As mentioned earlier, our DL model was trained using experimental images taken at a specific setting, based on the CPI architecture and employing SPAD arrays as sensors as implemented in Massaro et al.<sup>16</sup>. In these refocused images, noise inherent to the image formation procedure, i.e., the computation of pixel-by-pixel intensity correlations, prevents a standard denoising algorithm from significantly increasing the SNR of the image. Our DL model significantly improves the performance with respect to the standard denoising algorithm despite the small training sample. A well known limitation of DL algorithms is that they have poor performances when the number of observations in the training set is too low and thus not informative enough. Therefore, to improve the performance of our model, it would be necessary to have a larger number of images in the training phase, for example by using simulated images based on realistic noise models. Through simulations, the algorithm can also be trained on a richer variety of possible scenarios and objects that are difficult to deal with experimentally, such as complex objects requiring a much larger statistical pool than a ground glass disk, or fast-moving objects. Furthermore, the use of more complex simulated images is reasonably expected to increase the discriminative power of the deep model.

It is worth remarking that a sampling ratio of 5% allows to perform CPI at 10 Hz with a satisfactory SNR level, as demonstrated by the SSIM reported in Fig. 3. However, the application of our DL model enables to obtain SNR values fully comparable with the ground truth, as demonstrated in Fig. 2. This primarily leads a general improvement of image quality under condition of low-SNR, but more specifically it shows that it is possible to further reduce the required number of frames and achieve video rate acquisition speed, as demonstrated by the cases with a lower sampling ratio in Fig. 2. In fact, if we take into account the actual frame rate of the SPAD array used in our experiment (almost 100.000 frames per second), it is possible to estimate a potential acquisition speed of the CPI setup of 200 volumetric images per second for  $S = 0.25\%$ , and even 2000 volumetric images per second for  $S = 0.025\%$ , therefore more than a factor of 20 compared to the fastest CPI demonstrated so far.

A direct comparison of our results with the literature is difficult to make, because our work is the first application of a DL model to CPI. However, as mentioned in the “Introduction”, DL methods for noise reduction have been applied to ghost imaging. In particular, for computational ghost imaging (CGI), Rizvi et al.<sup>33</sup> used deep convolutional autoencoder network to achieve imaging at a frame rate of 4–5 Hz with 10–20% sampling ratios, for reconstructing good-quality  $96 \times 96$  images. Earlier, Lyu et al.<sup>30</sup> and He et al.<sup>31</sup> used DL approaches, also with the support of Compressive Sensing<sup>44</sup>, to reconstruct, still in the CGI framework, good-quality  $32 \times 32$  and  $64 \times 64$  images, respectively, with sampling ratios between 5 and 20%. Recently, Hu et al.<sup>36</sup> demonstrated the possibility to reconstruct both the trajectory and a clear image of a moving object via GI, by using a convolutional denoising auto-encoder network; the quality of images was enhanced with a sampling ratio even down to 3.7%. Our work shows that the employed DL method can achieve a sampling ratio smaller than 1%, with an SSIM of about 0.7. As mentioned before, this result demonstrates the potential of our approach to retrieve volumetric images at a frame rate larger than 200 Hz, well beyond video rate.

We acknowledge that our work presents some other limitations. Mainly, in the training phase of our DL model, the computational demands, in terms of RAM, GPU and computation time, rapidly increase with image resolution. So using higher image resolution (for example  $1024 \times 1024$  pixels) would require a different approach (e.g. image patch calculations).



**Figure 4.** Schematic representation of the construction of each data set: for each sampling ratio  $S$ , 100 images were reconstructed, each one containing a different random noise due to the random choice of the frames from the acquired out-of-focus images of the test target.

Further research will be dedicated to the application of Artificial Intelligence methods directly on the images acquired by the two sensors of CPI setup, in the attempt to entrust DL with both the data analysis, and the denoising stage, by feeding the algorithm the raw data and obtaining the denoised 3D stack of refocused images as the output. This type of approach could revolutionize the CPI technique, because it would definitely overcome the problem of image acquisition and reconstruction times that currently represents the main bottleneck towards real-time volumetric imaging.

