Imaging Through Turbid Media With Vague Concentrations Based on Cosine Similarity and Convolutional Neural Network
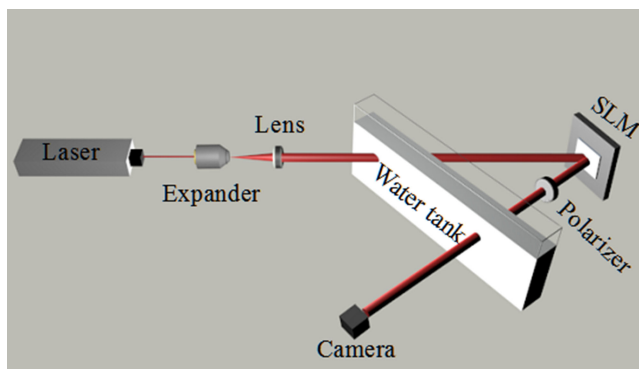
Lina Zhou
Yin Xiao
Wen Chen

Imaging Through Turbid Media With Vague Concentrations Based on Cosine Similarity and Convolutional Neural Network

Lina Zhou [ID],[1] **Yin Xiao** [ID],[1] **and Wen Chen** [ID] [1,2]

[1]Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, China
[2]The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, China

Abstract: Underwater imaging has been extensively studied to bypass the limitation aroused by scattering and absorption of water solutions. It is highly meaningful to the development of optical imaging, especially in turbid media. The existing methods developed for reconstruction of original images from speckle patterns are applied in a stable medium, which obstruct wider applications in unpredictable media. Hence, it is crucial to take changeable environments into consideration to circumvent the limits of the extant methods. In this paper, we propose a new approach based on cosine similarity for speckle classification and convolutional neural network (CNN) for the reconstruction. The targets are placed in variant densities of turbid water mixed with certain of milk, and their corresponding intensity speckle patterns are recorded by a camera. It is verified that utilization of cosine similarity for the classification of patterns recorded in changeable media ensures high fidelity for label predictions. For a speckle pattern obtained in totally unidentified media, it can make a prediction of the density which has a high probability of accuracy. We can exploit the classified density to automatically select the most appropriate datasets to train a CNN model and then make predictions in real time with the trained CNN model. The combined model presented in this paper is tolerant to the uncertainty of turbidity. Moreover, it guarantees high-accuracy pattern classification and high-quality image reconstruction. It is feasible for potential applications in harsh water solutions with unknown perturbations of concentrations.

Index Terms: Image reconstruction, underwater imaging, turbid media, cosine similarity, deep learning.

1. Introduction

Studies on optical imaging have evolved rapidly from macroscopic imaging to microscopic imaging to keep up with further exploration of biological applications [1], [2]. Though numerous novel imaging algorithms and leading equipments keep springing up, the development of optical imaging techniques is highly restricted by their applications in simple environments [3]–[9]. Consequently, seeing through complex media is one of the hottest and trickiest topics in optical imaging [9].

So far, there has been abundant interest in probing practical methods to develop optical vision. Meanwhile, much research work has been implemented to reconstruct aboriginal images from noise-induced speckle patterns recorded under various environments. (i) For simple environments, the most frequently applied method is enhancing its contrast by means of preprocessing of lighting conditions, such as improvement of light source and increase of exposure time [10]. It works well in relatively uncomplicated environments but has a great tendency to show apparent weakness in more complex media such as murky water and cloudy conditions. (ii) Some research work has also been conducted to study the intrinsic relationship between speckle patterns and original images, which is known as transmission matrix (TM) [11]. With a great deal of pairs of speckles and original images collected, it may provide a precise representation of TM. Nevertheless, this approach exists an unavoidable constraint that it is extremely sensible to slight perturbations resulting in entirely different TMs. (iii) Some mathematical and theoretical algorithms based imaging techniques have been proposed for post-processing to recover clear images of the target placed in complex media [12], [13]. It is proven that holographic imaging technique can reconstruct objects from the holograms recorded under inhomogeneous environments. However, the background noise effect aroused by the complex media dramatically degrades the image quality of the reconstruction. The main reason for this is that the holographic reconstruction method strongly depends on the recorded interference pattern. (iv) Machine learning based technologies continue to flourish which open up a new era for image retrieval and provide an alternative technique to acquire satisfactory reconstructions under some simple environments, including: diffusers and slabs [14]–[19]. With pairs of speckle patterns as input and original images as output sent to a machine learning neural network, the well-trained network can successfully predict speckle patterns recorded in the same environment. Even so, the innovative approach has not been adequately utilized in situations of fluctuations of a specific factor. For example, a neural network is previously trained under a specific concentration of an aqueous solution, but it is out of service to predict the object placed in an absolutely unknown concentration of the same solution. In summary, the four groups of methods are narrowed by small ranges of applications either in simple conditions or in a known harsh environment.

The aforementioned environments in deep learning methods are simple since that they mainly put diffusers or slabs behind the objects to scatter transmission light [16], [19]. Even more, the scattering media are relatively stable compared to practical complex environments (e.g., water [20] and smoke [21]) which are varied by turbidity. A typical example of the complex environment to be discussed is turbid water which forms a new topic called underwater imaging. The biggest challenge in imaging through turbid media is the non-uniformity of such media which seriously interferes with regular propagation of light [22]–[24]. Accordingly, the light-carried valuable information cannot be precisely delivered leading to unrecognizable speckle patterns. Such external factor-induced artifacts would bring about irreparable damages including reduction of image quality and absolute breakage of the object images. In consideration of visualizing objects concealed behind an opaque medium, it is necessary to investigate applicable methods for extracting serviceable information from turbid media. Recently, some encouraging research work have been done to implement image restoration in sea tripods [25], depth estimation in clear water [26] and image classification in weakly scattered water [27]. Nevertheless, the camera recorded images in these water solutions are recognizable so that those water solutions cannot be strictly deemed as highly turbid water. Even more, these solutions are static water with steady concentrations which are not applicable in variant water conditions. Hence, currently prevalent techniques still have a long way to go to seek for a more universal scheme for a complex environment with vague factors.

