



# One-step robust deep learning phase unwrapping

KAIQIANG WANG,<sup>1</sup> YING LI,<sup>1</sup> QIAN KEMAO,<sup>2,3</sup> JIANGLEI DI,<sup>1,4</sup> AND JIANLIN ZHAO<sup>1,5</sup>

<sup>1</sup>MOE Key Laboratory of Material Physics and Chemistry under Extraordinary Conditions, Shaanxi Key Laboratory of Optical Information Technology, School of Science, Northwestern Polytechnical University, Xi'an 710072, China

<sup>2</sup>School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore

<sup>3</sup>MKMqian@ntu.edu.sg

<sup>4</sup>jiangleidi@nwpu.edu.cn

<sup>5</sup>jlzhao@nwpu.edu.cn

**Abstract:** Phase unwrapping is an important but challenging issue in phase measurement. Even with the research efforts of a few decades, unfortunately, the problem remains not well solved, especially when heavy noise and aliasing (undersampling) are present. We propose a database generation method for phase-type objects and a one-step deep learning phase unwrapping method. With a trained deep neural network, the unseen phase fields of living mouse osteoblasts and dynamic candle flame are successfully unwrapped, demonstrating that the complicated nonlinear phase unwrapping task can be directly fulfilled in one step by a single deep neural network. Excellent anti-noise and anti-aliasing performances outperforming classical methods are highlighted in this paper.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. Introduction

Phase calculation is required in many measurement and imaging techniques, such as synthetic aperture radar (SAR) interferometry [1], optical interferometry [2], wave-front compensation [3,4], and magnetic resonance imaging (MRI) [5], to yield a physical quantity of interest. However, the phase of a wavefront directly calculated from a complex exponent is wrapped in the range of  $(-\pi, \pi]$  [6]. The wrapped phase must be unwrapped to provide an estimate of the underlying physical quantity. There are two classical spatial strategies to solve the phase unwrapping problem: minimum-norm methods, path-following methods [7,8]. However, when the wrapped phase contains noise or aliasing (undersampling), the minimum-norm methods cannot limit the propagation of phase error caused by unreliable data points in space, getting an incorrect unwrapped phase, while the path-following methods will encounter the integration-path inconsistency problem [6]. In addition to the spatial method, there are temporal phase unwrapping methods which cannot be implemented in a single wrapped phase [9].

Deep learning, whose motivation lies in building and simulating the neural network of the human brain to analyze learning, is a new field in machine learning research. The earliest idea of neural networks originated from the MCP (McCulloch and Pitts) artificial neuron model in 1943 [10]. The first use of the MCP artificial neuron model for machine learning (classification) was the perceptron invented by Rosenblatt in 1958 [11]. In 1986, Hinton invented the back-propagating (BP) algorithm for multi-layer perceptron (MLP) and adopted sigmoid for nonlinear mapping, effectively solving the nonlinear classification and learning problems [12]. Since 1991, the development of deep learning has encountered bottlenecks due to the gradient disappearance problem [13]. Fortunately, Hinton proposed the solution to the gradient disappearance problem in 2006 and the rectified linear unit (ReLU) in 2010 [14,15].

In 2015, Ioffe and Szegedy proposed another solution called batch normalization (BN) [16] for the gradient disappearance problem. In recent years, deep learning has become more and more popular owing to the introduction of new network architectures (i.e., AlexNet, VGG, GoogLeNet, Inception V3, ResNet) [17–20] and the rapid development of hardware computing speed (such as graphics processing unit of AMD and NVIDIA). At the same time, deep learning has been applied to solve inverse problems in imaging science, such as super-resolution [21–24], computed tomography [25], magnetic resonance imaging [26], photoacoustic tomography [27], holography [28–34], and imaging through scattering media [35–39]. Therefore, it is a natural curiosity whether the deep learning is also suitable for phase unwrapping. In fact, as early as 2000, it has been suggested to use a supervised feedforward multilayer perceptron neural network (25 inputs, 5 hidden units, and 3 outputs) to detect phase discontinuities in optical Doppler tomography images, which is a classification-based pixel-by-pixel low-complexity phase discontinuities detection method [40]. Recently, an analogously upgraded method was proposed, in which a pixel-wise classification layer followed encoder-decoder network is used to directly classify all pixels into different wrap-count instead of pixel by pixel. However, because the classification result of the network is not accurate enough, it is necessary to use post-processing (clustering-based smoothness) to optimize it. Then, this optimized result is added to the wrapping phase to get the true phase [41]. What's more, if a pixel is incorrectly classified, the error of this pixel will be integer multiple of  $2\pi$  ( $2\pi$  at least), which is unacceptable in practical applications. Dardikman and Shaked proposed to use a neural network to establish the mapping relationship between wrapped phase and unwrapped phase, but they only demonstrated preliminary results [42].

Aiming to fundamentally tackle the phase unwrapping problem, a one-step deep learning phase unwrapping (DLPU) method and a database generation method specifically for phase-type objects are proposed in this paper. This network is trained to statistically learn the mapping relation between wrapped phase and corresponding unwrapped phase. That is, a network can be trained to learn the phase unwrapping operation. The training stage only needs to be performed once, and then the network can be used to accurately unwrap actual wrapped phase images. The network runs quickly (less than 30ms for a wrapped phase image with size of  $256 \times 256$  pixels) and automatically (without any pre-processing or post-processing). By comparing with the least square (LS, a minimum-norm method) [7] and the quality-guided (QG, a path-following method) methods [8], the DLPU method shows an excellent robustness for the wrapped phase containing noises and aliasing in varying degrees. Finally, the generalization capability of the DLPU method is demonstrated by successful unwrapping of the phase images of living mouse osteoblasts and dynamic candle flame which has not been seen by the neural network.

## 2. Materials and methods

### 2.1. Training and testing data sets

Assuming that  $\varphi(x, y)$  is the real phase image to be measured, and  $\psi(x, y)$  is the wrapped phase provided by optical systems, they have the following relationship:

$$\psi(x, y) = \text{angle}\{\exp[j\varphi(x, y)]\}, \quad (1)$$

where  $j$  is the imaginary unit;  $\text{angle}(\cdot)$  takes the argument of a complex number and effectively wraps the real phase into the range of  $(-\pi, \pi)$ . Reconstructing  $\varphi(x, y)$  from  $\psi(x, y)$  is called phase unwrapping, which suffers many challenges such as noise and aliasing.

There are two steps in the generation of the real phase. The first step is to generate an initial square matrix whose size ( $2 \times 2$  to  $25 \times 25$  for this experiment), range of values (2 to 100 for this experiment) and distributed type (one of uniformly and Gauss distributed random matrix) are all random. The second step is to amplify (by one of the nearest, bilinear and

bicubic interpolations) the matrix into a larger matrix ( $256 \times 256$  for this experiment). There are four random parameters in the real phase generation. The size of the initial matrix determines the number and position of the extreme points in the final real phase. The range of the initial matrix determines the gradient of the final real phase. The distribution type of the initial matrix and the interpolation method determine the distribution of the final real phase. Four examples of the real phase generation process are shown in Fig. 1, from which we can see different real phase distributions with different initial random square matrices.