## Materials and methods

### Correlation plenoptic imaging

The dataset used to train and test the network is composed of images acquired in a setup based on the concept of *correlation plenoptic imaging between arbitrary planes*<sup>13</sup> (CPI-AP). Massaro et al.<sup>16</sup> describes both its working principle and experimental realization in detail: the conjugate planes of the two high-resolution sensors are located at general axial distances from an imaging lens, in the surroundings of the object of interest. A beam-splitter is used to deflect the chaotic light from the object onto the two sensors. Unlike a conventional light-field camera, which uses both the usual camera lens and a microlens array, our setup is implemented with a single lens that captures light from the selected planes and focuses it on the sensors. To avoid the need for synchronization, the two sensors are realized by using two halves of the same SwissSPAD2 sensor<sup>17,18</sup>; each acquired frame, thus, consists of a binary matrix identifying the pixels triggered by at least one detected photon. The software evaluates the correlations between the photon-number fluctuations, pixel by pixel, between the two halves of the sensor, and reconstructs the volumetric image of the scene. As the light from the scene is chaotic, by calculating the simultaneous pixel-by-pixel correlation between the number of photons detected by the sensors, we obtain the correlation function:

$$\Gamma(\rho_a, \rho_b) = \langle N_a(\rho_a)N_b(\rho_b) \rangle - \langle N_a(\rho_a) \rangle \langle N_b(\rho_b) \rangle, \quad (1)$$

where  $N_a$  ( $N_b$ ) and  $\rho_a$  ( $\rho_b$ ) are the number of photons and the coordinates denoting the pixel positions on the sensors  $a$  ( $b$ ) respectively, while  $\langle \dots \rangle$  indicates the averaging process. The correlation function in Eq. (1) represents the correlation between the intensity fluctuations reaching two points, one placed on the first, and the other on the second detector.  $\Gamma(\rho_a, \rho_b)$  contains plenoptic information and thus allows the reconstruction of features of a 3D object that can lie both between and beyond the two selected planes imaged on the detectors<sup>12,45</sup>.  $\Gamma(\rho_a, \rho_b)$  encodes a collection of multi-perspective volumetric images; proper processing of these volumetric images provides the *refocused image* of a specific transverse plane in the scene.

### Image generation

In our specific case, the CPI device is used for imaging several transmissive planar test targets placed out of focus on both sensors. The targets are illuminated by a chaotic light source with controllable polarization, intensity, and coherence time, generated by a laser scattered by a rotating ground glass disk. We performed a series of acquisitions with the object at different axial positions. For each target, we acquired a large number of frames (larger than 200 k). Following the workflow of the refocusing algorithm<sup>45,46</sup>, we used these data to create a dataset for training and testing the network. Basically, we used 6 acquisitions of different transmissive targets and we exploited each full dataset to achieve a  $128 \times 128$  pixels refocused image of the sample. The retrieved images can be considered as our *ground truth*. However, the behaviour of the SNR of a refocused image deviates from the  $\sqrt{N_t}$  scaling, where  $N_t$  is the number of the acquired frames, since our source can only provide a finite number of statistically independent realizations<sup>16</sup>. For this reason, we estimated the full sampling rate at 200 k detections. Thus, we define the sampling ratio  $S$  as the ratio between the considered *number of acquired frames* and the *total number of acquired frames at the full sampling rate*. To test our DL approach, we generated three sets of undersampled refocused images, for each considered test target, using three different sampling ratios  $S$  (0.025%, 0.25% and 5%). For a given value of the sampling ratio  $S$ , each data set was generated by randomly extracting, from the images of the target directly retrieved by the sensor, a number of frames corresponding to the considered sampling ratio. By repeating the random procedure 100 times, we were able to increase the variability of the data sets while keeping constant the introduced noise level. Figure 4 schematically shows the



composition of each dataset: 5 datasets of  $128 \times 128$  pixels images have been used to train the network for each value of  $S$ , and the remaining one has been used for testing the model. It is worth noting here that all the data sets were built in the same modality, namely, starting from images of different targets placed in different axial positions within the same setup described in the dedicated section.