In this paper, the objective of our work is to abstain from the restrictions of narrow scopes of applications in practice by implementing imaging through turbid water (mixed with milk) based on the techniques of cosine similarity and CNN. Cosine similarity is a conventional probabilistic approach used to measure the similarity between two non-zero vectors [28]. Correspondingly, it is applicable to estimate the interrelation between two images, and then it can be used for image classification. Related work for image recognition has been reported but limited to completely irrelevant classification results which severely hamper subsequent applications for image reconstruction. CNN is one of the deep learning (DL) frameworks combined with a processing of convolution calculation which
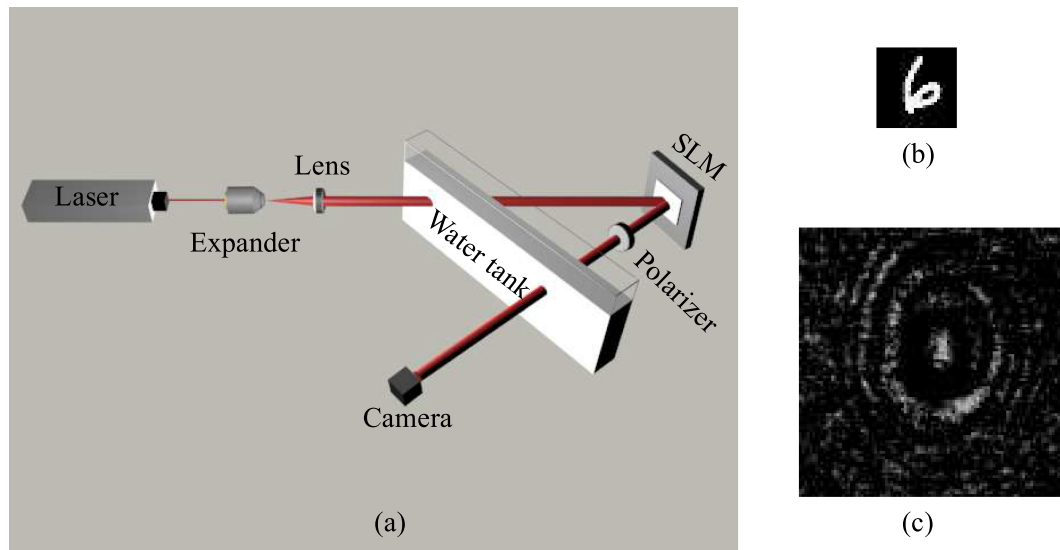
Fig. 1. (a) Experimental setup for imaging through turbid water. An example of the input digit transported to SLM and the camera captured speckle pattern is presented in (b) and (c), respectively.

can abstract senior representations from the given data [29], [30]. For the tested image recorded in an unknown concentration of milk solution, we can sort the tested image into a certain group of concentration of water solution based on cosine similarity. By means of the labeled density, the most appropriate datasets can be validly selected to train a CNN model. Finally, the unknown objects related to the tested patterns can be effectively reconstructed by using the trained CNN model. It is verified that the combined model is highly acceptable to wrong classifications and is robust to variant concentrations of turbid media.

2. Experimental Demonstration

To demonstrate capability of the proposed method, an experimental validation of imaging through inhomogeneous media is depicted in Fig. 1(a). The turbid medium tested in our experimental setup is a mixture emulsion of pure water and milk. A laser beam emitted from a He-Ne laser source (Newport, R-30993, 633 nm) is first expanded by a microscope objective (Newport, M-40X, 0.65 NA) and then collimated by a collimating lens ($f = 50$ mm). Subsequently, the expanded and collimated light transmits through a water tank illuminating on an intensity spatial light modulator (SLM) (Holoeye, LC-R 720, reflective). The transparent water tank (100 mm in length, 50 mm in width and 300 mm in height) made by polymethyl methacrylate (PMMA) contains a certain amount of milk and pure water. The SLM is embedded with object images. The input target images sequentially displayed on SLM are handwritten digits from MNIST database being widely used for demonstration in machine learning [31]. Image size of the input digits is 28×28 pixels. The reflected wave passes through the murky medium again and then the strongly scattered light is captured by a camera (Thorlabs, DCC3240M, 1280×1024 pixels and pixel size of 5.30 μm $\times$ 5.30 μm). The camera is placed 70 mm away from the water tank. It is worth noting that the laser beam passing through the water tank becomes a diffused beam and then further passes through the water tank again to illuminate on the camera. In conventional experimental setups [13], [14], [16], [17], light illuminated on the SLM is directly emitted from the laser without any barriers. In comparison with conventional experiments, the situation in our experiments can be viewed as objects placed in the water. By means of the advanced programming techniques such as LabVIEW, it is controllable to synchronously embed the target images to SLM and record the corresponding output images on a camera. To verify the reliability and validity of our methods, room temperature for the whole

experimental environments is set as 18 °C. An example of the handwritten digit input to SLM is shown in Fig. 1(b), and its diffraction pattern recorded by the camera is shown in Fig. 1(c). To reduce the computational load without significantly decreasing valuable information, the captured pattern is cropped to be of 100×100 pixels.

3. Methodology

To obtain better performance from imaging through murky water with vague concentrations, it is essential that the specific concentration of the turbid water can be exactly confirmed. And then a suitable CNN model can be well characterized and developed for image reconstruction in the extreme environment. The main problem is that there are limited technologies to ascertain the explicit density of the unstable solution leading to failed model chosen for subsequent processing of image recovery. Motivation of this paper derives from the objective reality mentioned above. Our research work aims to avoid high requirements for precision measurement techniques and accurate predictions of densities. In this paper, we describe a cosine similarity based method to make predictions of the concentration and a CNN-based method to directly reconstruct the ground truth of the diffraction pattern recorded in an unknown density of murky water.

3.1 Cosine Similarity for Pattern Classification

For speckle patterns captured in variable environments, the existing methods face a limited capacity to react to perturbations aroused by uncertain factors, e.g., concentration or the lighting time. Especially for the situations where there are only slight disturbances, it is fatal to the current theoretical and mathematical algorithms, since those small changes may result in huge computational load and wrong predictions. These problems roused by undetermined interferences need to be tackled urgently on account of the constraints of successive applications for image reconstruction in reality.

At the beginning, our general idea is to seek for a universal approach for speckle classification which has advantages of high robustness and great tolerance to error predictions. The proposed method is based on cosine similarity to classify the pattern into a certain category. Given a totally unknown speckle recorded in turbid environments, cosine similarity can be implemented to determine an optimal group that the pattern may belong to. Cosine similarity is a straightforward technique that directly measures the similarities between the speckles. Detailed statements of the principles for cosine similarity [23] based speckle classification is given as follows.

The intrinsic theory originates from the Euclidean dot product formula which is defined as

$$\vec{a} \bullet \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos \theta, \tag{1}$$

where $\vec{a}$ and $\vec{b}$ are non-zero vectors, and $\cos \theta$ is cosine value of two vectors. Then, cosine similarity is described by the transformation of the Euclidean dot product formula as

$$\begin{aligned} cosine\ similarity = \cos \theta &= \frac{\vec{a} \bullet \vec{b}}{\|\vec{a}\| \left\|\vec{b}\right\|} \\ &= \frac{\sum_{i=1}^{n} (a_i * b_i)}{\sqrt{\sum_{i=1}^{n} a_i^2} \sqrt{\sum_{i=1}^{n} b_i^2}}, \end{aligned} \tag{2}$$

where a_i and b_i denote the i_{th} element of $\vec{a}$ and $\vec{b}$, respectively. It is derived from the essential attribute of cosine function that the value of cosine similarity ranges from -1 to 1. Here, '-1' means that the two vectors are exactly in the opposite direction, and correspondingly '1' denotes that they are in the same direction. '0' stands for that they are mutually independently from each other. Except all the special values mentioned above, other values are used to indicate the similarity. The higher the value of cosine similarity is, the higher level of similarity will be.