We generated 37,500 real phase images and obtained corresponding wrapped phase images by Eq. (1), among which, 80%, 10% and 10% are used as training image set (partly shown in Supplementary Visualization 1), validation image set and testing image set, respectively. Noise with a random type (one of Gaussian, salt & pepper or multiplicative noises) and a random level (standard deviations of Gaussian and multiplicative noises from 0.01 to 0.20, or density of salt & pepper noise from 0.01 to 0.20) was then added into wrapped phase images of the training image set.

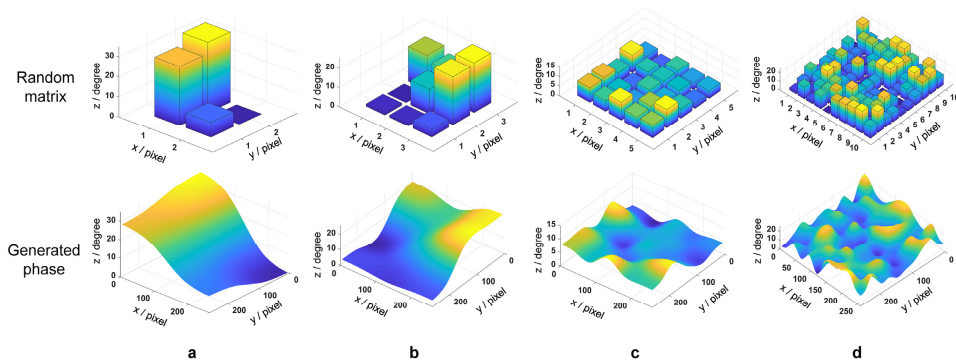


Fig. 1. Four examples of the random square matrices and the corresponding generated real phases. (a)  $2 \times 2$  matrix to real phase, (b)  $3 \times 3$  matrix to real phase, (c)  $5 \times 5$  matrix to real phase, (d)  $10 \times 10$  matrix to real phase.

## 2.2. The process of DLPU method

The proposed DLPU method has two stages, network training shown in Fig. 2(a) to establish the nonlinear mapping between  $\psi$  and  $\varphi$  and network testing shown in Fig. 2(b) to reconstruct  $\varphi(x, y)$  from  $\psi$ , directly.

For the network training, real phase images are generated firstly as the ground truth, and their wrapped versions are computed by Eq. (1) as the input, as shown in the orange part of Fig. 2(a). During training, as illustrated in the blue part of Fig. 2(a), the MSE of network output by comparing with the ground truth is minimized by iterative back-propagation using the adaptive moment estimation (ADAM) based optimization with a learning rate of  $10^{-3}$  to update the network's parameters (weights and biases) [43]. The network training only needs to be executed once. The orange and blue parts are in the opposite direction.

For network testing, as illustrated in Fig. 2(b), an unseen wrapped phase image is fed into the trained network to rapidly obtain the corresponding unwrapped phase image.

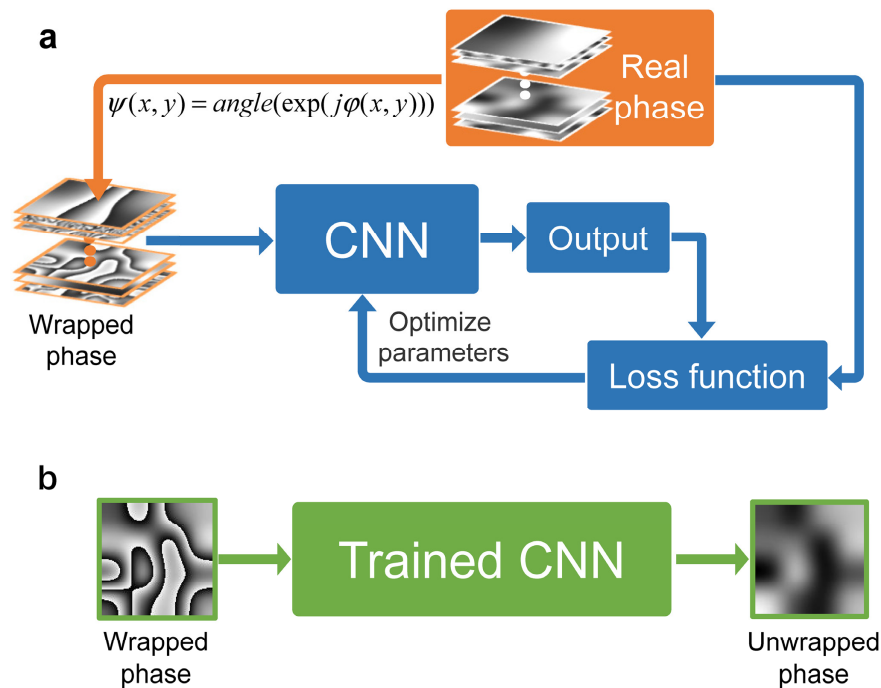


Fig. 2. Training and testing of the network. (a) Network-training flow diagram. The orange part is the data-preparation stage, and the blue part is the training stage. (b) Network-testing flow diagram. After training, the trained network is fed a wrapped phase image which is not included in the training image set and rapidly outputs the corresponding unwrapped phase image.

### 2.3. The deep neural network architecture

The convolutional neural network (CNN) architecture designed for DLPU as illustrated in Fig. 3 is inspired by U-Net [44] and residual network [20]. It consists of a contracting path in left side, an expansive path in right side and a bridge path in the middle (connecting contracting and expansive paths). The contracting path consists of five repeated uses of two  $3 \times 3$  convolution operations (each followed by a BN and a ReLU), a residual block (see Ref. 20 for details) between the two convolution operations, and a  $2 \times 2$  max pooling operation with stride 2 (for down sampling). We increase the number of feature channels by the first convolution operation in every repeat (the first from 1 channel to 8 channels with 8 convolution kernels, the rest for double channels with 2 convolution kernels). The bridge path is gained by removing the max pooling operation from contracting path. Each step in expansive path consists of a deconvolution (transposed convolution) concatenating with corresponding feature map at contracting path by skip connection, two  $3 \times 3$  convolutions followed by a BN and a ReLU, and a residual block between the two convolutions. The two convolutions in each repeat decrease the number of feature channels (the last from eight channels to one channel, the rest for halving channels). The shortcut connections of the residual blocks and the skip connections speed up the convergence of the network by transferring the gradient of the back layers to the front layers. There are 49 convolutional layers including deconvolution in the network. Contracting path can be understood as a kind of downsampling, whose purpose is to convert the information of the input space into a more abstract and high-level feature space through feature extraction of multiple convolution layers. Correspondingly, the expansive path is to convert this abstract and high-level information into an expression of the output space through multiple deconvolution layers. For

the DLPU, it is to establish the mapping relationship from the wrapped form to the unwrapped form.

