PHOTONICS Research

Learning-based super-resolution interpolation for sub-Nyquist sampled laser speckles

HUANHAO LI,[1,2,†] ZHIPENG YU,[1,2,†] QI ZHAO,[1,2,†] YUNQI LUO,[3] SHENGFU CHENG,[1,2]
TIANTING ZHONG,[1,2] CHI MAN WOO,[1,2] HONGLIN LIU,[1,4] LIHONG V. WANG,[5,7] YUANJIN ZHENG,[3,8] AND
PUXIANG LAI[1,2,6,9]

[1]Department of Biomedical Engineering, Hong Kong Polytechnic University, Hong Kong, China
[2]Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518063, China
[3]School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639798, Singapore
[4]Key Laboratory for Quantum Optics, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China
[5]Caltech Optical Imaging Laboratory, Andrew and Peggy Cherng Department of Medical Engineering, California Institute of Technology, Pasadena, California 91125, USA
[6]Photonics Research Institute, Hong Kong Polytechnic University, Hong Kong, China
[7]e-mail: LVW@caltech.edu
[8]e-mail: yjzheng@ntu.edu.sg
[9]e-mail: puxiang.lai@polyu.edu.hk

Information retrieval from visually random optical speckle patterns is desired in many scenarios yet considered challenging. It requires accurate understanding or mapping of the multiple scattering process, or reliable capability to reverse or compensate for the scattering-induced phase distortions. In whatever situation, effective resolving and digitization of speckle patterns are necessary. Nevertheless, on some occasions, to increase the acquisition speed and/or signal-to-noise ratio (SNR), speckles captured by cameras are inevitably sampled in the sub-Nyquist domain via pixel binning (one camera pixel contains multiple speckle grains) due to finite size or limited bandwidth of photosensors. Such a down-sampling process is irreversible; it undermines the fine structures of speckle grains and hence the encoded information, preventing successful information extraction. To retrace the lost information, super-resolution interpolation for such sub-Nyquist sampled speckles is needed. In this work, a deep neural network, namely SpkSRNet, is proposed to effectively up sample speckles that are sampled below 1/10 of the Nyquist criterion to well-resolved ones that not only resemble the comprehensive morphology of original speckles (decompose multiple speckle grains from one camera pixel) but also recover the lost complex information (human face in this study) with high fidelity under normal- and low-light conditions, which is impossible with classic interpolation methods. These successful speckle super-resolution interpolation demonstrations are essentially enabled by the strong implicit correlation among speckle grains, which is non-quantifiable but could be discovered by the well-trained network. With further engineering, the proposed learning platform may benefit many scenarios that are physically inaccessible, enabling fast acquisition of speckles with sufficient SNR and opening up new avenues for seeing big and seeing clearly simultaneously in complex scenarios.   © 2023 Chinese Laser Press

https://doi.org/10.1364/PRJ.472512

1. INTRODUCTION

Light experiences strong scattering in complex media, and the interference of multiply scattered photons traveling along different paths leads to the formation of speckles, if the coherence length of light (e.g., laser light) is sufficiently long [1]. Information carried by light is therefore scrambled yet deterministically encoded in these speckles, which are visually observed as spatially isolated bright spots. Such grainy representations physically serve as the information carrier and require effective

processing and deciphering. Although challenging, it is feasible to analyze and/or retrieve the scrambled information through wavefront shaping [2–9] and phase retrieval [10–13]. Recent development of artificial intelligence [14,15] has, as well, inspired this field by fulfilling the image reconstruction from speckles via either purely learning-based methods [16–21] or their counterpart with physics prior [22,23].

Successful deciphering and reconstruction of the encoded information usually require sufficient acquisition of the speckle

details, although some recent studies implied that down-sampled speckles could also retain certain information [24,25]. That said, the lower limit of down sampling for information retaining is elusive and no clear criterion, such as with respect to the Nyquist criterion, has been studied. In principle, sub-Nyquist sampling to an image can be treated as a low-pass filter, causing loss of high-frequency information (or fine features) to the image. Once the speckles are sampled below the Nyquist criterion, the low-pass filter integrates several grains into one pixel and more importantly, the hidden information is assumed to be significantly lost (to be proved in the study). On the other hand, spatially sub-Nyquist sampled speckles are necessary in some scenarios, such as fast speckle acquisition of a large field of view (FOV) and under low-light conditions. Detection of more speckle grains or a large FOV is desired to encompass more effective information in the speckle pattern [25]. This requires more camera pixels (slows down the acquisition) or resolves individual speckle grains in a sub-Nyquist domain due to limited camera pixels. A similar dilemma also needs to be addressed under low-light conditions, where camera pixel binning might be necessary to increase the acquisition speed via bypassing the time-consuming averaging and/or to enhance the signal-to-noise ratio (SNR). Note that sparsity in detected signals and the insufficient sampling processes are irreversible. Pre-processing to convert those speckles to their well-sampled counterparts becomes inevitable before other post-processing procedures such as image reconstruction and transmission matrix calculation. To ensure the robustness of such reversion process, two hypotheses are spurred. (1) Explicit recovery: can the down-sampled or sparse speckle patterns be morphologically recovered to well-resolved counterparts? (2) Implicit recovery: can the recovered speckle patterns retrace the lost information due to insufficient sampling or pixel binning?

For the second hypothesis, the validation is equivalent to image reconstruction from speckles, which has been intensively explored via physical modeling [26] or deep learning [19,20,22]. Comparably, the first hypothesis has been rarely explored, probably due to the lack of physical tool to analyze and reverse such a down-sampling process. Intuitively, an image without fine features might correspond to many of well-resolved images. The issue arising from the first hypothesis is therefore an ill-posed problem. To address it, single-image super-resolution has been proposed to enhance the sampling frequency and imaging resolution in the community of deep learning [27,28]. For example, deep neural network (DNN) can automatically discover the implicit local/non-local correlation within a down-sampled image, which is hard to define in the current theoretical framework, and parametrically generate up-sampled output approaching the ground truth. The implementation has been successfully utilized in optical microscopies in the ballistic regime [29,30], where distortion of object information is relatively low. In the non-ballistic regime, e.g., strong scattering, the object information is entirely disordered and displayed in the form of speckle grains covering a large FOV. The more is the information complexity, the worse is the distortion. As a result, interpolating sub-Nyquist sampled speckle patterns, especially those encoded with complicated information, is highly challenging.

Few studies, if any, have explored whether sub-Nyquist sampled speckles can be effectively processed. In this work, we investigate how the sub-Nyquist sampling affects information reconstruction from the speckles and introduce an implementation of DNN, namely SpkSRNet, to parametrically interpolate sub-Nyquist speckle patterns to well-resolved representations and retrieve the lost information. The success of retrieval is validated by further feeding the interpolated speckles into another DNN, namely SpeckleNet, to accomplish generalized image reconstruction. Experimental results show that the proposed learning-based super-resolution implementation can handle very sparsely sampled speckles (below ∼1/10 of the Nyquist criterion). As a result, not only the high-frequency information (i.e., comprehensive grainy morphology) of the speckle patterns but also the lost information can be effectively recovered, allowing for robust retrieval of complicated information such as human face. Notably, the SpkSRNet is trained for super-resolution with no priors of the hidden information in the training. That means the SpkSRNet can learn the implicit constraints among speckle grains and the way information is encoded in the speckles and lost due to down sampling, which is achieved by adopting well-resolved speckles as the only target for training. This study provides a new perspective to understand the nature beneath speckles, such as information encoding and decoding, and may potentially serve a powerful platform to process or decipher speckled signals under extreme conditions, such as low light or high noise.

2. METHODS

A. Experimental Setup

The experimental apparatus is configured in Fig. 1. Images extracted from the "Flickr Faces High Quality" (FFHQ) database are used as the phase objects and displayed on a phase modulation spatial light modulator (SLM, HOLOEYE PLUTO VIS056 1080p, German). A collimated continuous-wave coherent laser beam at $\lambda = 532$ nm (EXLSR-532-300-CDRH, Spectra Physics, USA) is expanded, so that the screen of the SLM is fully illuminated. The laser beam is modulated by the SLM with phase objects uploaded in advance. The phase-only SLM converts 8-bit gray scale (0–255) to phase delay (0-2π in radian), and the images from the FFHQ database are rescaled to 0–127 for phase delay ranging from 0 to π to increase the modulation efficiency. After being modulated and reflected by the SLM, the laser beam is converged onto a diffuser (220-grid, DG10-220-MD, Thorlabs, USA) by an objective (RMS20X, Olympus, Japan). A total of 20,000 images from the FFHQ database are sequentially displayed on the SLM. The corresponding scattering-induced speckle patterns are one by one captured by a CMOS camera (FL3-U3-32S2M-CS, PointGrey, Canada). Speckles under different laser light conditions, characterized by SNR (the ratio of the speckle mean intensity to its variance) [1], are recorded by merely changing the output power from the laser: for the normal-light condition, optical power before entering the scattering medium is 240 μW, and the corresponding SNR = 1; for the low-light condition, the optical power is reduced by 100 times to ∼2.4 μW, with a corresponding SNR = 0.04 where only ∼0.14% pixels are non-minimal by statistics. The camera and the SLM are

synchronized via a MATLAB program in data acquisition. Note that in this study, it means "laser speckle" whenever "speckle" is mentioned.