Based on the property of cosine similarity, it is feasible to apply this theory to image classification. Then, the next step to be taken is to explore appropriate representations for the images. Instinctively,
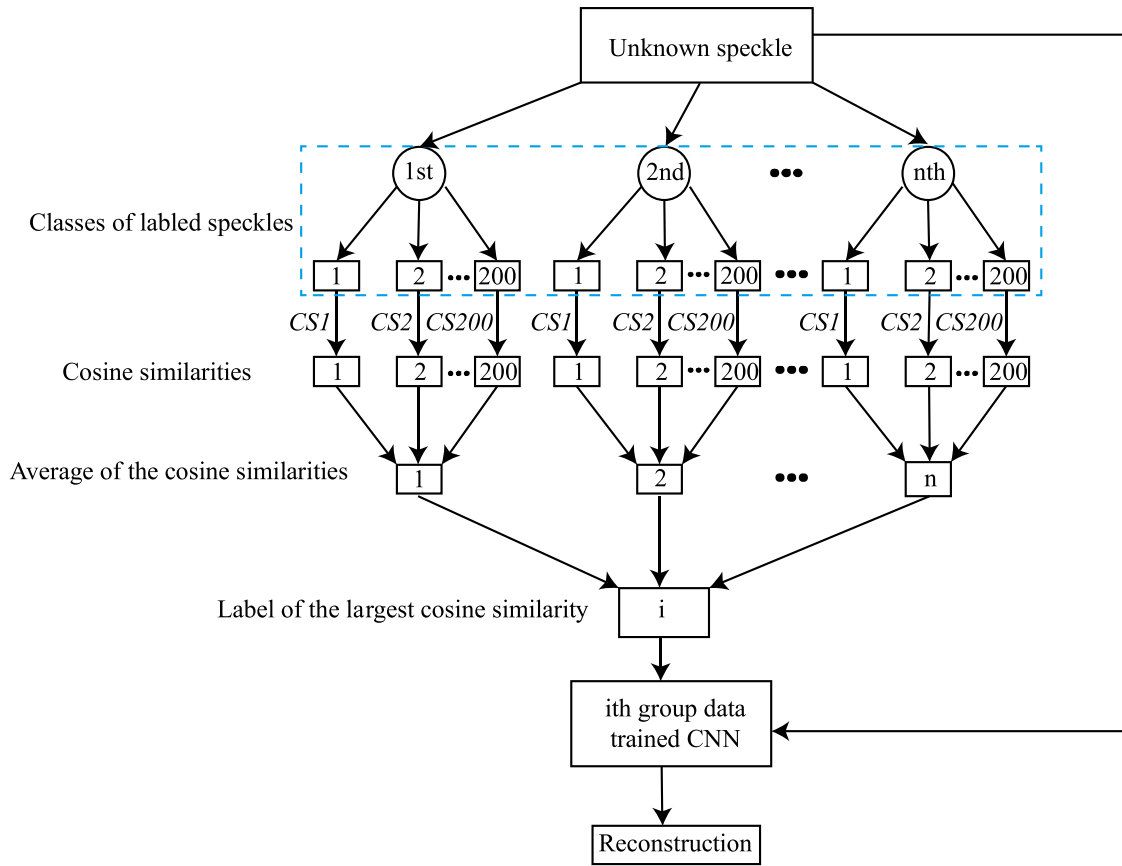
Fig. 2. Flow chart of cosine similarity based image classification for the reconstruction.

the main difference between two images is the brightness distribution which is mathematically defined as histogram. For each image, its histogram is distinguishable from any other histograms. From this point, the histogram is a perfect substitute for an image in the proposed method. Using the histograms as vectors, the similarity can be achieved by calculating the cosine similarity. Through the calculated coefficient of cosine similarity, we can judge the resemblance of the two images. Given a speckle captured in an unknown condition and patterns recorded under certain groups of known environments, comparisons of the cosine similarities can be made to pick up the largest value of cosine similarity. Here, the largest cosine similarity means that the tested speckle situated environment is close to current environment. With the methodology described above, a speckle recorded under unknown conditions can be classified into an optimal environment.

A detailed flow chart of cosine similarity for image classification in our case is shown in Fig. 2. Given a speckle recorded in an unknown density of turbid water and *n* classes of labeled speckles captured in certain specific densities of murky water, we can obtain 200 coefficients of cosine similarities and the mean value of cosine similarity in each typical environment. By comparing the *n* groups of average values of cosine similarities, the label (*i*) of the largest cosine similarity is achieved. Sequentially, the environment where the unknown speckle situated in is predicted to be close to the *ith* condition. Hence, the density of the unidentified speckle can be sorted into the most similar density of turbid media. The *ith* group of speckle patterns and their corresponding original images are used as the training dataset for CNN structure, and the given speckle can be reconstructed by using the trained CNN.

The methodology described here provides a reliable technique for image classification which will benefit the subsequent step for image recovery. Section 3.2 further introduces CNN architecture for the reconstruction.
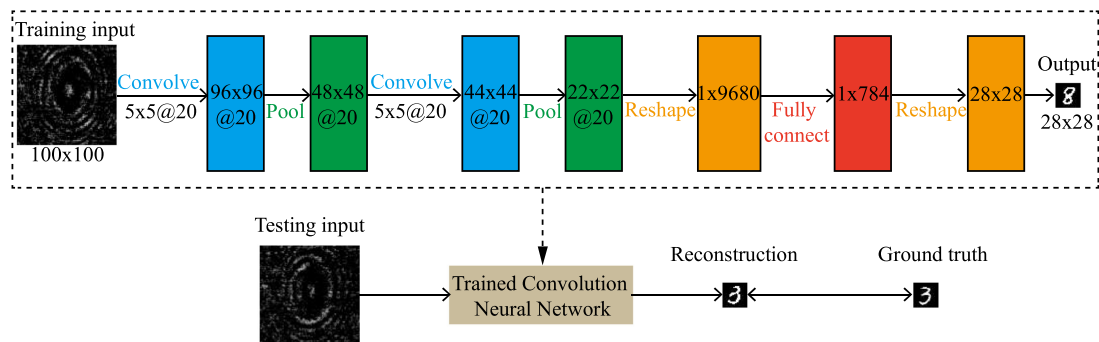
Fig. 3. A detailed schematic of our CNN architecture. Size of the input image is 100 × 100 which is cropped by using a 100 × 100 window from the camera captured image. Blue and green colors represent convolution processes and pooling processes, respectively. The orange color denotes the process of reshaping. The red block is the fully-connected layer obtained by processing of fully connecting. The output is final prediction of the input. The trained CNN structure can be used to recover unknown speckles in real time.

3.2 CNN Architecture for the Reconstruction

Machine learning based method usually consists of two phases including training neural networks with a given dataset and then predicting an unknown datum with the trained networks. Schematic of the proposed CNN structure is shown in Fig. 3. At first, the input images are downsized to 100 × 100 images which contain adequate information for our experiments. During the stage of training, the input image convolves with 20 kernels (size of 5 × 5) forming the first convolution layer. The activation function used in the convolution layer is sigmoid function. Then the first convolution layer (size of 96 × 96 × 20) is downsampled to 48 × 48 × 20 which generates the first pooling layer. Processing of pooling is an important step to be taken to reduce computational load. After that, the first pooling layer continues to convolve with 20 kernels which are of the same size with that used in the first convolution layer. The second convolution layer is of size 44 × 44 × 20 followed by the second pooling layer (size of 22 × 22 × 20). After two rounds of convolution and pooling processing, the second pooling layer is reshaped to a 1 × 9680 vector. To better build the relative relation between the input image (100 × 100) and ground truth (28 × 28), it is necessary to fully connect the reshaped vector to a vector of size 1 × 784. Then the second processing of transformation is taken to reshape the fully connected layer (1 × 784) to a 28 × 28 image which stands for the estimation of the input image. With pairs of input images and corresponding ground truths sent to the designed CNN model, the CNN structure is considered to learn underlying parameters of a specific environment. Although it cannot provide exact representations of the parameters nor their explicit values, the trained CNN model is capable to recover the object image from the raw intensity speckle pattern. The CNN architecture uses Matlab2009 on a PC with Nvidia Geforce GTX1080Ti GPU to guarantee normal operation of our CNN model.