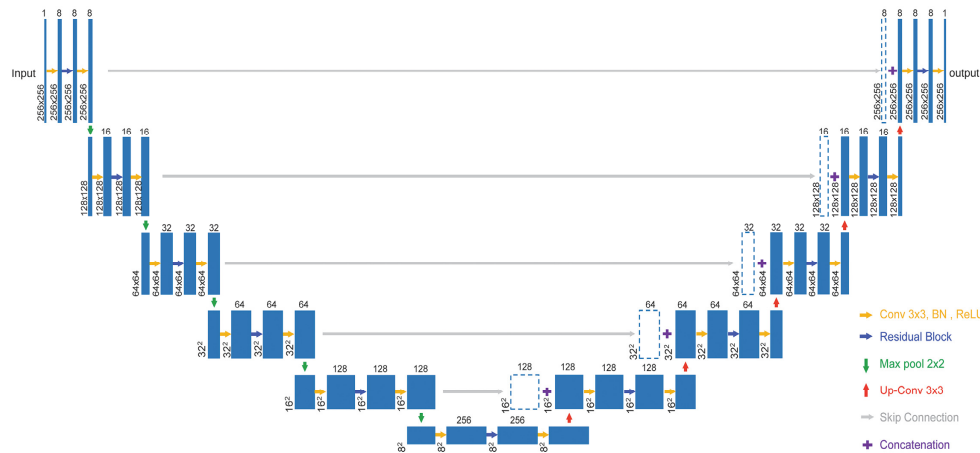


Fig. 3. Detailed schematic of the convolution artificial neural network architecture. Each blue box corresponds to a multi-channel feature map. The number of channels is provided on top of the box. The x-y-size is denoted at the lower left edge of the box. White-dotted boxes represent copied feature maps. The arrows and symbol denote the different operations.

## 2.4. Network implementation details

The CNN architecture is implemented using the TensorFlow framework version 1.1.0 based on Python 3.6.1. We performed the network training and testing on a PC with Core i7-8700K CPU (3.8 GHz) and 16 GB of RAM, using NVIDIA GeForce GTX 1080Ti GPU. The training process took  $\sim 12$ h for, e.g., 92 epochs ( $\sim 30,000$  pairs images size of  $256 \times 256$  pixels in batch size of 64). After training, the phase unwrapping time of the network (with size of  $256 \times 256$  pixels) was  $\sim 30$ ms.

## 3. Results

### 3.1. Feasibility and accuracy tests

The first step of the DLPU method is training the network to learn the statistical transformation between wrapped phase images and corresponding real phase images. To avoid overfitting of the neural network, the training was stopped when the network performance on the validation image set began to decline.

The trained network was tested by the samples included in the testing image set. The upper part of Fig. 4 shows an example of wrapped and real phase taken from the testing image set and its CNN result. Structural similarity (SSIM) [45] index of the CNN result with the ground truth is 0.991. We further compared the phase height across the horizontal directional lines (solid line for the horizontal direction, dashed line for vertical direction) indicated in real phase (red lines) and CNN result (blue lines), as illustrated in the lower part of Fig. 4. The root mean squared error (RMSE) and the maximum error between CNN result and real phase are  $0.09\pi$  and  $0.17\pi$ , respectively. The above comparisons demonstrate the feasibility and accuracy of the DLPU method.

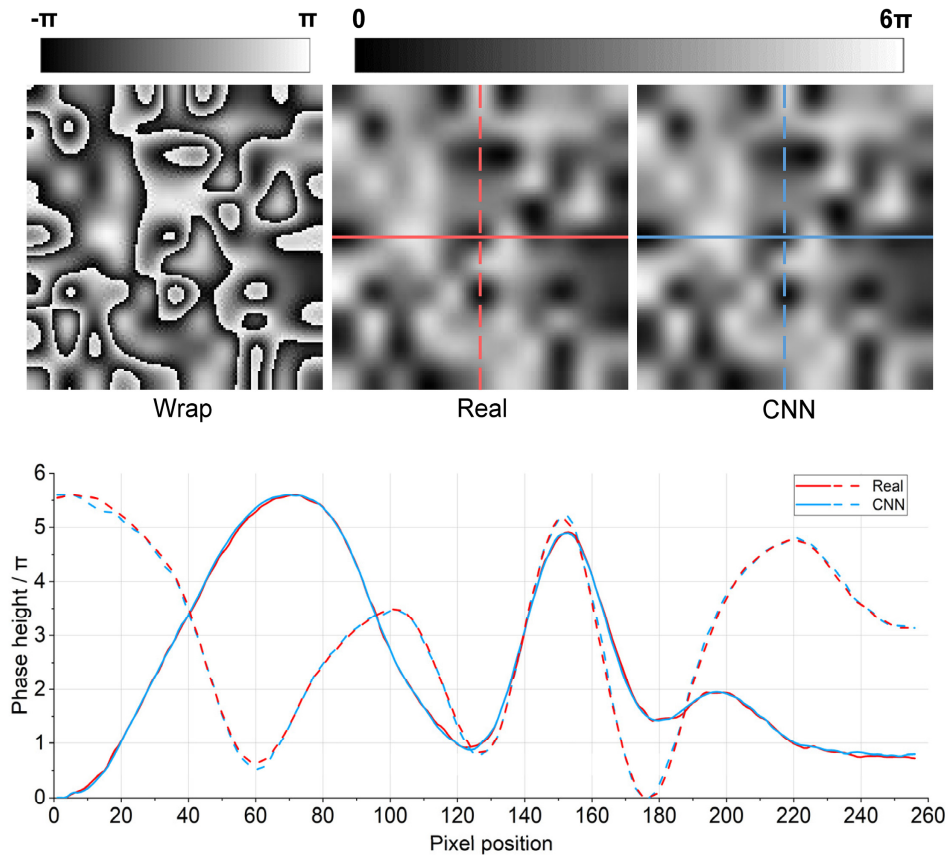


Fig. 4. CNN result of an example taken from the testing image set. Upper row: the wrapped, real and CNN-output unwrapped phase images; Lower row: the comparison of phase height across the central lines indicated in real phase (red lines) and CNN-output unwrapped phase (blue lines).

### 3.2. Anti-noise performance test

Noise inevitably presents in the wrapped phase images and thus the anti-noise performance of the DLPU is tested. A wrapped phase image is randomly selected from the testing image set. The range of the wrapped phase image is linearly mapped into  $[0, 1]$ , and then Gaussian, salt & pepper and multiplicative noises in different levels (the standard deviations of Gaussian and multiplicative noises from 0.01 to 0.40, and the density of salt & pepper noise from 0.01 to 0.40) are added together to the wrapped phase image which then is linearly mapped back to the same range as the input.

The noisy wrapped phase image was unwrapped by LS, QG and DLPU methods, respectively. As illustrated in Fig. 5, we calculated SSIM indices of the LS, QG and CNN results with the ground truth, shown by the solid lines, and Signal-to-Noise Ratio (SNR) of the noisy wrapped images, shown by dotted lines. With the increase of the noise level (SNR of the noisy wrapped images declining from 9 dB to less than  $-2$  dB), SSIM indices of the LS and QG results (the blue and cyan lines) declines from 1 to less than 0.27, but SSIM index of the CNN results (the green line) declines from 1 to 0.75 (2.78 times higher than those of the LS and QG results).