B. Data Preparation

The recorded speckle patterns (with the dimension of 252×252) and the corresponding images (with the dimension of 32×32) displayed on the SLM are paired up for DNN training. In this study, 20,000 human face samples for each sampling pitch are generated, 18,000 of which are used for DNN training while the rest are used for network validation. To ensure that differently sampled speckle patterns are evaluated under the same experimental conditions and not affected by the decorrelation of the medium, the down-sampled speckle patterns are computationally obtained through pixel binning, a process that is usually deemed irreversible as shown in Fig. 1. Thus, the sampling pitch (d_s) can be equivalently defined as $d_0 \times n$ μm by grouping n-by-n neighboring camera pixels as a macropixel, where n is the sampling factor and can be any divisor of 252. For example, in Columns I–VII in later shown Fig. 3(a), the speckle patterns are sampled with $d_s = d_0$, $4d_0$, $8d_0$, $12d_0$, $18d_0$, $21d_0$, and $28d_0$, respectively. For each sampling factor, a corresponding dataset is generated, namely nd_0-dataset, with each including 18,000 samples for network training and 2000 samples for the test. Note that the samples in the training and test sets do not overlap. Also, the average diameter of the speckle grains is 9.77 μm, as measured by the full

width at half-maximum of the autocorrelation of speckle patterns [31]. Therefore, the Nyquist criterion (d_c) to sample speckles in this study is $9.77/2 \sim 4.88$ μm, which corresponds to $1.952d_0$. Thus, for sampling factors larger than 2, the speckles are sampled in the sub-Nyquist domain.

C. DNN Architecture and Training Strategy

For speckle interpolation, DNNs, named as SpkSRNet, based on a modified ResNeXt-101 [32] with a pixel-shuffle layer [33] are used in this study, as shown in Fig. 2. The ResNeXt-101 is memory friendly and contains more parallel units than the U-net like networks have. Notably, the original ResNeXt-101 has 101 layers, six of which are designed for down sampling. To match this study, these down-sampling layers are modified or removed so that the output from ResNeXt-101 can retain the dimension of the images and expand the number of channels from 1 (in the input) to n^2 (in the output), where n is the sampling factor. Therefore, incorporating the parameter-free pixel-shuffle layer, a widely used layer for pixel super-resolution in video community, the proposed network can be trained efficiently with limited computational resources.

The down-sampled speckles from six nd_0-datasets ($n = 4$, 8, 12, 18, 21, 28) are individually paired up with corresponding speckles from the d_0-dataset as the input–output pairs, and the resultant network is denoted as nd_0-trained SpkSRNet. A two-step training is applied to each SpkSRNet as follows for optimal performance: (1) the networks are pre-trained by
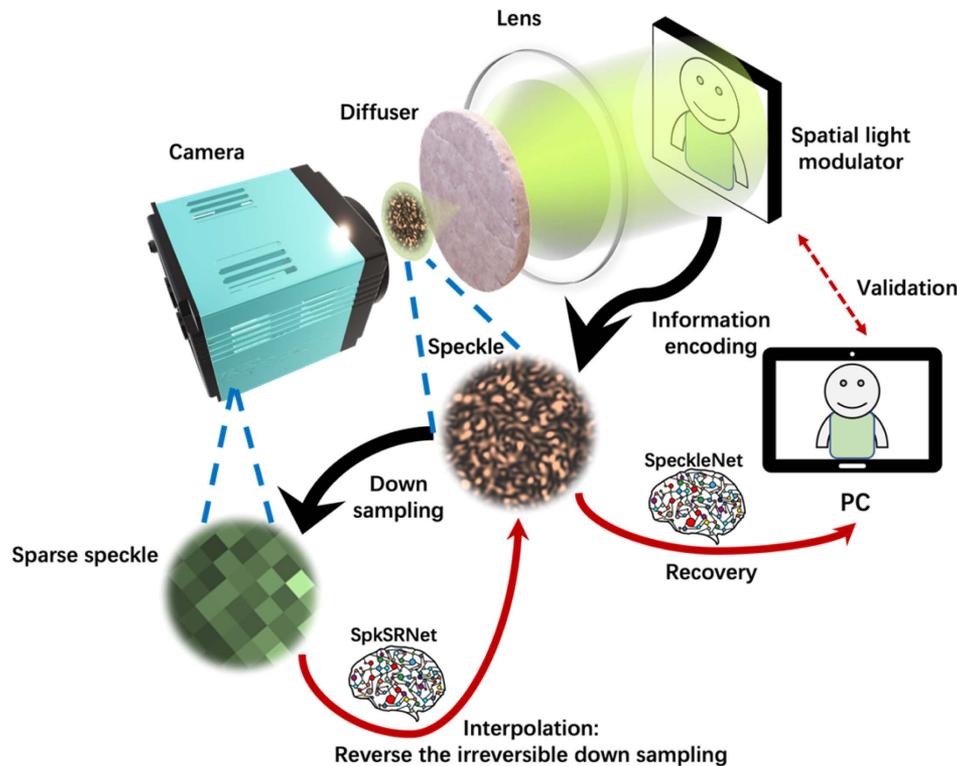

Fig. 1. Conceptional diagram of speckle collection, deep-learning-based speckle super-resolution interpolation processing and information recovery. Phase objects are displayed on the SLM, which is illuminated by an expanded continuous coherent laser beam ($\lambda = 532$ nm). Speckle patterns behind the scattering medium are down sampled for rapid recording, transmission, and storage, or enhanced SNR. Inevitably, intrinsic correlations among speckle grains and hence the encoded information are irreversibly impaired. In signal processing, the down-sampled speckles are interpolated via a deep-learning based super-resolution framework to reverse the "irreversibility" and retrace with high-fidelity the speckle comprehensive morphology and finally the encoded object information.
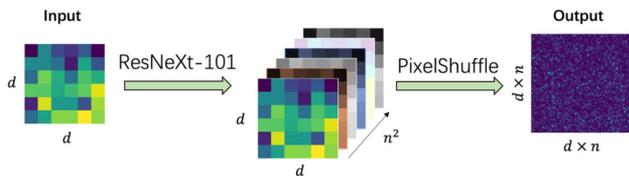
Fig. 2. Architecture of SpkSRNet is the combination of ResNeXt and PixelShuffle layers. nd_0-sampled speckles with the dimension of $d \times d$ ($d = 252/n$) are input into the SpkSRNet for speckle super-resolution (i.e., interpolation) processing ($n = 4, 8, 12, 18, 21, 28$), with d_0-sampled speckles (i.e., the original speckles with the dimension of 252×252) as the target. For example, when the down-sampling factor (n) is 12, the network is called $12d_0$-trained SpkSRNet, whose input dimension is 21×21 and output dimension is 252×252.