It is worth mentioning that the input pattern fed to the CNN structure is pre-processed by removing its DC component through subtracting the mean value. To evaluate quality of the reconstructed image, mean squared error (MSE) and peak signal-to-noise ratio (PSNR) are used. The MSE and PSNR are respectively described by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2, \tag{3}$$

$$PSNR = 20\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right), \tag{4}$$

where n (784) denotes the whole pixel number of the recovered image which is the same with that of the ground truth, Y_i and $\hat{Y}_i$ respectively denote the pixel value of the ground truth and the predicted

image, and MAX_I represents the maximum pixel value of the image. For the grayscale images used in our experiments, MAX_I is defaulted to be 255. The MSE is average of all the squares of the differences between the predicted image and the original image, and PSNR is further defined based on MSE. To update the parameters (including weights and bias) of the CNN model, stochastic gradient descent (SGD) is applied as an optimization function to minimize the MSE values. The updating rule of SGD is defined as

$$\theta = \theta - v, \tag{5}$$

$$v = m \times v + \alpha \times d\theta, \tag{6}$$

where θ denotes the weights and bias which are updated by velocity v. The velocity v is initiated to 0 and is renewed by the fixed momentum (-9.5×10^{-5}) and the updated learning rate (initialized from 10^{-6} and doubled by every 200 speckles sent to the network). The weight is initialized to 10^{-3} for the processing of first convolution layer, to be 10^3 for the second convolution layer and to be 10^{-3} for the fully connected layer. After 5 epochs of iterations, the MSE decreases gradually and stabilizes at an optimal order of magnitude. Finally, the CNN model is trained to learn the relationship between the input speckles and ground truths, and then it can be applied to make predictions of unknown speckles in real time.

4. Experimental Results and Discussions

For speckles captured under turbid media, it is difficult for the existing methods to abstract the explicit representations of the relationship between the speckles and ground truths. At this point, machine learning based method provides a substitute to avoid tedious computation of theoretical expressions. However, a precondition for success of the CNN based method is that the speckle pattern to be reconstructed should situate in the same environment with that of the dataset used to train the CNN model. In practice, complex environments are varied in various factors which may obstruct the development of these learning-based methods.

In our experiments, the turbid media are varied in the densities. At the beginning, the murky medium is made by mixing 1200 ml of water with 50 ml and 60 ml of milk. The camera is placed 70 mm away from the water tank and 180 mm away from the SLM. 2000 handwritten digits from the MNIST database are sent to the SLM. Consequently, 2000 diffraction patterns are separately recorded under two different densities of the turbid media, where 1900 speckle patterns and their corresponding ground truths are used as the training dataset and the left 100 are used to test the performance of the trained CNN model. With sophisticated equipments to precisely measure the density of the turbid media, speckles recorded in this condition can be clearly reconstructed using the CNN model trained in the same environment. An example is given in Fig. 4. The speckle pattern in the first block is recorded under the turbid medium with 50 ml of milk and reconstructed by two CNN models. The upper CNN model is trained in the same environment and the lower is trained in a different condition with 60 ml of milk mixed with the same amount of water. The second block shows the pattern recorded under the turbid medium with 60 ml of milk and reconstructions made by two different CNN models. It is obvious that quality of the reconstruction recovered by a CNN model trained under a totally different condition is poor in comparison with that of the recovery given by the CNN model trained under the same environment. Explanation for the poor-quality and wrong reconstruction can be ascribed to deficiency of the precision measurement techniques and lack of generality of the proposed CNN. To avoid the influence aroused by densities, an attempt to combine the training datasets obtained at all concentrations to train a big CNN model is done but failed to reconstruct the speckles in these concentration ranges. Reason is that these training datasets with different densities are quite different so that it is not possible to build only one CNN model for all the densities. Here, it is experimentally verified that our CNN model is robust to multi-range of turbidity and can be manually grouped into certain classes. Each CNN model is trained by certain groups of datasets since that the diffracted patterns recorded under nearby densities share some similarities.
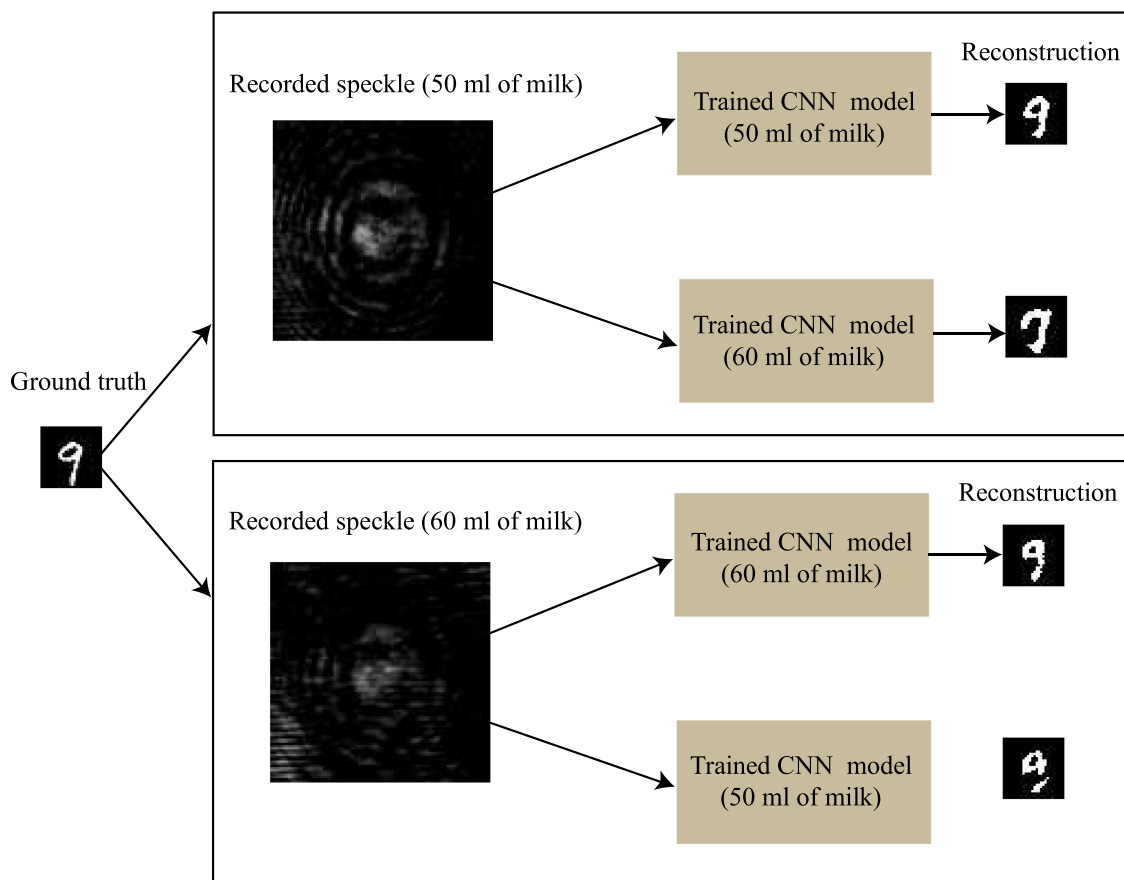
Fig. 4. An example of speckle reconstructed by the CNN model trained in the same environment and that trained in a different circumstance.