To reduce the adverse impact of noises, the usual practice is applying a denoise operation before phase unwrapping. Therefore, the wrapped phase images were denoised by window Fourier transform (WFT) before using LS and QG methods to do phase unwrapping [46]. As

illustrated in Fig. 5, SSIM indices of the WFT-LS and WFT-QG results (the purple and orange lines) decline from 1 to less than 0.53 and 0.40, respectively. Although WFT does make LS and QG results better, CNN still outperform them (in terms of SSIM index, CNN is 1.41 times higher than that of the WFT-LS results, and 1.88 times higher than that of the WFT-QG results). For noises levels less than 0.15, SSIM index of the CNN results is about 0.4% lower than that of the WFT-LS and WFT-QG results. To make it more intuitive, the error maps of the unwrapping results with noise level from 0.05 to 0.40 with an interval of 0.05 are shown in the lower part (full results in Supplementary [Visualization 2](#)). All the comparisons in Fig. 5 surprisingly demonstrate that the DLPU method does much better than the LS and QG methods, even better than WFT-LS and WFT-QG methods, in anti-noise performance.

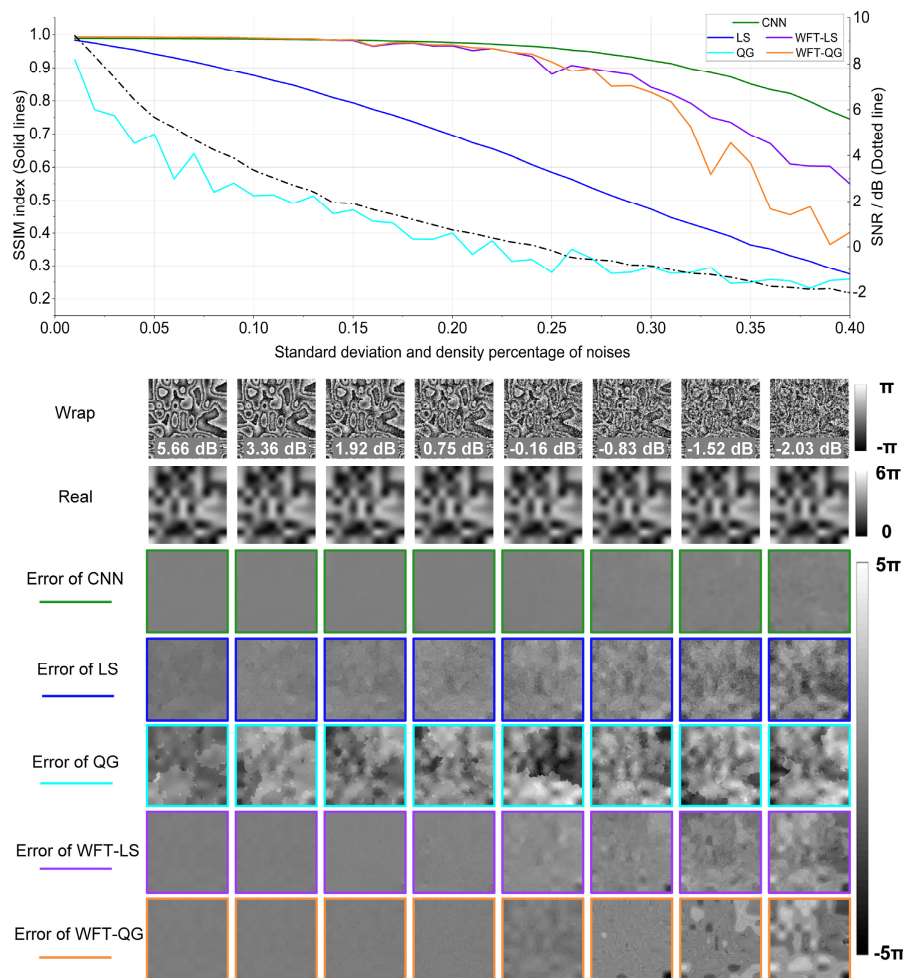


Fig. 5. SSIM indices (solid lines) of CNN, LS, QG, WFT-LS and WFT-QG results with the ground truth, SNR (dotted lines) of the noisy wrapped images while the level of the noises increasing from 0.01 to 0.40 (standard-deviations of Gaussian and multiplicative noises from 0.01 to 0.40, and density of salt & pepper noise from 0.01 to 0.40). The wrapped phase, real phase, the error maps of the CNN results (in green box), the error maps of the LS results (in blue box), the error maps of the QG results (in cyan box), the error maps of the WFT-LS results (in purple box) and the error maps of the WFT-QG results (in orange box) of the example are shown in the lower part from 0.05 to 0.40 with an interval of 0.05. Also see Supplementary [Visualization 2](#) for a detailed comparison.

### 3.3. Anti-aliasing performance test

Another often encountered problem is aliasing, which happens when the phase field distorted dramatically. To test anti-aliasing performance of the DLPU method, we randomly chose a real phase image from the testing image set, linearly mapped the height of the real phase image into the range  $[0, H]$ , where  $H$  is from 5 ( $\sim 1.6\pi$ ) to 100 ( $\sim 31.8\pi$ ) radians, and then wrapped the real phase image by Eq. (1). After that, the series of wrapped phase images were unwrapped by LS, QG and DLPU methods.

In Fig. 6, we calculated SSIM indices of the LS, QG and DLPU results, shown by the solid lines, and counted the percentage of aliasing pixels in the real phase image and wrong unwrapping pixels in the results of the three methods, shown by dotted lines. A pixel is considered aliased when the transverse or longitudinal difference value of the real phase image is greater than  $\pi$  (which is derived from sampling theorem). The unwrapping result at a pixel is considered wrong when the absolute difference between results of the three methods and ground truth is greater than  $0.20\pi$ .

As illustrated in Fig. 6, with the increase of the real phase height (i.e., degree of the aliasing is getting worse), the DLPU method performs much more robustly than the others. In detail, SSIM indices of the LS and QG results (the blue and cyan solid lines) start to decline at the phase height of 27 radians rapidly from 0.99 to 0.20, while SSIM index of the DLPU method (the green solid line) starts to decline at the phase height of 70 radians from 0.99 to 0.80 (4.00 times higher than those that of LS and QG results). Correspondingly, with increase of the aliasing pixel proportion (the magenta dotted line from 0 to 0.56), incorrect pixel proportion of the DLPU results (the green dotted line) increases from 0 to 0.20, but that of the LS and QG results (the blue and cyan dotted lines) increases from 0 to 0.71 (3.55 times higher than that of the DLPU results). More intuitively, the error maps of the example from different methods are shown in the lower part from 11 to 99 with an interval of 11 units, except for the case that the second column is the results' error maps at the phase height of 27 where the aliasing starts to appear (full results in Supplementary [Visualization 3](#)). The LS and QG methods are seen very sensitive to the aliasing points, the DLPU on the contrary, shows stronger immunity to aliasing. The WFT is not involved in this comparison as it does not help to solve the aliasing problem.