### Deep learning

After the image collection phase we developed a DL model based on CNN framework to remove image noise. Figure 1 shows a schematic overview of the performed analysis: first, we trained 3 DL models, one for each sampling ratio used (see “Image generation” section for details), then we tested our algorithms on a dataset independent of the training sample.

#### Data augmentation

When there is only a limited number of training samples available, it is crucial to use data augmentation techniques in order to train the network on the required invariance and robustness characteristics. Therefore, we implemented Data Augmentation by rotating the source images at various angles, and also using the transposed image and its rotations. Specifically, for each training image we built: 6 rotations ( $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ,  $180^\circ$ ,  $225^\circ$  and  $270^\circ$ ), the transposed image, 3 rotations of the transposed image ( $90^\circ$ ,  $180^\circ$ ,  $270^\circ$ ). In this way, our training set has reached the number of 4400 images (from the initial 400) for each sampling ratio used.

#### Learning model

CNNs are a class of DL algorithms that have been specifically developed to address various computer vision and image processing tasks<sup>47,48</sup>. CNNs structure is inspired by the visual cortex of some animals<sup>49,50</sup>, and is comprised of three main types of layers: convolutional, pooling, and fully connected layers. In contrast to conventional Artificial Neural Networks, CNNs eliminate the need for a feature engineering and extraction process as the convolutional layer automatically performs these functions. This layer uses both linear and nonlinear operations by applying a fixed-dimension filter (known as a kernel) across the layer’s input during each linear operation. The resulting output is then transferred to a nonlinear activation function. The pooling layer conducts a down-sampling operation over the feature maps’ spatial dimensions. The most frequently utilized pooling operation is max pooling, which extracts fixed-dimension blocks from the input feature maps and retains only the maximum value in each block. Over the last years, deep convolutional networks overcame previous best practices in various visual recognition assignments. In particular, the newly suggested CNNs have significantly improved the removal of image noise procedures because of their powerful expressive capabilities and speedy performance<sup>51</sup>. In this work, we used a transfer learning approach with a model composed by two different algorithms (U-Net and VGG-19 architectures) for a noise reduction task. The U-Net is a type of CNN that was created specifically for the purpose of biomedical image segmentation in 2015<sup>52</sup>. Unlike typical CNNs, the U-Net is designed to be trained using a smaller number of images. Its architecture consists of four coding blocks connected to four decoding blocks via a bridge and four “skip connections” that bring directly in the decoding block the spatial information from the encoding blocks. The encoding part works as a feature extractor and learns, during the training phase, an abstract representation of the image, which the decoding part expands back to its initial size. By using noisy input images and corresponding clean images as labels during the training phase, U-Net can learn to denoise similar images within the same domain as the input images. The VGG-19<sup>53</sup> is a DL model consisting of 19 layers, of which 16 layers are convolutional and the remaining 3 are fully connected. Its main goal is to classify images into 1000 different categories using the ImageNet database, which includes a vast collection of images. The 16 convolutional layers are employed for feature extraction and are divided into five groups, each followed by a max-pooling layer. Finally, the last three layers of the model are used for classification. In this work we used the U-Net architecture, with the encoding part formed by the pre-trained VGG-19 Feature Extraction block. In our configuration we implemented the default parameters: the binary crossentropy loss and Adam optimizer. The U-NET was composed by a final convolutional layer with 1 filter, 1 pixel kernel and sigmoid as activation function. Starting from the weights of VGG-19, pretrained on ImageNet database, we trained our DL model through a cross validation procedure, as detailed in the following section. We reported the network structure in Fig. S1 of the Supplementary Materials section.