Without precision techniques to measuring accurate concentration, it is possible to recover the speckles by using one of the CNN models but it is more or less at the cost of image quality.

The requirement of advanced instruments to precisely measure the variables greatly restricted the development of machine learning based method for image reconstruction. It is a big challenge to meet the demand for high-level image quality. Hence, research on an alternative strategy to avoid the limit mentioned above is an essential topic for future study of imaging under turbid media. Section 4.1 demonstrates the cosine similarity based method for image classification to ascertain a proper CNN model for the reconstruction.

4.1 Speckle Classification With Vague Densities

Experiments are conducted in detail. The murky media used in our experiments vary from each other by densities. The volume of water is fixed with 1200 ml but that of the milk used to dilute the solution increases from 40 ml to 69 ml by steady rise of 1 ml. 40 ml is experimentally proven as the lowest limit of milk to mix the water. Otherwise, it is achievable to recover the object image by holographic reconstruction when the volume of milk is lower than 40 ml. The upper limit (69 ml) depends on the differentiation between the camera recorded speckles. The total number of handwritten digits sent to the SLM is 20000. Every 2000 digits are sent to the SLM with 1 ml increase of milk.

As mentioned above, the CNN model used in our experiments can be empirically divided into 3 general classes to make predictions. Dataset compromised by speckles recorded under turbid media with 40 ml to 49 ml of milk is used to train the first CNN model. Speckles recorded under 50 ml

to 59 ml of milk mixed turbid media are taken to train the second CNN model. The third CNN model is trained by speckles obtained from the turbid media with 60 ml to 69 ml of milk. For a speckle recorded under a totally unknown murky medium, 4 methods for image identification mentioned previously can be applied to decide which group of CNN model can be used for image recovery. Firstly, the grayscale object authentication based pattern recognition is proposed to implement classification by a nonlinear correlation algorithm which is described by [32]

$$C(\mu, \upsilon) = |IFT(|FT[S(\mu, \upsilon)]\{FT[O(\mu, \upsilon)]\}^*|^{p-1} \times FT[S(\mu, \upsilon)]\{FT[O(\mu, \upsilon)]\}^*)|^2 \qquad (7)$$

where p denotes nonlinearity strength which is set to 0.3 in our experiments, $S(\mu, \upsilon)$ denotes the speckle recorded under an unknown environment, $O(\mu, \upsilon)$ denotes the speckle captured in a known density of medium, and $C(\mu, \upsilon)$ represents nonlinear correlation output. The maximum value of the output is taken out to symbolize the final correlation coefficient. To verify reliability of the correlation coefficient, each group randomly chooses 200 speckles and we will obtain 3 groups of mean correlation outputs. By comparing 3 groups of correlation coefficients, the environment corresponding to the largest value of correlation coefficient is viewed as similar to that of the speckle. Secondly, there is a ready-made Matlab function 'corr2' to calculate the correlation which denotes the similarities between two images [33]. The coefficient is given by

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left[\sum_m \sum_n (A_{mn} - \bar{A})^2\right]\left[\sum_m \sum_n (B_{mn} - \bar{B})^2\right]}}, \qquad (8)$$

where A and B, $\bar{A}$ and $\bar{B}$ respectively represent the two images and their mean values, and r denotes the similarity coefficient of the two images ranged from 0 to 1. The larger r means the higher dependence between two images. Similarly, we calculate 200 coefficients in each group and take the mean value of 200 coefficients as the correlation coefficient. Then, density of the speckle is approaching to the density of the largest coefficient. Thirdly, machine learning based method for image classification is prevailing and achieves high performance in accuracy [18]. Here, we utilize a CNN model combined with softmax function and cross-entropy for image recognition. Cross entropy loss with softmax function is an output layer used to calculate the probability of every group. 19000 pairs of data in each group are selected to train CNN and then the trained model can classify the 200 inputs into particular groups in real time. For the higher label accuracy, we retrain the CNN model to fine tune originally trained parameters. The last method introduced to calculate the likeness between two images is cosine similarity. Taking advantage of the steps illustrated in Fig. 2, the most possible range of the densities is ensured by the largest mean cosine similarity. The number of the classes of labeled speckles is 3. To ensure the comparability between the four methods, 200 randomly selected speckles are in accordance with those used in previous approaches. In the end, the accuracies of the predictions made by these four methods are obtained as follows: 61.5% for authentication, 65.0% for direct calculation of Matlab function 'corr2', 65.5% for deep learning based method and 76.0% for cosine similarity. In essence, authentication, function of 'corr2' and cosine similarity are based on mathematic formulas to compute the intrinsic correlations between each other which is expected to have the higher acceptance to the variations of turbid environments. The machine learning method is widely used to resolve the problems for image classification in some typical fields such as: biomedicine and security on account of the high classification accuracy [18]. However, it fails to work perfectly in our turbid media with only 65.5% of correct classification rate and the rest wrong classifications are completely irrelevant with the correct labels. In comparison with other pattern classification results [18], the prediction accuracy rate of learning method in our turbid media is lower than expectation. Analysis for lower accuracy of the deep learning based method can be given as follows: there are 10 groups of densities in each class, the learning model is not powerful enough to learn all the characterizations of these speckle patterns. Except for the consideration of label accuracy rate, unconnected classifications are fatal to subsequent processing of image reconstruction on account of completely irrelevant predictions of CNN models. To balance high tolerance to variant environments and high rate of recognition

accuracy, cosine similarity based method for speckle classification is an optimal alternative in our case. Due to the superiority of cosine similarity for image recognition in 3 classes of speckles (each class contains 10 groups of densities), performance of the proposed method is further discussed and tested to classify the speckle into one of the 30 groups of densities.

There are 30 groups of densities of changeable turbid media to investigate the performance of cosine similarity for image classification. These 30 groups of densities are labeled from 1 to 30. We choose 200 images from the 30 groups of diffraction patterns as the testing images and separately calculate cosine similarity between the selected pattern and labeled patterns from 30 groups. As illustrated in Fig. 2, the variable n is 30. To ensure stability of the computed cosine similarity in each environment, each testing image compares with 200 patterns from the same group and correspondingly each group has 200 cosine similarities. Mean value of 200 cosine similarities is considered as the ultimate similarity between the environment of the testing speckle and the environment where the labeled patterns situated in. After 30 rounds of calculations, 30 groups of average cosine similarities will be obtained and compared determining the largest cosine similarity. Finally, the environment matching to the largest cosine similarity (indicated as the ith group in Fig. 2) is deemed as the most similar environment with that of the testing pattern. In other words, density of the testing image is predicted to be approaching to that of the ith groups of patterns. For all of the 200 testing images, the correct prediction rate is 0.5200 which means that 52% of the testing images are accurately predicted with proper densities and classified into the most similar environment. As known previously, the general CNN model trained by 10 groups of densities has the ability to make predictions, although the difference of the density might be 10 ml. Here, a value of fault tolerance (a) is set to define the error between the predicted class and real class. The symbol a denotes an integer, and $|a|$ ranges from 0 to 5. If we set $|a| = 0, 1, 2, 3, 4, 5$, accuracy of the prediction rate is 52%, 83.5%, 90%, 92%, 93.5% and 94%, respectively. From the speckle classification rates, the fault tolerance factor plays an important role for significant enhancement of the accuracy. It indicates that the predicted label is highly possible to be close to its real label within the range. Accordingly, we can validly choose optimal datasets to train the CNN model, and this is a major breakthrough. Compared with the classification accuracy of 3 general classes (i.e., 76%), the accuracy of 30 groups of classes is improved greatly with the fault tolerance. Moreover, it provides an approach to selecting proper datasets for training a CNN model, and then the trained model will benefit the reconstruction. The method of cosine similarity based image classification for the recovery is verified to be robust and adaptive to turbulence [34], [35] of the murky media in this study.