minimizing the loss function (F_1) given by Eq. (3); (2) the pre-trained networks are further optimized via minimizing the loss function (F_2) expressed by Eq. (4). In both steps, the stochastic gradient descent algorithm (SGD) is used for optimization with a batch size of 16 and a weight decay of 0.01, and the learning rates are set to be 0.01 and 0.001 for the first and second steps, respectively, with cosine annealing for 50 epochs:

$$f_1(y, \hat{y}) = \text{NPCC}(y, \hat{y}) + \text{MSE}(y, \hat{y})$$
$$= -\frac{\langle (y - \langle y \rangle)(\hat{y} - \langle \hat{y} \rangle) \rangle}{\sigma_y \sigma_{\hat{y}}} + \langle ||y - \hat{y}||^2 \rangle, \quad \textbf{(1)}$$

$$f_2(y, \hat{y}) = -\text{SSIM}(y, \hat{y}) + \text{CL}(y, \hat{y})$$
$$= -\frac{(2\langle y \rangle \langle \hat{y} \rangle + c_1)(\sigma_{y\hat{y}} + c_2)}{(\langle y \rangle^2 + \langle \hat{y} \rangle^2 + c_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + c_2)}$$
$$+ \langle \sqrt{(y - \hat{y})^2 + \epsilon^2} \rangle, \quad \textbf{(2)}$$

$$F_1(y, \hat{y}) = f_1(y, \hat{y}) + f_1[S_n(y), S_n(\hat{y})] + f_1(\nabla y, \nabla \hat{y}), \quad \textbf{(3)}$$

$$F_2(y, \hat{y}) = f_2(y, \hat{y}) + f_2[S_n(y), S_n(\hat{y})] f_2(\nabla y, \nabla \hat{y}), \quad \textbf{(4)}$$

where y is the target and $\hat{y}$ is the output from the network with average operation $\langle \rangle$, σ is the variance, $|| ||$ is absolute operation, S_n is the pixel binning operator with the down-sampling factor (n), and ∇ is to calculate the gradient of the image. In Eq. (1), the sum of the negative Pearson's correlation coefficient (NPCC) and mean square error (MSE) contributes to the quick convergence. In Eq. (2), the sum of the negative structural similarity index measure (SSIM) with $c_1 = 0.0001$ and $c_2 = 0.0003$ and the Charbonnier loss (CL) with $\epsilon = 0.001$. Specifically, the SSIM can better reflect the visual quality of the two-dimensional features and the CL helps to confine the fine textures in generated images during optimization [34], which matches the speckle interpolation task. Nevertheless, these two perceptually metrics are hard to be directly optimized so they were used to fine-tune the network in the second step.

For image reconstruction, with a network called SpeckleNet, a complex fully connected layer is applied, with the speckles as the input and the corresponding human face as the target, since one complex fully connected layer can mostly represent the physical model, i.e., transmission matrix, of the strong optical

scattering and it works well for complicated information [35]. Notably, this model is not fit for the speckle interpolation since it probably provides diverse optimization with less efficiency to analyze two-dimensional information. Likewise, the reconstruction network is individually trained by seven datasets for each d_s and is denoted as nd_0-trained SpeckleNet ($n = 1, 4, 8, 12, 18, 21, 28$). All the SpeckleNets are optimized by minimizing f_1 via the SGD, whose learning rate is 0.001 with cosine annealing for 500 epochs. The training framework is Pytorch 1.9.0 with Python 3.7, using CUDA for GPU acceleration. The computing unit is a Dell precision workstation with E5-1620v3, 64 GB RAM, and an RTX3090 GPU.

3. RESULTS

A. Sampling Effect and Classic Interpolation

First, fully resolved speckle patterns [Fig. 3(a), Column I] are collected by the camera as shown in Fig. 1. The original speckle grains are well resolved by 12 pixels on average (each pixel area is d_0^2). With a larger down-sampling factor, more speckle grains are combined into one pixel and the grainy representations fade, as seen from Columns I to VII in Fig. 3(a). These speckle patterns sampled at different sampling factors are then fed into the corresponding SpeckleNet, i.e., nd_0-sampled speckles from the test set are input into the nd_0-trained SpeckleNet. By doing so, whether the object information is included in the down-sampled speckles can be validated. Their corresponding outputs are the reconstructed human faces [Fig. 3(b)]. As seen, when fed with d_0-sampled speckles [Fig. 3(a), Column I], the SpeckleNet can successfully reconstruct the human face [Fig. 3(b), Column I)], with the Pearson's correlation coefficient (PCC) of 0.9841 and MSE of 0.0047 compared with the ground truth [Fig. 3(c)], which is the baseline performance in this study. When fed with speckles with nd_0-sampled speckles [$n \geq 4$, Columns II–VII in Fig. 3(a)], the detailed features of the human face are lost [Columns II–VII in Fig. 3(b)]. For $n = 4$, although the PCC is around 0.8, details in the visualized image are significantly lost and the MSE becomes 6 times larger than the baseline [Fig. 3(b), Column II)]. Furthermore, only a blurry outline of the face is left when $n \geq 8$ [Columns III–VII in Fig. 3(b)]. The quantified metrics [Fig. 3(d)] also validate the information loss with the increasing sampling factor. An enlarged tendency is also observed for the standard deviation [the color-filled region in Fig. 3(d)]. Such fluctuating performance might indicate that the intrinsic correlations or constraints within one speckle pattern are undermined by the down sampling, which relaxes the exclusive transformation between speckles and encoded information. Therefore, to retrace the lost intrinsic correlations or information, the down-sampling process needs to be effectively reversed, such as via interpolation. Note that in the following sections, the **SpeckleNet will act as the tool to validate** whether the object information is contained in the interpolated speckles.

Two classic interpolation methods, bilinear and bicubic, are first selected to up sample the poorly resolved speckle patterns [Fig. 4(a)], aiming to generate d_0-sampled speckles. As seen in Figs. 4(b1) and 4(c1), morphological representations achieved by these two methods are similar, which tend to smoothen the morphology compared to the down-sampled ones [Fig. 4(a)].
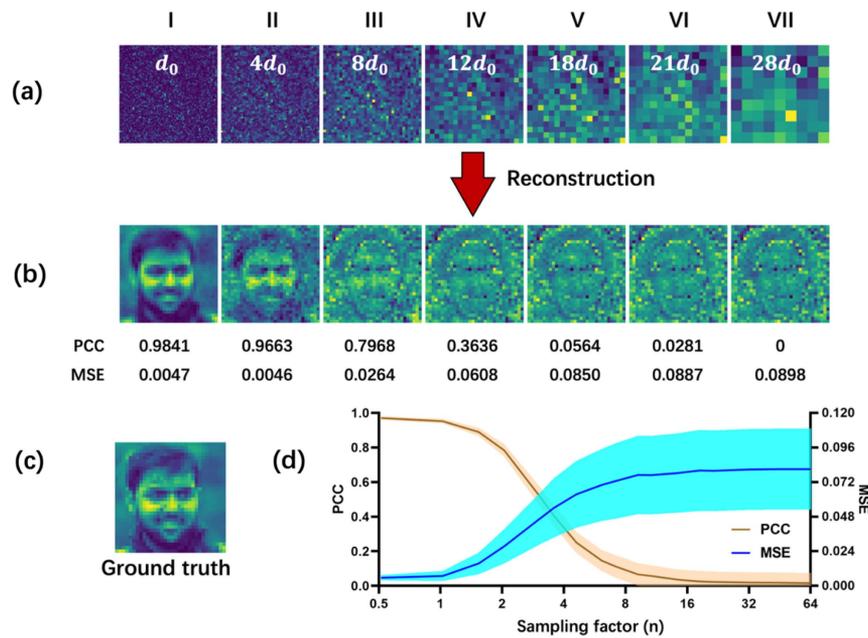
Fig. 3. Sampling effect for information recovery. (a) Representative speckle patterns with different sampling factors. From Columns I to VII, speckles are sampled with sampling factors of 1, 4, 8, 12, 18, 21, and 28, respectively. (b) Reconstructed images via the corresponding SpeckleNet, with the PCC and MSE information with respect to the ground truth (c). (d) PCC and MSE of the reconstructed face images with respect to the ground truth as a function of the down-sampling factor (n). The solid line and the colored region represent the mean value and the standard deviation, respectively, among the 2000 test samples. The ground truth image (c) is reproduced under terms of the Public Domain Mark 1.0 license, and captured by DOCB Bengaluru on 2017-08-06, Flickr (https://www.flickr.com/photos/docb400/35592623433/). The original image is cropped and converted to gray scale. Images in (b) are generated via the deep learning output based on the transformation of (c) for signal processing purpose.