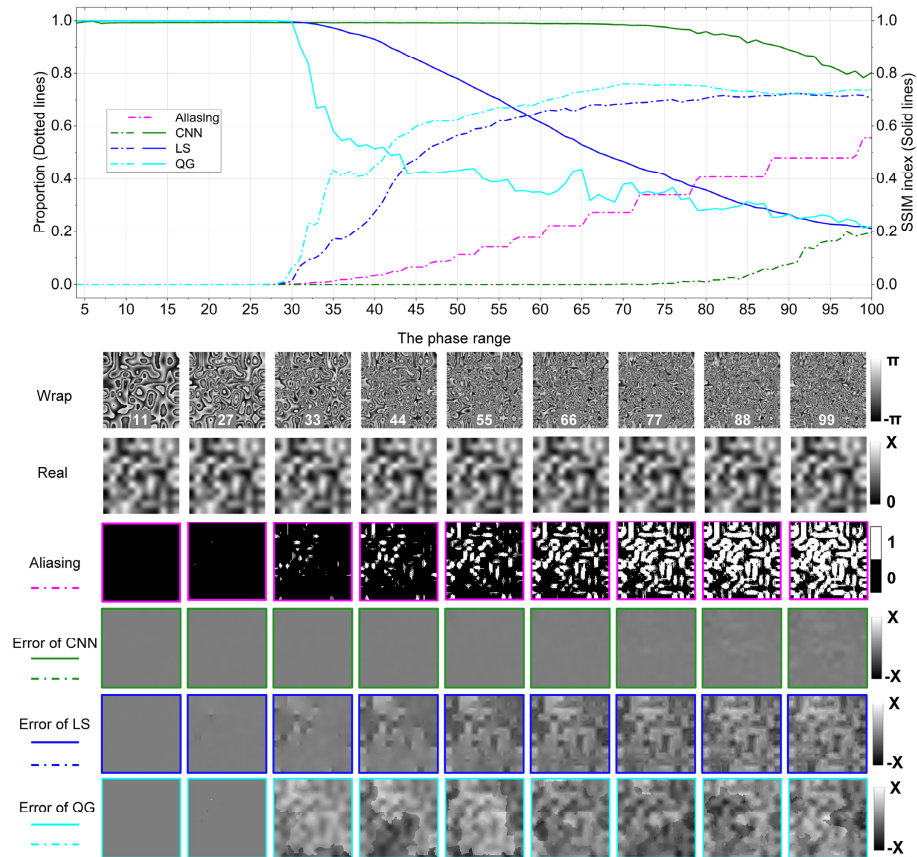


Fig. 6. SSIM indices (solid lines) of the CNN, LS and QG results with the ground truth, aliasing pixel percentage (dotted lines) of the real phase image and incorrect pixel percentage (dotted lines) of the three methods' results while the height of the real phase image increasing from 5 ( $\sim 1.6\pi$ ) to 100 ( $\sim 31.8\pi$ ). The wrapped phase, real phase, aliasing maps (in magenta box), the error maps of the CNN results (in green box), the error maps of the LS results (in blue box) and the error maps of the QG results (in cyan box) of the example are shown in the lower part from 11 to 99 with an interval of 11 units, except for the case that the second column is the results at the phase height of 27 where the aliasing starts to appear. Also see Supplementary Visualization 3 for a detailed comparison.

### 3.4. Generalization capability test

To verify the generalization capability of the DLPU method, we tested the CNN (which is trained by simulated training image set) by the phase data of living mouse osteoblasts and compared the unwrapping result with LS and QG methods. Firstly, we calculated the real phase of living mouse osteoblasts directly by the transport of intensity equation (TIE) [47] method which has been widely used in the field of optical microscopy to achieve quantitative phase microscopic imaging [48–55]. The TIE was solved by use of the fast Fourier transform algorithm. Then, the corresponding wrapped phase gained from the real phase of living mouse osteoblasts by Eq. (1) was unwrapped by the DLPU, LS and QG methods, respectively. The results of the three methods, their error maps and the aliasing maps of TIE are exemplified in Fig. 7. The DLPU method still can unwrap the wrapped phase of living mouse osteoblasts, although the phase images of living mouse osteoblasts are not included in the training image set. This comparison illustrates that the image transformation of the trained CNN not only reflects the relation between simulative wrapped and unwrapped phases, but also the statistical mapping relation between the wrapped and unwrapped phases of different

phase distributions where the generalization ability lies. Interestingly, as shown by the red arrow in Fig. 7, when the TIE images appear aliasing, the LS and QG results show a large error, but the DLPU method performs well.

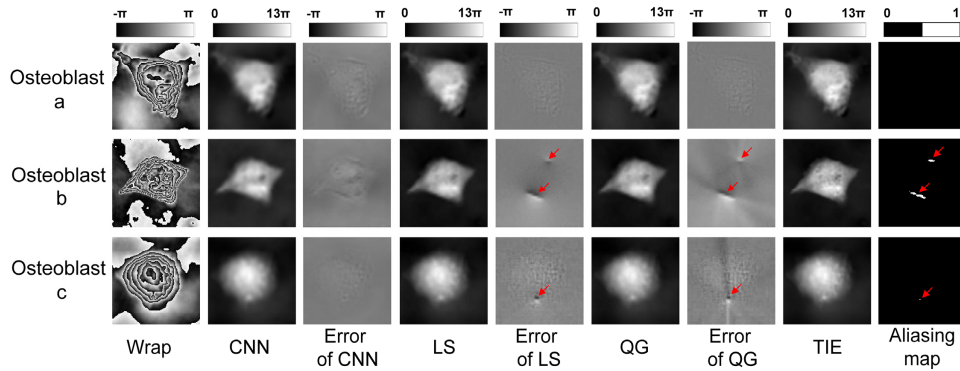


Fig. 7. Comparison of the results between CNN, LS and QG methods.

To further verify the generalization capability of the DLPU method, we unwrapped the wrapped phase of the dynamic candle flame which was obtained by off-axis digital holography with a Mach-Zehnder interferometer. During this experiment, the flame was disturbed by a fan to generate different phase distribution. The wrapped phase of dynamic candle flame, corresponding unwrapped phase reconstructed by the DLPU and LS methods in different frames within 20s are shown in Supplementary Visualization 4, and the extractions at 1st, 63th, 152th, 209th, 238th, 299th, 374th and 400th frames are shown in Fig. 8. The CNN again successfully reconstructs the unwrapped phase from these different and unseen wrapped phase images.

To verify that the anti-aliasing ability is also extended to the phase that the network has not seen. We randomly select one from the flame phase and linearly map its real phase range to 200 radians, and then wrap it with Eq. (1). The wrapped phase with the aliasing point is unwrapped using the LS, QG and DLPU methods, whose results are shown in Fig. 9.