Hence, cosine similarity based speckle classification offers a feasible scheme to predict the density and then choose appropriate datasets to train a CNN model to recover the objects in complex environments.

4.2 Reconstruction With Vague Densities

When we obtain the label of the speckle recorded under totally unknown media, image reconstructions can be made by using the proposed CNN model. Figure 5 shows some representative examples of the reconstructions given after correct classification. The first row shows the object image sent to the turbid media and correspondingly the second row shows the speckles recorded under unknown densities of turbid water. It is obvious that the patterns are captured under different densities of the murky water due to the discriminable characterizations in each speckle. Then, the cosine similarity based pattern classification is implemented to estimate the most similar density of the speckle. Once the training dataset for CNN model is properly predicted, quality of reconstructed images is satisfactory as illustrated in the third row respectively having PSNR values of 26.25 dB, 18.11 dB and 26.82 dB.

The accuracy of totally correct classification is relatively lower than expectations, but the classification results can be used to serve for selecting datasets for training the CNN model. It is demonstrated in Fig. 6 that the speckle can be successfully recovered although wrong classification is made. The reconstructions predicted by classified dataset trained CNN model
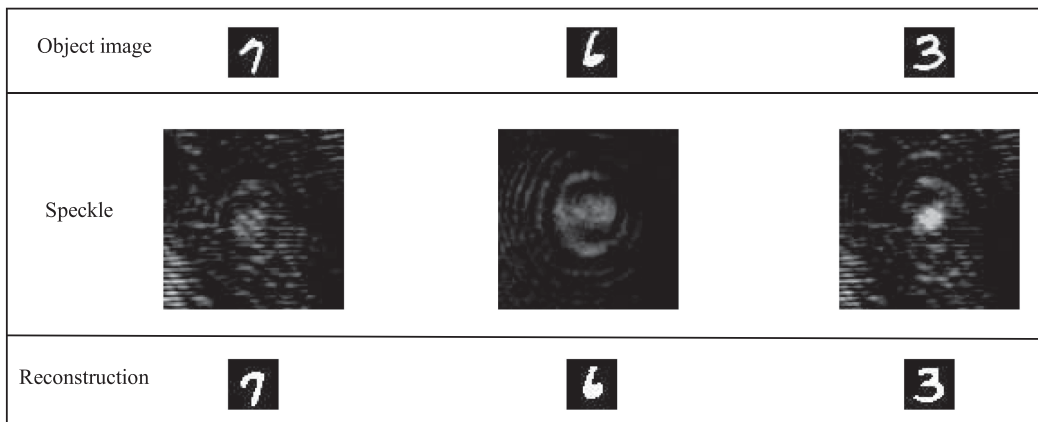
Fig. 5. Reconstructions obtained by using the CNN model trained with correct dataset.
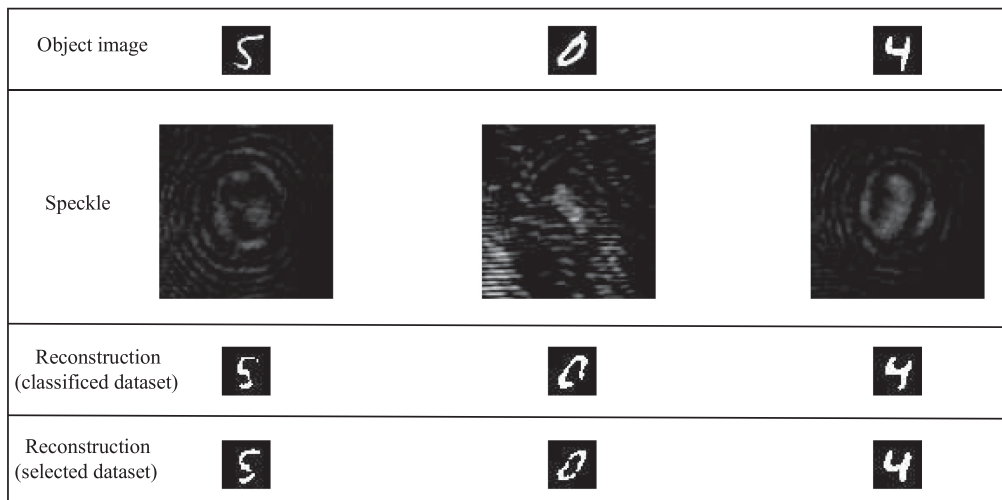

Fig. 6. Comparisons of reconstructions obtained by using CNN models trained by the classified dataset and selected dataset after wrong classification.

and selected datasets trained CNN model are given in the third row and fourth row, respectively. Label of the classified dataset is directly given by the cosine similarity based speckle classification. The PSNR values of the recovered images in the third row are 27.32 dB, 16.43 dB and 23.54 dB, respectively. The selected dataset is comprised by 11 groups of densities following the rule of fault tolerance ($|a| = 5$). The difference between the predicted label and the real label is 1, 3, 2, respectively. The PSNR values of the images reconstructed in the fourth row are 29.89 dB, 21.61 dB and 25.11 dB, respectively. There are two main reasons for setting the fault tolerance as $|a| = 5$. One point is that the CNN model trained by multi-range of densities is experimentally verified to be robust and applicable in variant media. Another explanation is that the classification accuracy under this condition is the highest which ensures better quality of reconstructed images with higher possibility. The existence of slight shape changes of the reconstructions is accessible, since it is quantitative pixel-wise prediction for image reconstruction method compared with the prediction of a serial number for image classification. These results show that reconstruction quality of the speckles classified into wrong labels can still be guaranteed.