For $4d_0$-sampled speckle patterns [one speckle grain sampled by around one pixel in Fig. 4(a), Column I], the interpolated features are similar to those of d_0-sampled speckles [Fig. 4(d)], based on which reconstructed information can be more or less

recognized as a face [Column I in Figs. 4(b2) and 4(c2)]. That said, for larger sampling factors ($n \geq 8$) that combine several speckle grains into one pixel, both classic interpolation methods fail to decompose individual speckle grains from one pixel, and
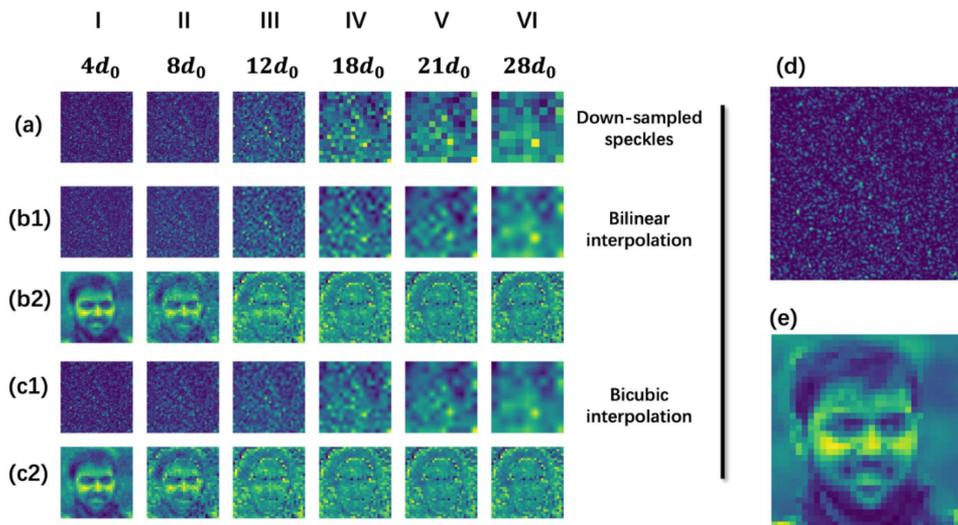

Fig. 4. Speckle interpolation based on classic methods and the corresponding learning-based imaging reconstruction. (a) Down-sampled speckle patterns with a sampling factor from 4 to 28. (b1) and (c1) are the up-sampled (i.e., interpolated) speckle patterns through bicubic and bilinear interpolations, respectively. (b2) and (c2) are the reconstructed images by feeding (b1) and (c1) are into the SpeckleNet. (d) d_0-sampled speckles. The ground truth image (e) is reproduced under terms of the Public Domain Mark 1.0 license, and captured by DOCB Bengaluru on 2017-08-06, Flickr (https://www.flickr.com/photos/docb400/35592623433/). The original image is cropped and converted to gray scale. Images in (b2) and (c2) are generated via the deep learning output based on the transformation of (e) for signal processing purpose.

the reconstructed information is far from the baseline [Fig. 3(b), Column I]. Moreover, through comparison the reconstructed information based on bilinearly or bicubically interpolated speckles [Figs. 4(b2) and 4(c2)] is almost the same as those from speckles without processing [Fig. 3(b)]. It suggests that significant down sampling cannot be overcome by classic bilinear or bicubic interpolations. Non-linear interpolation with higher orders (say, larger than 3) is therefore required to effectively up sample the down-sampled speckles. Deep learning, featured by its non-linear characteristics, is hence called for in the next section.

B. Learning-Based Speckle Interpolation and Information Recovery

A DNN-based network called SpkSRNet was trained, and the interpolation as well as information reconstruction results are shown in Fig. 5. As seen, the learning-based interpolated speckles exhibit promising representations like the original speckles [Fig. 5(a), Column I]. It indicates that the six SpkSRNets are able to decompose multiple speckle grains from one binned pixel and recover the grainy features from the down-sampled speckle patterns; the randomly scattered bright spots can be clearly identified in all the realizations, which apparently outperforms the classic interpolations (Fig. 4). For example, the $28d_0$-sampled speckles with the dimension of only 9×9 can be up sampled to speckles with the dimension of 252×252, which are highly consistent with the original speckles (PCC of 0.9557, MSE of 0.0008), as denoted in Column VII in Fig. 5(b). In comparison, if up-sampled by the classic methods, the same speckles merely become smoother, as shown in Column VI in Figs. 4(b1) and 4(c1), and the resultant PCCs ($\sim$0.15) are below one-sixth of the DNN realizations ($\sim$0.95). Moreover, the morphological features of these interpolated

patterns from SpkSRNet are highly similar due to their high PCC values [>0.95 in Fig. 5(b)] with the original speckle pattern.

Reconstructing the implicit information hidden in these speckles is the other critical approach to validate the accuracy of the super-resolution interpolation processing. By feeding the interpolated speckles into the d_0-trained SpeckleNet, human faces can be seen with fine features for all cases, as shown in Columns II–VII in Figs. 5(c). Meanwhile, it should be noted that the quality of the reconstructed images, gauged by either visual perception or PCC/MSE, decreases with increased sampling factors. For example, the reconstructed image generated from the interpolated $28d_0$-sampled speckles [Fig. 5(c), Column VII]] apparently deviates from the $4d_0$ realization [Fig. 5(c), Column II)], with a PCC of 0.8963 versus 0.9783, although the differences between the corresponding interpolated speckles [Fig. 5(b), Column VII and Fig. 5(b), Column II] are hardly recognized, with a PCC of 0.9557 versus 0.9869.

To further compare the performance of speckle super-resolution with different interpolation methods, the quantitative metrics, including PCC, MSE, and SSIM with respect to the corresponding ground truths based on the 2000 testing samples at each sampling factor, are plotted in Fig. 6. As seen, no matter which interpolation method is adopted, better quantities of metrics (i.e., larger PCC/SSIM and smaller MSE) can be obtained with a smaller sampling factor, which is intuitive and consistent with the visual observations. With increasing sampling factor, however, the curves of metrics are noticeably divided into two groups: one for speckle interpolations with classic methods and no processing, and the other for the learning-based interpolations that are much less sensitive to the sampling factor. In the first group, the performances with classic
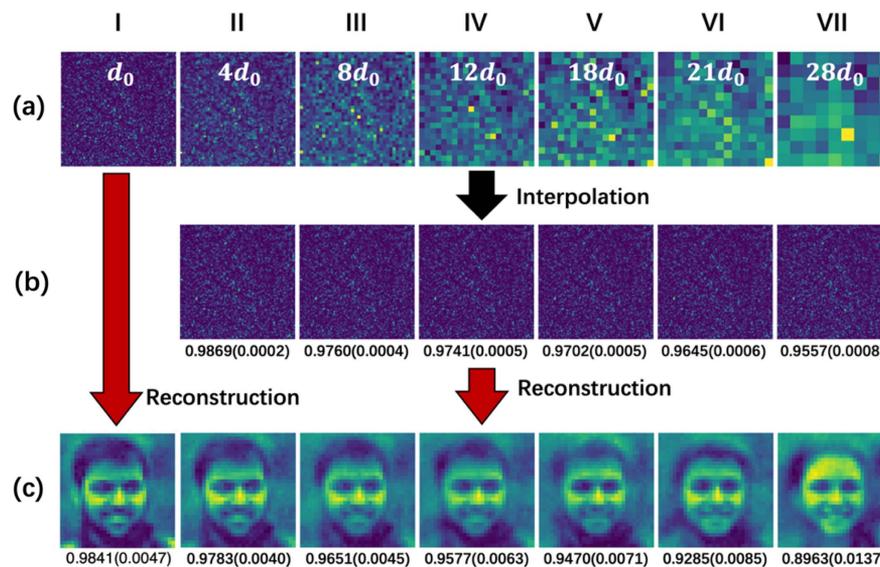