#### Cross validation

To increase the robustness of our DL model, we implemented a 5-fold cross validation (CV) framework using 5 of the 6 generated data sets. It is worth emphasizing that the further data set (the sixth) has been used to provide an external validation of the DL algorithm. This technique entails partitioning the original data set into five non-overlapping subsets consisting of the same number of cases, which are assigned to each fold on a random basis. We employed four of the five subsets for training purposes and reserved the remaining portion for validation. On the training sample we applied the data augmentation procedure described in a previous section. We repeated CV 100 times so the average of the 100 performance values is a reliable indicator of the overall model accuracy.

#### Performance metrics

As we anticipated, refocusing images starting from an undersampled correlation function leads to a worsening of the image quality due to statistical noise; in fact, it is well known that the SNR of correlation-based imaging techniques improves with the square root of the number of correlated frames. In CPI, however, attributing a single numerical value to the statistical SNR for estimating the image quality can be ambiguous: from Eq. (1), we see that the SNR of the correlation function, defined as the correlation function itself over its statistical variance, is a local non-homogeneous four-dimensional quantity, depending on all four coordinates at the detectors.

Because of its local nature, the SNR cannot be used as a global image quality estimator for the refocused images as is. Here, we choose a straightforward approach, that is to assess the quality of our image reconstruction in terms of Structural Similarity (SSIM) index<sup>54</sup>, used as a proxy of the statistical SNR, which has the advantage of being a global estimator for the image quality. This index quantifies the degradation of structural information in an image and evaluates the similarity measurement through 3 comparisons: luminance, contrast and structure. Given an image  $\mathbf{x}$  considered to have perfect quality, we can measure quantitatively the quality of a second image  $\mathbf{y}$  by means of a similarity measure with  $\mathbf{x}$ ,

$$SSIM(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^\alpha \cdot [c(\mathbf{x}, \mathbf{y})]^\beta \cdot [s(\mathbf{x}, \mathbf{y})]^\gamma, \quad (2)$$

where  $\alpha$ ,  $\beta$  and  $\gamma$  are positive parameters used to modify the relative importance of the three components. The first term in Eq. (2) indicates the luminance comparison

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2\mu_y^2 + C_1}, \quad (3)$$

with  $\mu_x$  and  $\mu_y$  the mean intensity of  $\mathbf{x}$  and  $\mathbf{y}$  respectively and  $C_1$  a constant. The second term in Eq. (2) represents the contrast comparison function

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2\sigma_y^2 + C_2}, \quad (4)$$

that is expressed as the comparison between the standard deviation of  $\mathbf{x}$  ( $\sigma_x$ ) and  $\mathbf{y}$  ( $\sigma_y$ ). The last component of Eq. (2) defines the structure comparison function

$$s(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad (5)$$

where  $C_3$  is a constant and  $\sigma_{xy}$  is

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y). \quad (6)$$

### Conventional denoising

To demonstrate the effectiveness of our approach, we compared the results with the ones achieved by combining two conventional denoising approaches: Bilateral filter<sup>41</sup>, and a Gaussian filter. The bilateral filter is a well known type of non-linear noise reduction filter having the peculiarity to preserve edges; it has been used in the most diverse contexts, including correlation imaging<sup>32</sup>. Here, it was applied to the reconstructed testing images. Because of the binary nature of the target used in the experiment (i.e. a negative transmissive resolution mask), a Gaussian filter was then applied to the previously filtered images in a minimally invasive way, with the aim to close the possible artificial gap between adjacent pixels<sup>55</sup> inside the correlated regions. We applied this combined noise reduction method to each image of the same data sets generated under the three sampling ratio conditions used for testing our DL model.

### Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 22 June 2023; Accepted: 4 November 2023

Published online: 10 November 2023

### References

- D'Angelo, M., Pepe, F. V., Garuccio, A. & Scarcelli, G. Correlation plenoptic imaging. *Phys. Rev. Lett.* **116**, 223602 (2016).
- Pepe, F. V., Scarcelli, G., Garuccio, A. & D'Angelo, M. Plenoptic imaging with second-order correlations of light. *Quant. Meas. Quant. Metrol.* **3**, 20–26 (2016).
- Pepe, F. V., Di Lena, F., Garuccio, A., Scarcelli, G. & D'Angelo, M. Correlation plenoptic imaging with entangled photons. *Technologies* **4**, 17 (2016).
- Pepe, F. V., Vaccarelli, O., Garuccio, A., Scarcelli, G. & D'Angelo, M. Exploring plenoptic properties of correlation imaging with chaotic light. *J. Opt.* **19**, 114001 (2017).
- Pepe, F. V. *et al.* Diffraction-limited plenoptic imaging with correlated light. *Phys. Rev. Lett.* **119**, 243602 (2017).
- Massaro, G. *et al.* Light-field microscopy with correlated beams for high-resolution volumetric imaging. *Sci. Rep.* **12**, 16823 (2022).
- Lippmann, G. Épreuves réversibles donnant la sensation du relief. *J. Phys. Theor. Appl.* **7**, 821–825 (1908).
- Adelson, E. H. & Wang, J. Y. Single lens stereo with a plenoptic camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 99–106 (1992).
- Ng, R. *et al.* Light field photography with a hand-held plenoptic camera. *Comput. Sci. Tech. Rep. CSTR 2*, 1–11 (2005).
- Scattarella, F., D'Angelo, M. & Pepe, F. V. Resolution limit of correlation plenoptic imaging between arbitrary planes. *Optics 3*, 138–149 (2022).
- Scagliola, A., Di Lena, F., Garuccio, A., D'Angelo, M. & Pepe, F. V. Correlation plenoptic imaging for microscopy applications. *Phys. Lett. A* **1**, 126472 (2020).
- Di Lena, F., Pepe, F. V., Garuccio, A. & D'Angelo, M. Correlation plenoptic imaging: An overview. *Appl. Sci.* **8**, 1958 (2018).
- Di Lena, F. *et al.* Correlation plenoptic imaging between arbitrary planes. *Opt. Express* **28**, 35857–35868 (2020).
- Abbattista, C. *et al.* Towards quantum 3d imaging devices. *Appl. Sci.* **11**, 6414. <https://doi.org/10.3390/app11146414> (2021).