In case of wrong classification of the training model for the speckle pattern recorded at unknown condition, it is indispensable to figure out a relatively robust and adaptive architecture. As mentioned
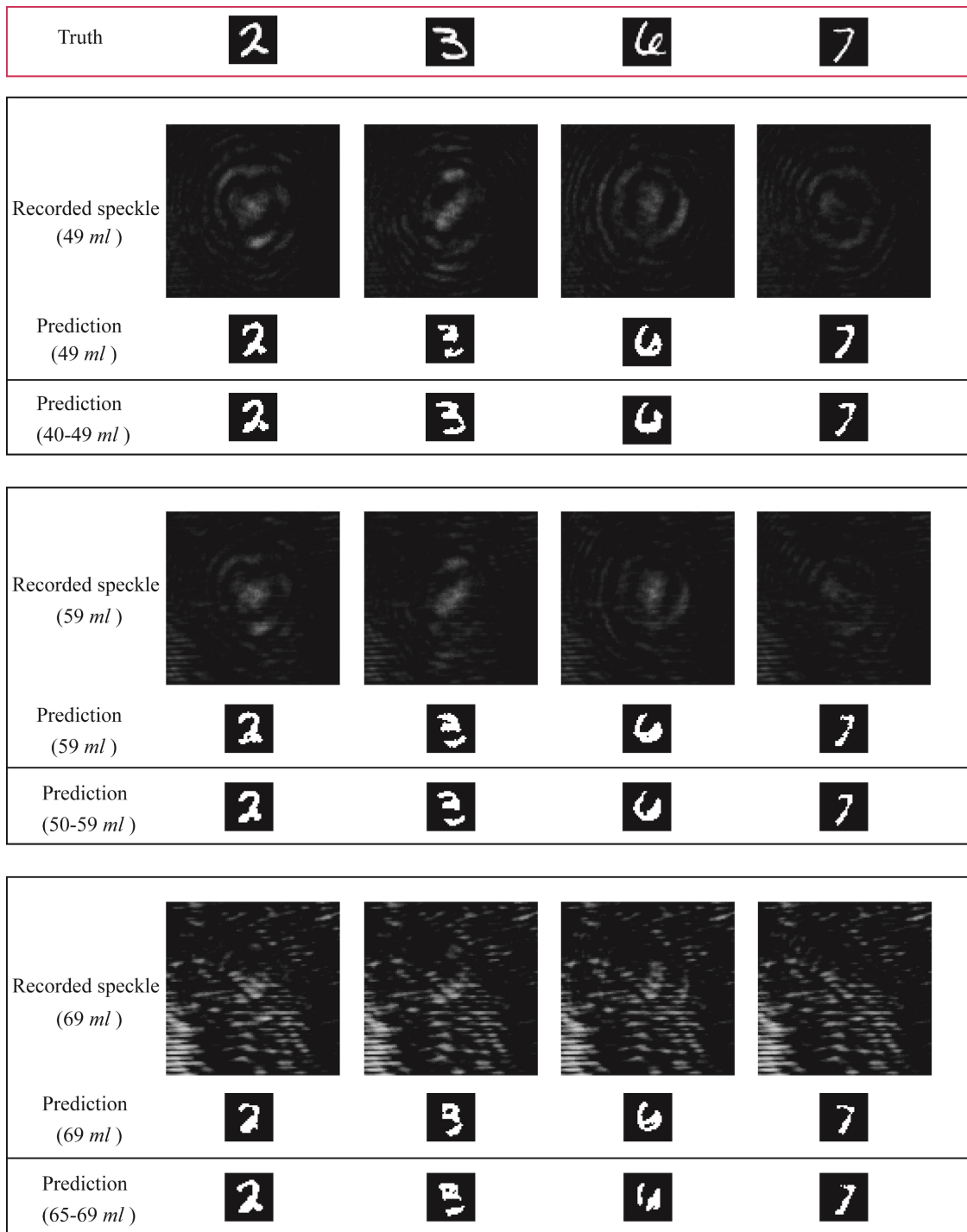
Fig. 7. Comparisons of reconstruction results predicted by using two different training datasets. The first block presents the ground truth sent to the SLM. The second and third blocks respectively show the reconstructions made by 1900 pairs of data trained CNN model (single range density) and the reconstructions given by 19000 pairs of data (multi-range densities) trained CNN model. The fourth block compares the predictions made by CNN trained with data collected under 69 ml of milk with those made by CNN trained with 5 groups of milk amount, ranging from 65 ml to 69 ml.
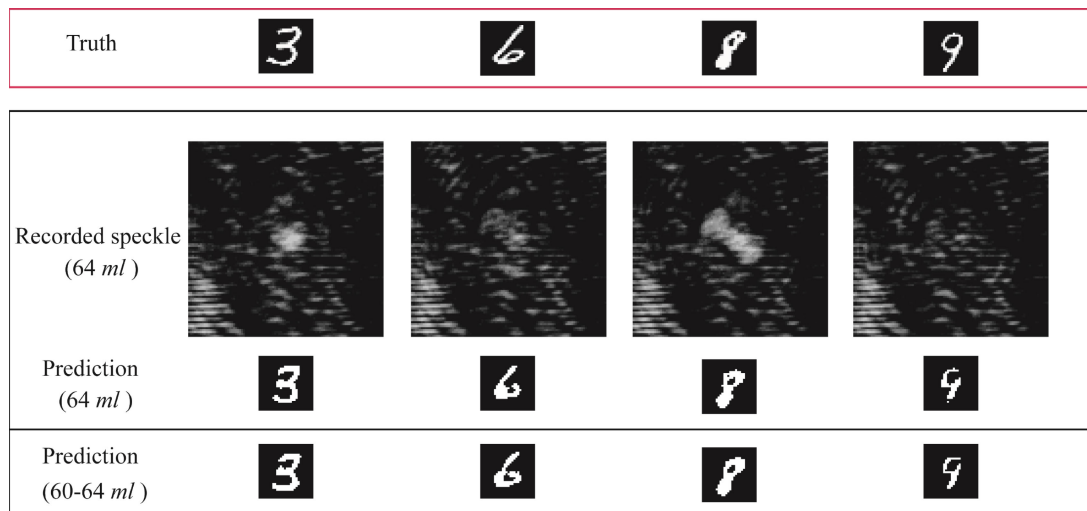
Fig. 8. Comparisons of reconstruction results predicted by two different training datasets. The first block presents the ground truth sent to the SLM. The second block compares the predictions made by CNN trained with 64 ml of milk with those made by CNN trained with 5 groups of milk amount, ranging from 60 ml to 64 ml.

above the diffracted patterns under nearby densities share some similarities. This variation can be deemed as turbulence of the media in practice. Here, it is demonstrated in Fig. 7 that a locally generalized CNN model is fitted to a certain range of perturbations. Comparisons of reconstruction results predicted by two different training dataset are presented in Fig. 7. The first block shows the original handwritten digits sent to the SLM. The second block shows that the CNN model trained by 10 groups of data collected under different densities of turbid water (40 ml to 49 ml of milk) is efficient for reconstruction. Contrast to the results given by the single group of data (49 ml of milk) trained CNN model, the hybrid data trained CNN model performs equally for reconstruction but it is far more universal to the uncertainty of learning environments. Likewise, the quality of reconstructions made by CNN trained using 19000 pairs of data (50 ml to 59 ml of milk) is comparable to that trained only by using 1900 pairs of data (59 ml of milk) as shown in the third block of Fig. 7. Along with the increase of milk, generality of the widely trained CNN model is slightly degraded. As shown in the fourth block of Fig. 7, the predictions made by the CNN model trained with multi-range of milk (65 ml to 69 ml) are similar to those made by the single group of data (69 ml of milk) trained CNN architecture. Figure 8 shows the reconstructions made by the CNN model trained with 64 ml of milk and 60 ml to 64 ml of milk. Both results given by two training dataset resemble the ground truths. Time used for training these multi-range data is about 7 hours for training 10 groups of dataset and about 3.5 hours for training 5 groups of dataset, and then predictions are both made in real time. Hence, there are four groups of CNN models for this type of environment. Although cosine similarity based image recognition wrongly predicts the learning model for image reconstruction, it is probable to recover the original images of the object by choosing one typical trained CNN model from the four groups of models. Rather than putting much effort to seek for a local and contextual method, the proposed deep learning-based technique provides an efficient and relatively stable model for complex and variant environments in practice. Overall, the generalized CNN model trained by multi-range of data is capable to recover the images in scope, thus avoiding repetitive work compared to the existing methods.