Fig. 5. Learning-based super-resolution interpolation and imaging reconstruction. (a) Speckle patterns with a sampling factor from 1 to 28. (b) The corresponding interpolated speckles via SpkSRNet for down-sampled speckle patterns. The insets are the PCC (MSE) with respect to the ground truth speckle pattern in Column I of panel (a). (c) The corresponding reconstructed images via SpeckleNet for d_0-sampled and interpolated speckles. The insets are the PCC (MSE) with respect to the ground truth face image [Fig. 3(c)]. Images in (c) are generated via deep learning output based on the transformation of Fig. 3(c), which is reproduced under terms of the Public Domain Mark 1.0 license, and captured by DOCB Bengaluru on 2017-08-06, Flickr (https://www.flickr.com/photos/docb400/35592623433/).
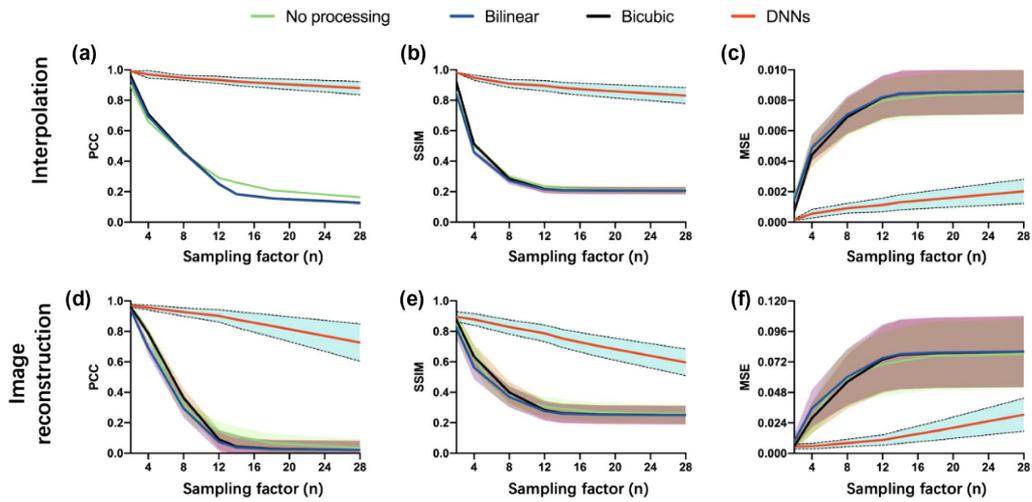
Fig. 6. Performance analysis of speckle interpolation and image reconstruction. (a)–(c) The PCC (a), SSIM (b), and MSE (c) between the interpolated speckles of interest and the original speckles (d_0-sampled). (d)–(f) The PCC (d), SSIM (e), and MSE (f) between the corresponding reconstructed images and the target human face. The solid lines and the shadowed regions represent the mean value and the standard deviation, respectively, from the 2000 testing samples.

interpolation methods and no processing almost overlap and decay sharply with increased sampling factor. These once again confirm that the classic bilinear and bicubic interpolation methods hardly contribute to the improvement of morphology and information recovery for the down-sampled speckles. In the other group, with learning networks, the similarities of interpolated speckles and reconstructed information with respect to the corresponding ground truths, as measured by PCC and SSIM, are significantly boosted than those in the first group. For example, for $28d_0$ realizations, the learning-based PCC and SSIM for interpolation are increased by more than 5 times, while for image reconstruction the enhancement is also striking. Meanwhile, the variations from the ground truths,

as measured by the MSE, are considerably suppressed with the learning networks.

All these demonstrate the effectiveness and superior performance of learning-based networks on speckle interpolation and recovery of information that is lost or undermined due to insufficient sampling, which may be necessary in speckle recording and/or storage for higher SNR, faster frame rate, or storage space saving.

C. Speckle Super-Resolution Interpolation under Low-Light Condition

Speckle-based applications under low-light condition constitute such scenes where camera pixel binning may be necessary
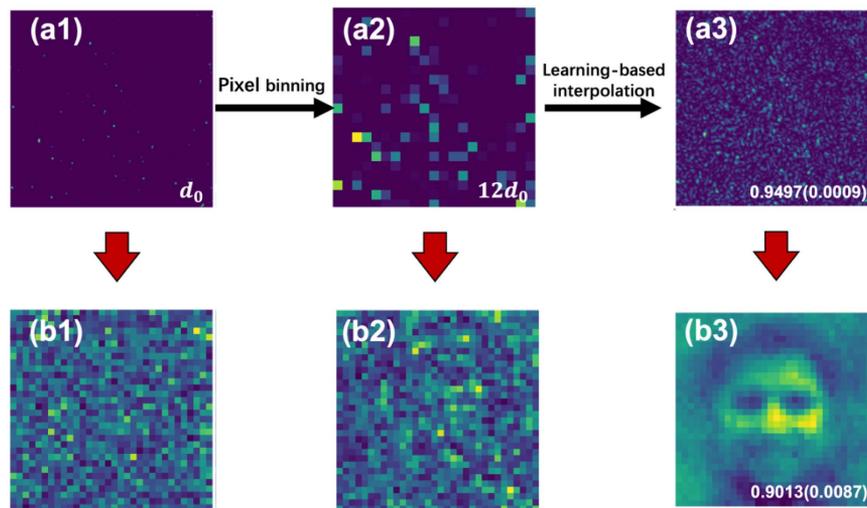

Fig. 7. Learning-based speckle interpolation and information recovery under low-light condition. (a1) Speckles collected in low-light condition (optical power = 2.4 μW and SNR = 0.04). (a2) Down-sampled speckles (SNR = 0.28) through combining 12 × 12 pixels in (a1). (a3) The learning-based interpolated speckles by feeding (a2) into the SpkSRNet trained by the $12d_0$-sampled speckles. (b1)–(b3) are the reconstructed images of (a1)–(a3), respectively, via the SpeckleNet. The red downward arrows represent learning-based image reconstruction. The image in (b3) is generated via the deep learning output based on the transformation of Fig. 3(c), which is reproduced under terms of the Public Domain Mark 1.0 license, and captured by DOCB Bengaluru on 2017-08-06, Flickr (https://www.flickr.com/photos/docb400/35592623433/).

to ensure fast acquisition and usable SNR, which, however, causes insufficient sampling to resolve the speckle grains well. As an example, Fig. 7(a1) shows a speckle pattern recorded by the camera under low-light condition (i.e., optical power = 2.4 µW, SNR ≈ 0.04 as defined in Section 2.A) with $n = 1$ for the sampling factor. The corresponding speckle pattern under normal-light condition (SNR ≈ 1.00) is shown in Column I in Fig. 3(a) and the encoded information in Fig. 3(c). As seen, the speckle pattern under low-light condition is of high-level sparsity with only several scattered bright spots that can be visualized in the FOV; the encoded human face information is entirely gone with the learning-based information reconstruction method [Fig. 7(b1)]. Through combining 12 × 12 pixels into one mega pixel [Fig. 7(a2)], which is also a down-sampling processing (sampling factor $n = 12$), the speckle SNR is enhanced from 0.04 to 0.28. Based on such pixel-binned speckles, the reconstructed image only exhibits a very vague outline of a human face [Fig. 7(b2)]. The application of SpkSRNet, however, brings the hope. As shown in Fig. 7(a3), the SpkSRNet, trained with the $12d_0$-sampled low-light speckles as the input and the corresponding d_0-sampled normal-light speckles as the target, generates an ultra-fine textured speckle pattern. This speckle pattern not only accomplishes the task of super-resolution interpolation, but also compensates for the sparsity with a PCC as high as 0.9497 with respect to the ground truth [Fig. 3(a1)]. Further, based on such learning-based interpolated speckles, the encoded human face information can be reconstructed with high fidelity [Fig. 7(b3)]. This group of experiments shows that under extreme conditions such as low light, complex information encoded in speckles is hard to be accurately extracted through learning the sparse speckles, including the direct low-SNR speckles or their down-sampled copy with higher SNR. Learning-based speckle interpolation, however, provides a promising strategy to effectively process down-sampled speckles, turning the seemingly useless very sparse speckles into valuable data from which encoded information can be recovered.

4. DISCUSSION

Through experiments, with sampling factor $n \geq 4$ or several speckle grains binned into a mega-pixel, direct and classically interpolated image reconstructions fail even with deep learning (Figs. 3 and 4). The disability has been, to some extent, overcome by the proposed learning-based interpolation by improving the grain features of the intermediate speckle patterns (Fig. 5). To validate whether or not they preserve the correct object information, visual observation based on speckle PCC (MSE) and information consistency based on reconstructed face images are both evaluated. Performance of the interpolation degrades with increased sampling factor, which essentially introduces more sampling insufficiency into the captured speckles and hence the information loss. With acceptable information loss (e.g., PCC drops to 0.9 for the human face recovered from the interpolated speckles), the largest down-sampling factor supported in this study is around 21 (Fig. 5) and 12 (Fig. 7) under normal- and low-light conditions,

respectively. More examples can be checked in Appendix A (Figs. 8–11).