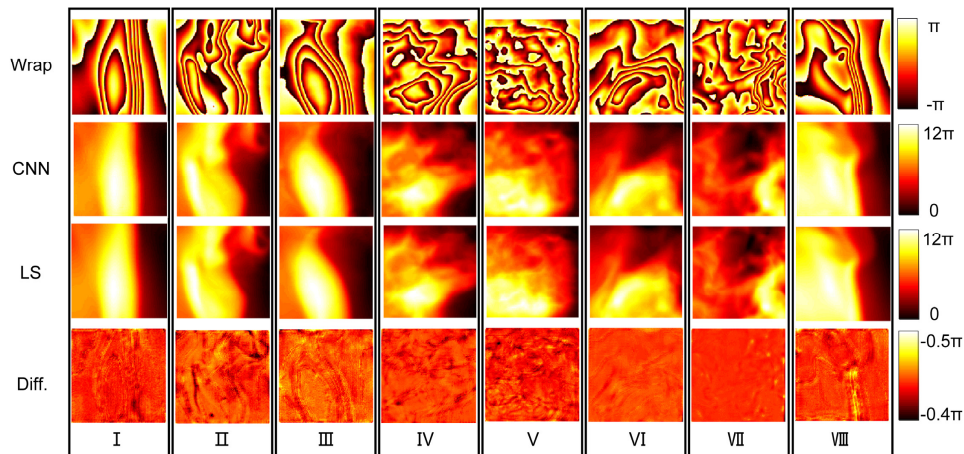


Fig. 8. Wrapped phase, corresponding unwrapped phase of dynamic candle flame reconstructed by the DLPU and LS methods, and their different maps in different frames within 20s. (I) 1th frame, (II) 63th frame, (III) 152th frame, (IV) 209th frame, (V) 238th frame, (VI) 299th frame, (VII) 374th frame, (VIII) 400th frame. Also see Supplementary Visualization 4 for full results.

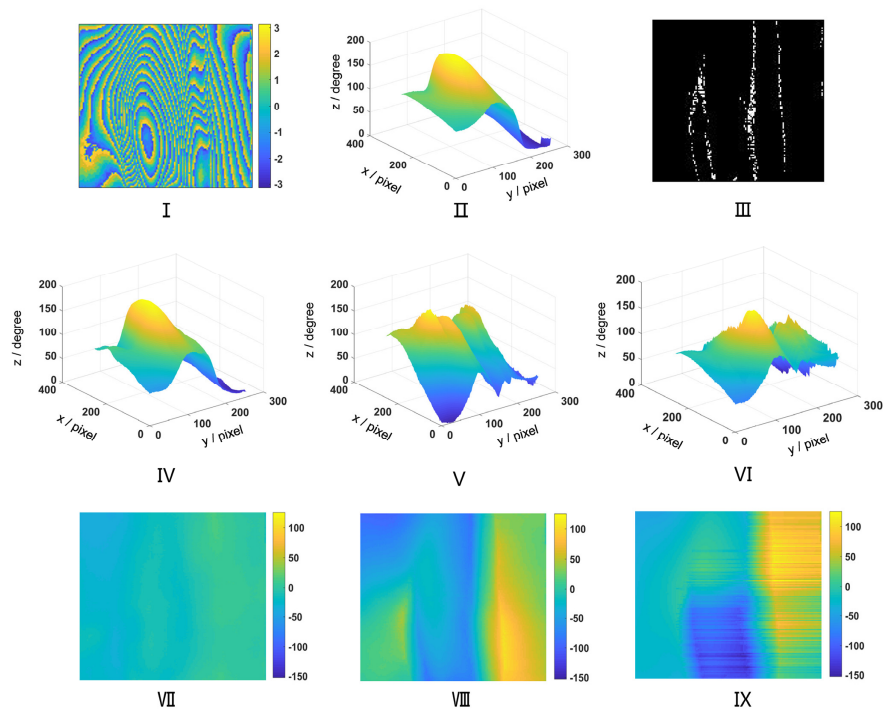


Fig. 9. Anti-aliasing ability test for a flame phase. (I) wrapped phase, (II) Real phase, (III) Aliasing map, (IV) 3D display of the DLPU result, (V) 3D display of the LS result, (VI) 3D display of the QG result, (VII) Error map of the DLPU result, (VIII) Error map of the LS result, (IX) Error map of the QG result.

#### 4. Discussions and conclusion

An accurate and appropriate training set is important for deep learning. The currently published image data sets, such as ImageNet and MNIST, only have the amplitude information and are used for image recognition, classification, understanding, etc. Hence, we propose a database generation method specifically for phase-type objects, such as cell phase and flow field phase. Four random parameters are set in the real phase generation to get a big variety of phase distribution, ensuring the generalization of the training set. More specifically, the size of the initial random matrix determines the number and position of the extreme points in the final real phase, which can be clearly seen in Fig. 1. As shown in Fig. 10(a), for the real phase with the same distribution, when the height increases, the high-frequency information of its wrapped phase also increases. As shown in Fig. 10(b), for the real phase with the same height, when the size of the matrix used to generate the real phase increases, the high-frequency information of its wrapped phase also increases. After real phase generation, we wrap the corresponding real phase by the Eq. (1), ensuring that wrapped phase of the training set is 100% correct. In training, the network learns the inverse operation of the Eq. (1) guided by training set.

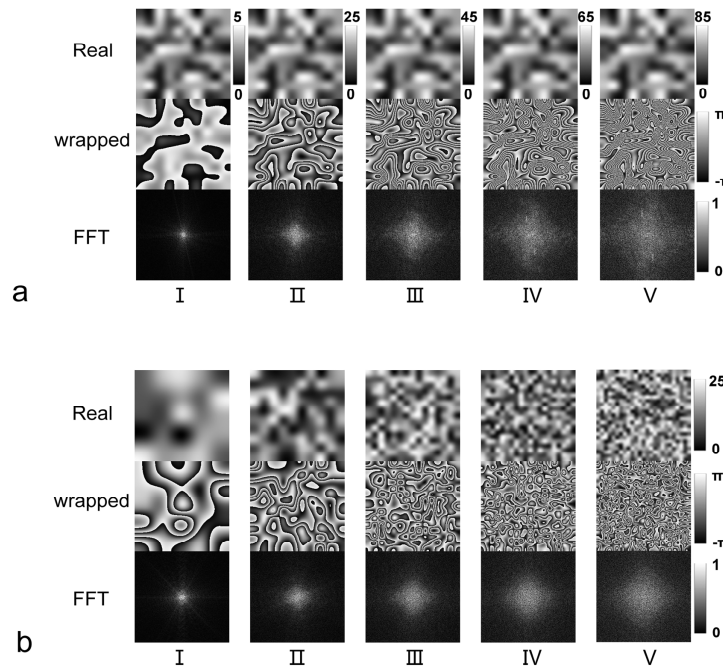


Fig. 10. Spectrum analysis of training sets. (a) Real phase, wrapped phase and corresponding Fast Fourier Transform (FFT) Spectrogram with same distribution but different height from 5 to 85 radians (I-V). (b) Real phase, wrapped phase and corresponding Fast Fourier Transform (FFT) Spectrogram with same height but different size of random square matrices in  $5 \times 5$ ,  $10 \times 10$ ,  $15 \times 15$ ,  $20 \times 20$ ,  $25 \times 25$  (I-V).