15. Massaro, G., Scala, G., D'Angelo, M. & Pepe, F. V. Comparative analysis of signal-to-noise ratio in correlation plenoptic imaging architectures. *Eur. Phys. J. Plus* **137**, 1123. <https://doi.org/10.1140/epjp/s13360-022-03295-1> (2022).
16. Massaro, G. *et al.* Correlated-photon imaging at 10 volumetric images per second. *Sci. Rep.* **13**, 12813. <https://doi.org/10.1038/s41598-023-39416-8> (2023).
17. Ulku, A. C. *et al.* A  $512 \times 512$  SPAD image sensor with integrated gating for widefield FLIM. *IEEE J. Sel. Top. Quant. Electron.* **25**, 6801212 (2019).
18. Ulku, A. C. *et al.* Wide-field time-gated SPAD imager for phasor-based FLIM applications. *Methods Appl. Fluoresc.* **98**, 024002 (2020).
19. Antolovic, I. M. *et al.* Photon-counting arrays for time-resolved imaging. *Sensors* **16**, 1005 (2016).
20. Amoroso, N. *et al.* Deep learning and multiplex networks for accurate modeling of brain age. *Front. Aging Neurosci.* <https://doi.org/10.3389/fnagi.2019.00115> (2019).
21. Bellantuono, L. *et al.* Predicting brain age with complex networks: From adolescence to adulthood. *NeuroImage* **225**, 117458. <https://doi.org/10.1016/j.neuroimage.2020.117458> (2021).
22. Liang, J. & Liu, R. Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network. in *8th International Congress on Image and Signal Processing (CISP)*, IEEE 697–701. <https://doi.org/10.1109/CISP.2015.7407967> (2015).
23. Xu, Q., Zhang, C. & Zhang, L. Denoising convolutional neural network. in *8th International Congress on Image and Signal Processing (CISP)*, IEEE, 1184–1187. <https://doi.org/10.1109/ICInfA.2015.7279466> (2015).
24. Mao, X., Shen, C. & Yang, Y.-B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Adv. Neural Inf. Process. Syst.* **1**, 2802–2810 (2016).
25. Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **26**, 3142–3155 (2017).
26. Lefkimmiatis, S. Non-local color image denoising with convolutional neural networks. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3587–3596 (2017).
27. Niu, Z. *et al.* Photon-limited face image super-resolution based on deep learning. *Opt. Express* **26**, 22773–22782. <https://doi.org/10.1364/OE.26.022773> (2018).
28. Moodley, C. & Forbes, A. Super-resolved quantum ghost imaging. *Sci. Rep.* **12**, 10346. <https://doi.org/10.1038/s41598-022-14648-2> (2022).
29. Wang, F. *et al.* Far-field super-resolution ghost imaging with a deep neural network constraint. *Light Sci. Appl.* **11**, 1. <https://doi.org/10.1038/s41377-021-00680-w> (2022).
30. Lyu, M. *et al.* Deep-learning-based ghost imaging. *Sci. Rep.* **7**, 17865 (2017).
31. He, Y. *et al.* Ghost imaging based on deep learning. *Sci. Rep.* **8**, 6469 (2018).
32. Shimobaba, T. *et al.* Computational ghost imaging using deep learning. *Opt. Commun.* **413**, 147–151. <https://doi.org/10.1016/j.optcom.2017.12.041> (2018).
33. Rizvi, S., Cao, J., Zhang, K. & Hao, Q. Deepghost: Real-time computational ghost imaging via deep learning. *Sci. Rep.* **10**, 11400. <https://doi.org/10.1038/s41598-020-68401-8> (2020).
34. Li, Z.-M. *et al.* Fast correlated-photon imaging enhanced by deep learning. *Optica* **8**, 323–328. <https://doi.org/10.1364/OPTICA.408843> (2021).
35. Moodley, C., Sephton, B., Rodríguez-Fajardo, V. & Forbes, A. Deep learning early stopping for non-degenerate ghost imaging. *Sci. Rep.* **11**, 8561. <https://doi.org/10.1038/s41598-021-88197-5> (2021).
36. Hu, H.-K., Sun, S., Lin, H.-Z., Jiang, L. & Liu, W.-T. Denoising ghost imaging under a small sampling rate via deep learning for tracking and imaging moving objects. *Opt. Express* **28**, 37284–37293. <https://doi.org/10.1364/OE.412597> (2020).
37. Cimini, V. *et al.* Deep reinforcement learning for quantum multiparameter estimation. *Adv. Photon.* **5**, 016005. <https://doi.org/10.1117/1.AP.5.1.016005> (2023).
38. Gianani, I. & Benedetti, C. *Multiparameter Estimation of Continuous-Time Quantum Walk Hamiltonians Through Machine Learning* (2022). <http://arxiv.org/abs/2211.05626>.
39. Scala, G., D'Angelo, M., Garuccio, A., Pascasio, S. & Pepe, F. V. Signal-to-noise properties of correlation plenoptic imaging with chaotic light. *Phys. Rev. A* **99**, 053808 (2019).
40. De Scisciolo, E. *et al.* Nonclassical noise features in a correlation plenoptic imaging setup. *Int. J. Quant. Inf.* **18**, 1941017. <https://doi.org/10.1142/S021974991941017X> (2020).
41. Tomasi, C. & Manduchi, R. Bilateral filtering for gray and color images. in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, 839–846. <https://doi.org/10.1109/ICCV.1998.710815> (1998).
42. Mäkinen, Y., Azzari, L. & Foi, A. Collaborative filtering of correlated noise: Exact transform-domain variance for improved shrinkage and patch matching. *IEEE Trans. Image Process.* **29**, 8339–8354. <https://doi.org/10.1109/TIP.2020.3014721> (2020).
43. Kruskal, W. & Wallis, W. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**, 583–621 (1952).
44. Katz, O., Bromberg, Y. & Silberberg, Y. Compressive ghost imaging. *Appl. Phys. Lett.* **95**, 131110 (2009).
45. Massaro, G., Pepe, F. V. & D'Angelo, M. Refocusing algorithm for correlation plenoptic imaging. *Sensors* **22**, 6665 (2022).
46. Massaro, G., Di Lena, F., D'Angelo, M. & Pepe, F. V. Effect of finite-sized optical components and pixels on light-field imaging through correlated light. *Sensors* **22**, 778. <https://doi.org/10.3390/s22072778> (2022).
47. Yamashita, R., Nishio, M., Do, R. & Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **9**, 611–629 (2018).
48. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
49. Hubel, D. & Wiesel, T. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **195**, 215–243 (1968).
50. Fukushima, K. Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**, 193–202 (1980).
51. Bansal, M., Kumar, M., Sachdeva, M. & Mittal, A. Transfer learning for image classification using vgg19: Caltech-101 image data set. *J. Ambient Intell. Hum. Comput.* <https://doi.org/10.1007/s12652-021-03488-z> (2021).
52. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention: MICCAI 2015*. MICCAI 2015 Vol. 9351 (eds Navab, N. *et al.*) (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
53. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. in *3rd International Conference on Learning Representations (ICLR 2015)*, 1–14 (2015).
54. Wang, Z., Bovik, A., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612. <https://doi.org/10.1109/TIP.2003.819861> (2004).
55. Kim, J. *et al.* Ghost imaging with Bayesian denoising method. *Opt. Express* **29**, 39323–39341. <https://doi.org/10.1364/OE.438478> (2021).