These successful demonstrations may essentially provide some physical implications or new knowledge about speckles and speckle correlations. For example, as seen in Column VI in Fig. 3(a), the grainy morphology of the $21d_0$-sampled speckles has been entirely smoothed out because the signal at each macropixel equivalently represents $21^2 = 441$ pixels in the d_0-sampled speckles and integrates quite a few speckle grains. Accordingly, the object information is gone [Fig. 3(b), Column VI]. The learning-based interpolation [Fig. 5(b), Column VI] can effectively decompose the former pixel integration by searching the optimal solution among $[(2^{16})^{21 \times 21}]^{12 \times 12} \sim 2^{106}$ possibilities (for up sampling 16-bit $21d_0$-sampled speckles with 12×12 pixels to 16-bit d_0-sampled speckles with 252×252 pixels). The seemingly irreversible down sampling (i.e., pixel binning) is reversed after learning the transformation from the down-sampled to well-resolved data. Also, in the low-light realizations, such as the $12d_0$-sampled low-light speckles [Fig. 7(a2)], bright speckle grains left in the FOV comprise only ~0.14% non-minimal signals, from which the signals in the rest of the FOV can be still "derived" by the well-trained SpkSRNet [Fig. 7(a3)]. The feasible transformation proves the objective existence of implicit yet strong correlation among speckles. Therefore, as seen, sparsely sampled or partial speckled signals can contain the overall information of the target, although with compromised resolution.

The implicit correlation among speckles might also define the high-order fluctuating intensity and gauge the global distribution of the speckle grains. Without such physics prior, it is therefore impossible for the classic interpolation methods to recover the grainy morphology as well as the global distribution. The inherent non-linearity in the DNNs, on the other hand, provides feasibility to overcome the lack of physics knowledge about the implicit speckle correlation. Instead of producing uncorrelated random signals, the DNNs allow for adaptively inserting the values to form the morphological feature of speckle grains to decompose the integrated pixels. Such distinguished capability should be attributed to the supervised configuration of the deep learning method. With supervised learning, targets in the training dataset specifically guide the update of parameters in the DNNs. By minimizing the discrepancy between the network output and the target (the well-resolved speckles) through backpropagation optimization, the information from the target flows back across each layer and updates their parameters to relate the down-sampled speckles. That is why the DNN is able to learn the effective transformation for speckle interpolation. Therefore, for speckles under either normal- or low-light conditions, the proposed learning-based interpolation approach provides a robust and promising solution to retrieve object information from down-sampled speckles.

Before concluding, the difference of super-resolution interpolation for regular and laser speckled images should be emphasized. As known, some essential information may be lost during the process of down or sparse sampling. For regular images, such as natural scenes and human faces, resolution

improvements by 4 to 8 times (the same definition as the sampling factor in this study) have been reported [28,36]. However, implementation of super-resolution based on the low-resolution images leads to diverse and non-unique high-resolution results [36]. Such uncertainty is induced by the delta-like point spread function (PSF) of the lens-based imaging system for regular images, which localizes the information from the object to a diffraction-limited narrow region on the image plane. The correlation among neighboring signals is therefore inevitably weak so that it is much more difficult to fulfill the super-resolution interpolation for regular images, which may result in incorrect artifacts [37]. For laser speckled images, the scattering medium defines a speckle-like PSF [10] and sufficiently delocalizes the information. Theoretically, the whole region on the image plane is conjugated to one point on the object plane, and vice versa, making the plane wavelets emanate from the object remotely correlated on the image plane. Thus, the delocalization

significantly strengthens the correlation over the FOV or among the speckle grains, and only a few speckle signals can encode the complex information. Consequently, even with down sampling, the preserved intrinsic correlation enables learning-based speckle interpolation, whose cut-off sampling factors correspond to 21 and 12 (i.e., 10.7 and 6.15 times below the Nyquist criterion), respectively, for normal- and low-light realizations in this study. Such effective information preservation from the sub-Nyquist sampled speckles indicates the information carriers in speckles are probably not limited to the fine features but, more essentially, its implicit correlation is buried among the spatial distribution of the speckle grains.

5. CONCLUSION

In this work, we propose a learning-based implementation (i.e., SpkSRNet based on a ResNeXt with a pixel-shuffle layer)
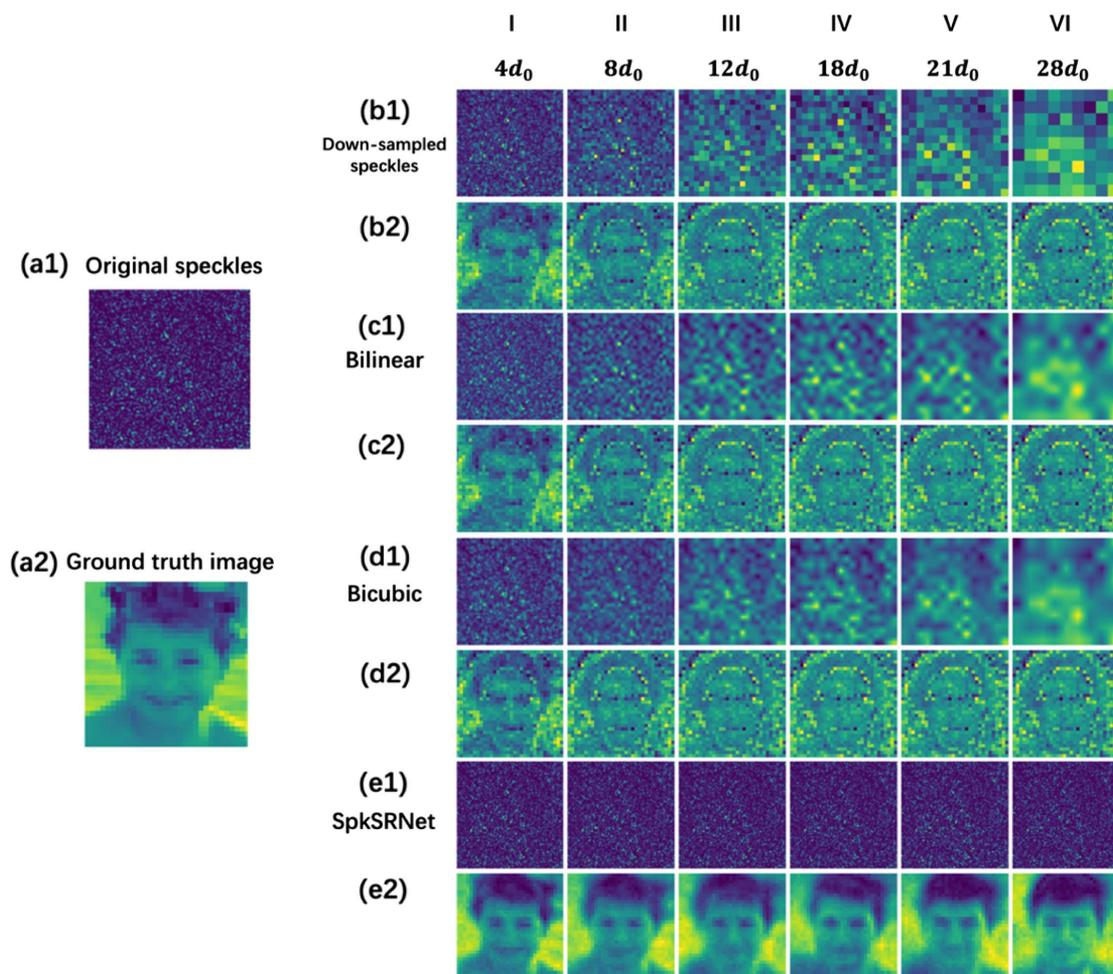

Fig. 8. Speckle interpolation and the corresponding learning-based imaging reconstruction. (a1) The original speckles generated with the image (a2). (b1) Down-sampled speckle patterns with sampling factors from 4 to 28. (c1), (d1), and (e1) are the up-sampled (i.e., interpolated) speckle patterns through bilinear, bicubic, and SpkSRNet methods, respectively. (b2), (c2), (d2), and (e2) are the reconstructed images by feeding (b1), (c1), (d1), and (e1) into the SpeckleNet. The ground truth image (a2) is reproduced under terms of the Public Domain Mark 1.0 license, and captured by Lionel AZRIA on 2018-05-15, Flickr (https://www.flickr.com/photos/157170122@N07/28252868748/). The original image is cropped and converted to gray scale. Images in (b2), (c2), (d2) and (e2) are generated via the deep learning output based on the transformation of (a2) for signal processing purpose.

to parametrically up sample sparse laser speckle patterns that are sampled below ~1/10 of the Nyquist criterion into well-resolved speckle patterns, which is impossible with classic interpolation methods. Experimental results show that the proposed method provides a robust and promising framework to recover the comprehensive speckle morphology as well as the lost complex information (human face in this study) with high fidelity from sub-Nyquist sampled speckles that are down sampled from the original ones by more than two magnitudes (21^2 and 12^2 for normal- and low-light conditions, respectively).