As the height range of the training set is from 0 to 70 radians, when the tested phase is beyond this range, the CNN starts to hallucinate and unwraps incorrect phase, with a gradually declining SSIM index, as illustrated in Supplementary Visualization 3 and Fig. 6. A similar behavior is also observed in Supplementary Visualization 2 and Fig. 5 for the anti-noise test. This behavior illustrates that the output images of the CNN are driven by the image transformation between wrapped phase and corresponding real phase of the trained neural network. Interestingly, SSIM index of the DLPU results maintains at a high level which is far superior to the two classical methods, even the noise or height degree has been beyond that of the training image set. In addition, we demonstrate that the anti-aliasing ability of neural networks is also feasible for a sample that has never been seen before. As shown in Fig. 9, for a flame phase, since the gradient of the real phase is too large, aliasing points occur in the wrapped phase. The DLPU, LS and QG methods are used to do unwrap and the results show that the DLPU result is much better than the other two classical methods. The quantitative SSIM indices comparisons are listed in Table 1.

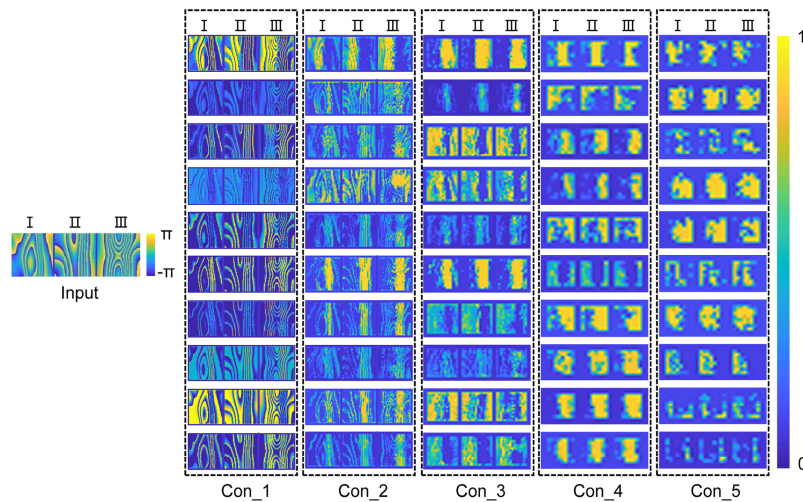


Fig. 11. Comparison of the flame wrapped phase and the two most similar phases in the training set. The five convolution layers before max pooling operation from Con\_1 to Con\_5, in which, the a is from flame wrapped phase, the b and c are from the two most similar phases in the training set.

The trained network can still unwrap the phases of the cell and candle flame which are not seen by the network during training, indicating that the network is not only pattern-matching, but also learning a generalizable model approximating the phase unwrapping operation. What the network learns from the data set is a universal prior feature information applicable for phase unwrapping. In order to find the reason why the network can unwrap the wrapped phase of unseen sample such as flame, we calculate the proportions of training set whose SSIM indices with the flame phase are higher than 0.5, shown in Table 2, and find that the similar proportion ranges from 6% to 10%. More intuitively, we pick out two wrapped phase (whose SSIM indices with the phase of the first frame flame is largest) from the training sets (shown in Fig. 8), feed these two similar phases and the phase of the first frame flame into the trained network, and visualize five middle convolution layers before the max pooling operation in Fig. 11. For a clearer display, the last four feature maps are enlarged appropriately. For these three inputs and their shallow convolution layers (Conv\_1 and Conv\_2), we can only roughly say that the distribution of the flame phase is a little similar to these two phases of the training set. However, as the convolution layer goes deeper, the receptive field of the network becomes larger, the feature space becomes more abstract, and the distribution of the convolution layers becomes more similar, further illustrating that the flame phase and the two similar phases of the training set are more similar in high-level feature space (which is the reason why the network can unwrap the wrapped phase of unseen sample such as flame). That is to say, looking to the input from the deep layer of the network, what you see is the result of filtering the input with a ‘big’ abstract convolution (whose receptive field is large). Therefore, in the perspective of two-dimensional spatial distribution, the flame phase and these two similar phases of the training set are only similar in style, but their deep convolution layers of the network have almost the same distribution. Quantitatively, the average SSIM indices of the flame phase and these two similar phases of the training set (layer by layer) are calculated in Table 3, in which, the deeper the network layer, the larger the SSIM indices.

Table 1. SSIM indices of the three methods with Real phase shown in Fig. 9.

Methods	DLP	LS	QG
SSIM	0.98	0.70	0.60

**Table 2. The proportions of training set whose SSIM indices with the flame phase shown in Fig. 8 are higher than 0.5.**

Frames	1th	63th	153th	209th	238th	299th	374th	400th
Similar proportions	8.01%	6.23%	7.21%	7.75%	9.72%	10.01%	5.71%	8.68%

**Table 3. SSIM indices of the flame phase and the two most similar phases in training set shown in Fig. 11. SSIM 1 is for the most similar phase in the training set. SSIM 2 is for the second similar phase in the training set.**

Layers	Conv _1	Conv _2	Conv _3	Conv _4	Conv _5
SSIM 1	0.28	0.39	0.62	0.77	0.79
SSIM 2	0.23	0.37	0.59	0.72	0.74

In this paper, we proposed and demonstrated a deep-learning-based method for phase unwrapping. The DLPU method has an overwhelming robustness, compared with two classical phase unwrapping methods (LS and QG) which are helpless for the condition of serious noises and aliasing. The generalization capability of DLPU method is verified by unwrapping the phase of living mouse osteoblasts and dynamic candle flame using the CNN which is trained by simulative phase-type image set. The results of this work provide compelling evidence that deep learning has much greater advantage in phase unwrapping than classical methods and can be widely applied to all the measurements and imaging techniques requiring phase unwrapping. We intend to undertake further studies in future work such as evaluating the effect of discontinuities.

## Funding

The Joint Fund of the National Natural Science Foundation of China and the China Academy of Engineering Physics NSAF (U1730137); The Fundamental Research Funds for the Central Universities (3102019ghxm018).

## References

1. R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: two-dimensional phase unwrapping," *Radio Sci.* **23**(4), 713–720 (1988).
2. D. W. Robinson, G. T. Reid, and P. de Groot, "Interferogram Analysis: Digital Fringe Pattern Measurement Techniques," *Phys. Today* **47**(8), 66 (1994).
3. D. L. Fried, "Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements," *J. Opt. Soc. Am.* **67**(3), 370–375 (1977).
4. R. H. Hudgin, "Wave-front reconstruction for compensated imaging," *J. Opt. Soc. Am.* **67**(3), 375–378 (1977).
5. S. Moon-Ho Song, S. Napel, N. J. Pelc, and G. H. Glover, "Phase unwrapping of MR phase images using Poisson equation," *IEEE Trans. Image Process.* **4**(5), 667–676 (1995).
6. M. D. Pritt and D. C. Ghiglia, *Two-dimensional phase unwrapping: theory, algorithms, and software* (Wiley, 1998).
7. M. D. Pritt and J. S. Shipman, "Least-Squares Two-Dimensional Phase Unwrapping Using Fréchet's," *IEEE Trans. Geosci. Remote Sens.* **32**(3), 706–708 (1994).
8. M. Zhao, L. Huang, Q. Zhang, X. Su, A. Asundi, and Q. Kemao, "Quality-guided phase unwrapping technique: comparison of quality maps and guiding strategies," *Appl. Opt.* **50**(33), 6214–6224 (2011).
9. J. M. Huntley and H. Saldner, "Temporal phase-unwrapping algorithm for automated interferogram analysis," *Appl. Opt.* **32**(17), 3047–3052 (1993).
10. W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in neurons activity," *Bull. Math. Biophys.* **5**(4), 115–133 (1943).
11. F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychol. Rev.* **65**(6), 386–408 (1958).
12. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature* **323**(6088), 533–536 (1986).
13. S. Hochreiter, "Untersuchungen zu dynamischen neuronalen Netzen," Diploma, Technische Universität München **91**(1) (1991).
14. G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.* **18**(7), 1527–1554 (2006).
15. V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10, 2010)* pp. 807–814.

16. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," <https://arxiv.org/abs/1502.03167>.
17. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2012), pp. 1097–1105.
18. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (IEEE, 2015), pp. 1–9.
19. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (IEEE, 2016), pp. 2818–2826.
20. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (IEEE, 2016), pp. 770–778.
21. C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016).
22. W. Shi, J. Caballero, F. Huszar, 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.
23. J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (IEEE, 2016), pp. 1646–1654.
24. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica* **4**(11), 1437–1443 (2017).
25. M. T. McCann, E. Froustey, M. Unser, M. Unser, and M. Unser; Kyong Hwan Jin, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.* **26**(9), 4509–4522 (2017).
26. S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang, "Accelerating magnetic resonance imaging via deep learning," in *Proceedings of IEEE International Symposium on Biomedical Imaging* (IEEE, 2016) pp. 514–517.
27. S. Antholzer, M. Haltmeier, and J. Schwab, "Deep learning for photoacoustic tomography from sparse data," *Inverse Probl. Sci. Eng.* **27**(7), 987–1005 (2018).
28. Y. Rivenson, Y. Zhang, H. Günaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light Sci. Appl.* **7**(2), 17141 (2018).
29. H. Wang, M. Lyu, and G. Situ, "eHoloNet: a learning-based end-to-end approach for in-line digital holographic reconstruction," *Opt. Express* **26**(18), 22603–22614 (2018).
30. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica* **4**(9), 1117–1125 (2017).
31. T. Pitkäaho, A. Manninen, and T. J. Naughton, "Focus prediction in digital holographic microscopy using deep convolutional neural networks," *Appl. Opt.* **58**(5), A202–A208 (2019).
32. Z. Ren, Z. Xu, and E. Y. Lam, "Learning-based nonparametric autofocusing for digital holography," *Optica* **5**(4), 337–344 (2018).
33. T. Shimobaba, T. Takahashi, Y. Yamamoto, Y. Endo, A. Shiraki, T. Nishitsuji, N. Hoshikawa, T. Kakue, and T. Ito, "Digital holographic particle volume reconstruction using a deep neural network," *Appl. Opt.* **58**(8), 1900–1906 (2019).
34. X. Yuan and Y. Pu, "Parallel lensless compressive imaging via deep convolutional neural networks," *Opt. Express* **26**(2), 1962–1977 (2018).
35. R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," *Opt. Express* **24**(13), 13738–13743 (2016).
36. M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen, and G. Situ, "Deep-learning-based ghost imaging," *Sci. Rep.* **7**(1), 17865 (2017).
37. Y. Sun, Z. Xia, and U. S. Kamilov, "Efficient and accurate inversion of multiple scattering with deep learning," *Opt. Express* **26**(11), 14678–14688 (2018).
38. P. Wang and J. Di, "Deep learning-based object classification through multimode fiber via a CNN-architecture SpeckleNet," *Appl. Opt.* **57**(28), 8258–8263 (2018).
39. S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," *Optica* **5**(7), 803–813 (2018).
40. W. Schwartzkopf, T. E. Milner, J. Ghosh, B. L. Evans, and A. C. Bovik, "Two-Dimensional Phase Unwrapping Using Neural Networks," in *Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation* (IEEE, 2000) pp. 274–277.
41. G. E. Spoorthi, S. Gorthi, and R. K. S. S. Gorthi, "PhaseNet: A Deep Convolutional Neural Network for Two-Dimensional Phase Unwrapping," *IEEE Signal Process. Lett.* **26**(1), 54–58 (2019).
42. G. Dardikman and N. T. Shaked, "Phase Unwrapping Using Residual Neural Networks," in *Imaging and Applied Optics, OSA Technical Digest* (Optical Society of America, 2018), paper CW3B.5.
43. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," <https://arxiv.org/abs/1412.6980>.

44. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer, 2015) pp. 234–241.
45. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
46. Q. Kemao, "Two-dimensional windowed Fourier transform for fringe pattern analysis: Principles, applications and implementations," *Opt. Lasers Eng.* **45**(2), 304–317 (2007).
47. M. R. Teague, "Deterministic phase retrieval: a Green's function solution," *J. Opt. Soc. Am.* **73**(11), 1434–1441 (1983).
48. Y. Li, J. Di, C. Ma, J. Zhang, J. Zhong, K. Wang, T. Xi, and J. Zhao, "Quantitative phase microscopy for cellular dynamics based on transport of intensity equation," *Opt. Express* **26**(1), 586–593 (2018).
49. A. Barty, K. A. Nugent, D. Paganin, and A. Roberts, "Quantitative optical phase microscopy," *Opt. Lett.* **23**(11), 817–819 (1998).
50. L. Waller, Y. Luo, S. Y. Yang, and G. Barbastathis, "Transport of intensity phase imaging in a volume holographic microscope," *Opt. Lett.* **35**(17), 2961–2963 (2010).
51. C. Zuo, Q. Chen, W. Qu, and A. Asundi, "High-speed transport-of-intensity phase microscopy with an electrically tunable lens," *Opt. Express* **21**(20), 24060–24075 (2013).
52. C. Zuo, Q. Chen, Y. Yu, and A. Asundi, "Transport-of-intensity phase imaging using Savitzky-Golay differentiation filter—theory and applications," *Opt. Express* **21**(5), 5346–5362 (2013).
53. W. Yu, X. Tian, X. He, X. Song, L. Xue, C. Liu, and S. Wang, "Real time quantitative phase microscopy based on single-shot transport of intensity equation (ssTIE) method," *Appl. Phys. Lett.* **109**(7), 071112 (2016).
54. C. Zuo, Q. Chen, W. Qu, and A. Asundi, "Noninterferometric single-shot quantitative phase microscopy," *Opt. Lett.* **38**(18), 3538–3541 (2013).
55. Z. Jingshan, R. A. Claus, J. Dauwels, L. Tian, and L. Waller, "Transport of Intensity phase imaging by intensity spectrum fitting of exponentially spaced defocus planes," *Opt. Express* **22**(9), 10661–10674 (2014).