## Acknowledgements

F.S. is supported by Research for Innovation REFIN - Regione Puglia POR PUGLIA FESR-FSE 2014/2020; G.M., F.V.P. and M.D. are supported by Istituto Nazionale di Fisica Nucleare (INFN) through project QUISS. M.D. G.M. and F.V.P. acknowledge funding under project ADEQUADE: this project has received funding from the European



Defence Fund (EDF) under grant agreement EDF-2021-DIS-RDIS-ADEQUADE (n°101103417). M.D. G.M and F.V.P. are supported by project Qu3D, funded by the Italian Istituto Nazionale di Fisica Nucleare, the Swiss National Science Foundation (Grant 20QT21 187716 “Quantum 3D Imaging at high speed and high resolution”), the Greek General Secretariat for Research and Technology, the Czech Ministry of Education, Youth and Sports, under the QuantERA programme, which has received funding from the European Union’s Horizon 2020 research and innovation programme. M.D. is supported by PNRR MUR project PE0000023—“National Quantum Science and Technology Institute”. F.V.P. is supported by PNRR MUR project CN00000013—“National Centre for HPC, Big Data and Quantum Computing”. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

### Author contributions

Conceptualization, F.S., A.M. and M.D.; methodology, F.S., D.D. and A.M.; software, F.S., D.D. and A.M.; formal analysis, F.S., D.D. and A.M.; writing-original draft preparation, F.S. and A.M.; writing-review and editing, F.S., D.D., A.M., N.A., L.B., G.M., F.V.P., S.T., R.B. and M.D.; visualization, F.S. and A.M.; supervision, R.B. and M.D. All authors have read and agreed to the published version of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-46765-x>.

**Correspondence** and requests for materials should be addressed to A.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