Notably, if the purpose is image reconstruction, the SpkSRNet and SpeckleNet can be merged so that sparse speckles can be fed into a single network to reconstruct the image. Yet, the application of the learning-based interpolation network can be generalized due to its independence from the image reconstruction procedure. In practice, the information about the target object to be imaged or sensed may be unknown, which impedes regular DNN training for image reconstruction. The proposed interpolation, however, exploits speckle patterns and its transformation only. Therefore, the training for SpkSRNet can proceed even with no prior knowledge of the target objects. Such a feature can considerably benefit many scenarios such as the measurement for a large transmission matrix of a scattering medium, where speckled acquisition is restricted within a specific temporal window, and biomedical applications whose aim is to reveal inhomogeneities that are optically inaccessible. This may also enable applications based on fast acquisition of speckles with sufficient SNR and open up new avenues for seeing big and seeing clearly simultaneously in complex scenarios.

APPENDIX A

Since each sample in this study is involved with multiple sampling conditions, only one speckles image is chosen to save the length of the text and brief demonstration. Therefore, two more examples, in Figs. 8, 9, and 10, 11, respectively, are listed here for reference, including normal-light conditions in Figs. 8 and 10 and low-light conditions in Figs. 9 and 11.

Author Contributions. H. L., Z. Y, and Q. Z. contributed equally to this work. Y. L., H. L., Z. Y, and Y. L. worked on simulation. H. L. and Z. Y. contributed to the experiments. C. W. and H. L helped to fulfill the low-light experiments. L. V. W., Y. Z., and P. L. conceived and supervised the research. All authors contributed to results discussion, manuscript writing, and revision.

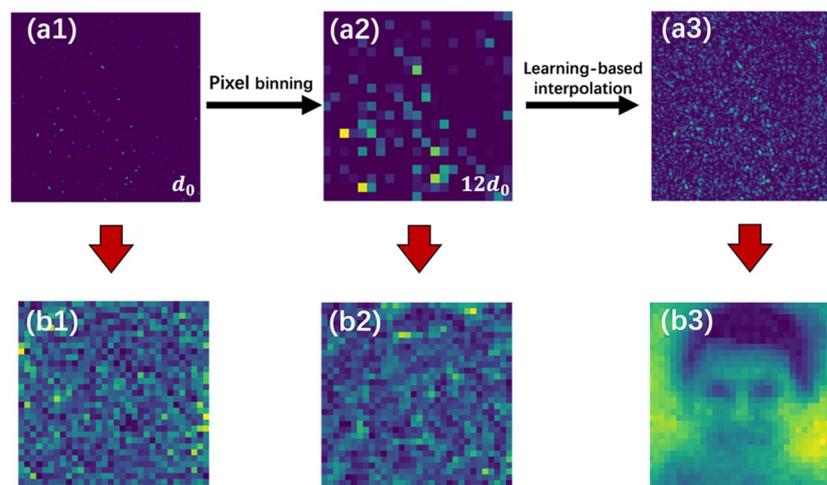**Disclosures.** The authors declare no conflicts of interest.

Fig. 9. Learning-based speckle interpolation and information recovery under low-light condition. (a1) Speckles collected in low-light condition (optical power = 2.4 μW and SNR = 0.04). (a2) Down-sampled speckles (SNR = 0.28) through combining 12×12 pixels in (a1). (a3) The learning-based interpolated speckles by feeding (a2) into the SpkSRNet trained by the $12d_0$-sampled speckles. (b1)–(b3) are the reconstructed images of (a1)–(a3), respectively, via the SpeckleNet. The red downward arrows represent learning-based image reconstruction. The image in (b3) is generated via the deep learning output based on the transformation of Fig. 8(a2), which is reproduced under terms of the Public Domain Mark 1.0 license, and captured by Lionel AZRIA on 2018-05-15, Flickr (https://www.flickr.com/photos/157170122@N07/28252868748/).
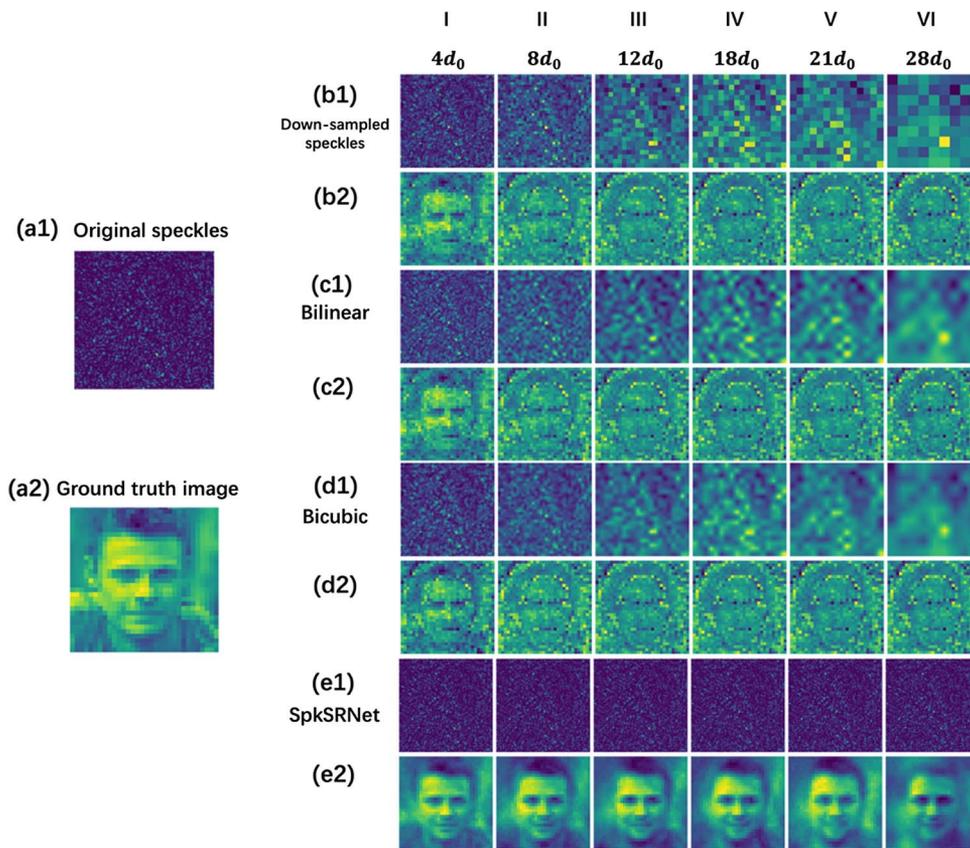
Fig. 10. Speckle interpolation and the corresponding learning-based imaging reconstruction. (a1) Original speckles generated with the image (a2). (b1) Down-sampled speckle patterns with sampling factors from 4 to 28. (c1), (d1), and (e1) are the up-sampled (i.e., interpolated) speckle patterns through bilinear, bicubic, and SpkSRNet methods, respectively. (b2), (c2), (d2), and (e2) are the reconstructed images by feeding (b1), (c1), (d1), and (e1) into the SpeckleNet. The ground truth image (a2) is reproduced under terms of the Public Domain Mark 1.0 license, and captured by Kya-Lynn on 2018-06-05, Flickr (https://www.flickr.com/photos/141074874@N05/41849168674/). The original image is cropped and converted to gray scale. Images in (b2), (c2), (d2), and (e2) are generated via the deep learning output based on the transformation of (a2) for signal processing purpose.
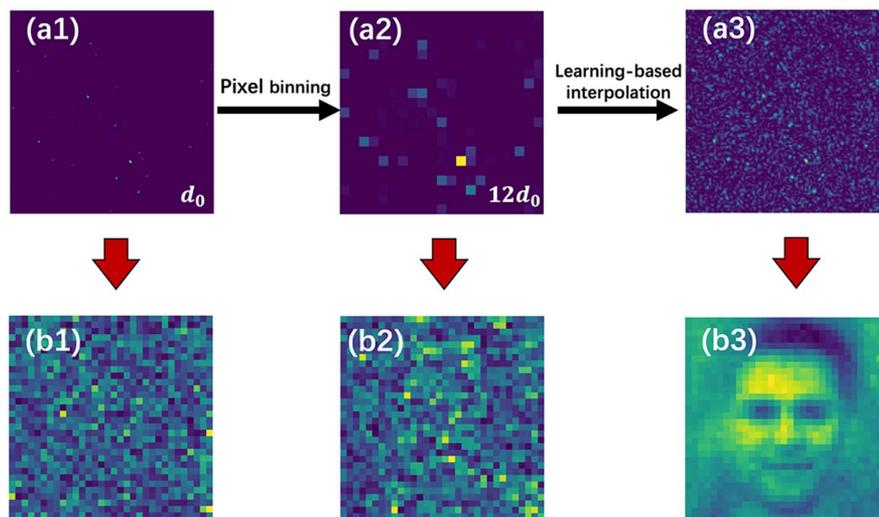

Fig. 11. Learning-based speckle interpolation and information recovery under low-light condition. (a1) Speckles collected in low-light condition (optical power = 2.4 μW and SNR = 0.04). (a2) Down-sampled speckles (SNR = 0.28) through combining 12×12 pixels in (a1). (a3) The learning-based interpolated speckles by feeding (a2) into the SpkSRNet trained by the $12d_0$-sampled speckles. (b1)–(b3) are the reconstructed images of (a1)–(a3), respectively, via the SpeckleNet. The red downward arrows represent learning-based image reconstruction. The image in (b3) is generated via the deep learning output based on the transformation of Fig. 10(a2), which is reproduced under terms of the Public Domain Mark 1.0 license, and captured by Kya-Lynn on 2018-06-05, Flickr (https://www.flickr.com/photos/141074874@N05/41849168674/).

Data Availability.   Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

†These authors contributed equally to this work.

REFERENCES

1. J. W. Goodman, *Speckle Phenomena in Optics: Theory and Applications* (SPIE, 2015).
2. I. M. Vellekoop and A. Mosk, "Focusing coherent light through opaque strongly scattering media," Opt. Lett. **32**, 2309–2311 (2007).
3. S. Popoff, G. Lerosey, R. Carminati, M. Fink, A. Boccara, and S. Gigan, "Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media," Phys. Rev. Lett. **104**, 100601 (2010).
4. P. Lai, L. Wang, J. W. Tay, and L. V. Wang, "Photoacoustically guided wavefront shaping (PAWS) for enhanced optical focusing in scattering media," Nat. Photonics **9**, 126–132 (2015).
5. J. Xu, H. W. Ruan, Y. Liu, H. J. Zhou, and C. H. Yang, "Focusing light through scattering media by transmission matrix inversion," Opt. Express **25**, 27234–27246 (2017).
6. J.-H. Park, Z. Yu, K. Lee, P. Lai, and Y. Park, "Perspective: wavefront shaping techniques for controlling multiple light scattering in biological tissues: toward *in vivo* applications," APL Photonics **3**, 100901 (2018).
7. H. Li, C. M. Woo, T. Zhong, Z. Yu, Y. Luo, Y. Zheng, X. Yang, H. Hui, and P. Lai, "Adaptive optical focusing through perturbed scattering media with a dynamic mutation algorithm," Photonics Res. **9**, 202–212 (2021).
8. Y. Luo, S. Yan, H. Li, P. Lai, and Y. Zheng, "Towards smart optical focusing: deep learning-empowered dynamic wavefront shaping through nonstationary scattering media," Photonics Res. **9**, B262–B278 (2021).
9. Z. Yu, H. Li, T. Zhong, J.-H. Park, S. Cheng, C. M. Woo, Q. Zhao, J. Yao, Y. Zhou, X. Huang, W. Pang, H. Yoon, Y. Shen, H. Liu, Y. Zheng, Y. Park, L. V. Wang, and P. Lai, "Wavefront shaping: a versatile tool to conquer multiple scattering in multidisciplinary fields," Innovation **3**, 100292 (2022).
10. O. Katz, P. Heidmann, M. Fink, and S. Gigan, "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations," Nat. Photonics **8**, 784–790 (2014).
11. T. Wu, O. Katz, X. Shao, and S. Gigan, "Single-shot diffraction-limited imaging through scattering layers via bispectrum analysis," Opt. Lett. **41**, 5003–5006 (2016).
12. J. Bertolotti, E. G. van Putten, C. Blum, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Non-invasive imaging through opaque scattering layers," Nature **491**, 232–234 (2012).
13. H. Yilmaz, E. G. van Putten, J. Bertolotti, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Speckle correlation resolution enhancement of wide-field fluorescence imaging," Optica **2**, 424–429 (2015).
14. Y. Xu, X. Liu, X. Cao, C. Huang, E. Liu, S. Qian, X. Liu, Y. Wu, F. Dong, C. W. Qiu, J. Qiu, K. Hua, W. Su, J. Wu, H. Xu, Y. Han, C. Fu, Z. Yin, M. Liu, R. Roepman, S. Dietmann, M. Virta, F. Kengara, Z. Zhang, L. Zhang, T. Zhao, J. Dai, J. Yang, L. Lan, M. Luo, Z. Liu, T. An, B. Zhang, X. He, S. Cong, X. Liu, W. Zhang, J. P. Lewis, J. M. Tiedje, Q. Wang, Z. An, F. Wang, L. Zhang, T. Huang, C. Lu, Z. Cai, F. Wang, and J. Zhang, "Artificial intelligence: a powerful paradigm for scientific research," Innovation **2**, 100179 (2021).
15. H. Li, Z. Yu, Q. Zhao, T. Zhong, and P. Lai, "Accelerating deep learning with high energy efficiency: from microchip to physical systems," Innovation **3**, 100252 (2022).
16. S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," Optica **5**, 803–813 (2018).
17. N. Borhani, E. Kakkava, C. Moser, and D. Psaltis, "Learning to see through multimode fibers," Optica **5**, 960–966 (2018).
18. B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, and C. Moser, "Multimode optical fiber transmission with a deep learning network," Light Sci. Appl. **7**, 69 (2018).
19. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," Optica **5**, 1181–1190 (2018).
20. P. Caramazza, O. Moran, R. Murray-Smith, and D. Faccio, "Transmission of natural scene images through a multimode fibre," Nat. Commun. **10**, 2029 (2019).
21. Q. Zhao, H. Li, Z. Yu, C. M. Woo, T. Zhong, S. Cheng, Y. Zheng, H. Liu, J. Tian, and P. Lai, "Speckle-based optical cryptosystem and its application for human face recognition via deep learning," Adv. Sci. **9**, e2202407 (2022).
22. S. Zhu, E. Guo, J. Gu, L. Bai, and J. Han, "Imaging through unknown scattering media based on physics-informed learning," Photonics Res. **9**, B210–B219 (2021).
23. M. Liao, S. Zheng, S. Pan, D. Lu, W. He, G. Situ, and X. Peng, "Deep-learning-based ciphertext-only attack on optical double random phase encryption," Opto-Electron. Adv. **4**, 200016 (2021).
24. Z. Zhou, J. Xia, J. Wu, C. Chang, X. Ye, S. Li, B. Du, H. Zhang, and G. Tong, "Learning-based phase imaging using a low-bit-depth pattern," Photonics Res. **8**, 1624–1633 (2020).
25. M. Lyu, H. Wang, G. Li, S. Zheng, and G. Situ, "Learning-based lensless imaging through optically thick scattering media," Adv. Photonics **1**, 036002 (2019).
26. S. Popoff, G. Lerosey, M. Fink, A. C. Boccara, and S. Gigan, "Image transmission through an opaque material," Nat. Commun. **1**, 81 (2010).
27. C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: a benchmark," in *European Conference on Computer Vision (ECCV)* (Springer International Publishing, 2014), pp. 372–386.
28. Z. Wang, J. Chen, and S. C. H. Hoi, *Deep Learning for Image Super-Resolution: A Survey* (ITPAM 2020).
29. H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Gunaydin, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," Nat. Methods **16**, 103–110 (2019).
30. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," Optica **4**, 1437–1443 (2017).
31. J. W. Goodman, *Statistical Optics* (Wiley, 2015).
32. S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2017), pp. 1492–1500.
33. W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 1874–1883.
34. W. S. Lai, J. B. Huang, N. Ahuja, and M. H. Yang, "Fast and accurate image super-resolution with deep Laplacian Pyramid networks," IEEE Trans. Pattern Anal. Mach. Intell. **41**, 2599–2613 (2019).
35. M. Pascucci, S. Ganesan, A. Tripathi, O. Katz, V. Emiliani, and M. Guillon, "Compressive three-dimensional super-resolution microscopy with speckle-saturated fluorescence excitation," Nat. Commun. **10**, 1327 (2019).
36. W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: a brief review," IEEE Trans. Multimedia **21**, 3106–3121 (2019).
37. S. Cheng, Y. Zhou, J. Chen, H. Li, L. Wang, and P. Lai, "High-resolution photoacoustic microscopy with deep penetration through learning," Photoacoustics **25**, 100314 (2022).