Based on these discussions, if density of the turbid media is approximately known, the general CNN models perform well enough to recover the object images. Even though the density is totally unknown, the proposed method is still capable to reconstruct the speckles through the CNN models trained by validly selected datasets. Moreover, four generalized learning models have been constructed to ensure correct image reconstruction after wrong selection CNN models. Hence,

the proposed method using cosine similarity based image classification for the reconstruction has superior adaptability to the variations of the complicated media.

5. Conclusions

A new approach has been proposed for speckle recognition with high classification accuracy, and the classification results play an important role in the reconstruction, especially in the turbid media with vague concentrations. In changeable and harsh environments, cosine similarity based image classification method has competitive edges in high accuracy for classification and larger possibility to approach to the correct label. This cosine similarity based classification technique is validated to perform well in our study compared with other existing methods. The technique for speckle classification benefits CNN model for the reconstruction for the convenience of validly selected training datasets. In consideration of the fact that some existing techniques can be utilized to ascertain the variables aroused by possible factors under complicated environments with unknown turbulences, it is experimentally demonstrated that general CNN models trained by using multi-ranges of environments are highly adaptive. Though the density of the turbid media is totally unknown, the proposed method is advanced enough to classify the speckle into a similar environment and the predicted labels can be utilized to choose the most suitable training datasets for CNN models. Moreover, it is demonstrated that the CNN model trained by using multi-range of data has superiority in imaging through uncertain environments after wrong selections of training models. Quality of the reconstructions is validated to be satisfying. In practice, the method for speckle classification leads to the decreased requirements of cutting-edge precision measurement techniques. It is experimentally verified that this cosine similarity based speckle classification for the reconstruction to retrieve the information hidden by complex environments is advanced, and has good performance in the turbid media with unknown perturbations. Moreover, it could be expected to be promising for the development of marine science.

References

[1] A. P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink, "Controlling waves in space and time for imaging and focusing in complex media," *Nature Photon.*, vol. 6, no. 5, pp. 283–292, 2012.

[2] V. Ntziachristos, "Going deeper than microscopy: the optical imaging frontier in biology," *Nature Methods*, vol. 7, no, 8, pp. 603–614, 2010.

[3] F. Ferri, D. Magatti, A. Gatti, M. Bache, E. Brambilla, and L. A. Lugiato, "High-resolution ghost image and ghost diffraction experiments with thermal light," *Phys. Rev. Lett.*, vol. 94, no. 18, pp. 183602–183606, 2005.

[4] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 3, pp. 14–20, Mar. 2008.

[5] S. R. Arridge and M. Schweiger, "A gradient based optimization scheme for optical tomography," *Opt. Express*, vol. 2, no. 3, pp. 213–22, 1998.

[6] H. Quan and Z. Guo, "Fast 3-D optical imaging with transient fluorescence signals," *Opt. Express*, vol. 12, no. 3, pp. 449–457, 2004.

[7] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.

[8] M. Duarte *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

[9] S. Popoff, G. Lerosey, M. Fink, A. C. Boccara, and S. Gigan, "Image transmission through an opaque material," *Nature Commun.*, vol. 1, no. 6, 2010, Art. no. 81.

[10] J. W. Goodman, W. H. Huntley, D. W. Jackson, and M. Lehmann, "Wavefront-reconstruction imaging through random media," *Appl. Phys. Lett.*, vol. 8, no. 12, pp. 311–313, 1996.

[11] A. V. Kanaev, K. P. Judd, P. Lebow, A. T. Watnik, K. M. Novak, and J. R. Lindle, "Holographic imaging through extended scattering media under extreme attenuation," in *Proc. Imag. Appl. Opt.*, San Francisco, CA, USA, Jun. 2017, Paper ITu3E.

[12] A. Liutkus *et al.*, "Imaging with nature: Compressive imaging using a multiply scattering medium," *Sci. Rep.*, vol. 4, pp. 5552–5559, 2014.

[13] R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," *Opt. Express*, vol. 24, no. 13, pp. 13738–13742, 2016.

[14] A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica*, vol. 4, no. 9, pp. 1117–1125, 2017.

[15] Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light Sci. Appl.*, vol. 7, no. 2, pp. 17141–17150, 2017.

[16] S. Li, M. Deng, J. Lee, A. Sina, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," *Optica*, vol. 5, no. 7, pp. 803–812, 2018.

[17] M. Lyu *et al.*, "Deep-learning-based ghost imaging," *Sci. Rep.*, vol. 7, no. 1, pp. 17865–17871, 2017.

[18] H. Chen, Y. Gao, X. Liu, and Z. Zhou, "Imaging through scattering media using speckle pattern classification based support vector regression," *Opt. Express*, vol. 26, no. 20, pp. 26663–26678, 2018.

[19] Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: A deep learning approach towards scalable imaging through scattering media," *Optica*, vol. 5, no. 10, pp. 1181–1190, 2018.

[20] R. Schettini and S. Corchs, "Underwater image processing: State of the art of restoration and image enhancement methods," *EURASIP J. Adv. Signal Process.*, vol. 14, 2010, Paper 746052.

[21] T. Çelik, H. Özkaramanlı, and H. Demirel, "Fire and smoke detection without sensors: Image processing based approach," in *Proc. 15th Eur. Signal Process. Conf.*, 2007, pp. 1794–1798.

[22] O. Katz, E. Small, and Y. Silberberg, "Looking around corners and through thin turbid layers in real time with scattered incoherent light," *Nature Photon.*, vol. 6, no. 8, pp. 549–553, 2012.

[23] A. Drémeau *et al.*, "Referenceless measurement of the transmission matrix of a highly scattering material using a DMD and phase retrieval techniques," *Opt. Express*, vol. 23, no. 9, pp. 11898–11911, 2015.

[24] C. D. Mobley, "Light and water," *Radiative Transfer in Natural Waters*. New York, NY, USA: Academic Press, 1994.

[25] J. Li and Y. Li, "Underwater image restoration algorithm for free-ascending deep-sea tripods," *Opt. Laser Technol.*, vol. 110, pp. 129–134, 2019.

[26] H. Lu, Y. Li, T. Uemura, H. Kim, and S. Serikawa, "Low illumination underwater light field images reconstruction using deep convolutional neural networks," *Future Gener. Comput. Syst.*, vol. 82, pp. 142–148, 2018.

[27] Y. Li, H. Lu, J. Li, X. Li, Y. Li, and S. Serikawa, "Underwater image de-scattering and classification by deep neural network," *Comput. Elect. Eng.*, vol. 54, pp. 68–77, 2016.

[28] H. V. Nguyen and L. Bai, "Cosine similarity metric leaning for face verification," in *Asia Conference on Computer Vision*. Berlin, Germany: Springer, 2010, pp. 709–720.

[29] S. Mallat, "Understanding deep convolution networks," *Philos. Trans. Roy. Soc. A*, vol. 374, no. 2065, 2016, Art. no. 203.

[30] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[31] [Online]. Available: http://yann.lecun.com/exdb/mnist/

[32] W. Chen and X. Chen, "Grayscale object authentication based on ghost imaging using binary signals," *Europhys. Lett.*, vol. 110, no. 4, 2015, Art. no. 44002.

[33] W. T. Chan, K. S. Sim, and C. P. Tso, "Application of optical character recognition in thermal image processing," *Infrared Phys. Technol.*, vol. 54, no. 4, pp. 353–366, 2011.

[34] X. D. Chen, *Computational Methods for Electromagnetic Inverse Scattering.* Hoboken, NJ, USA: Wiley, 2018.

[35] Y. Xiao, L. Zhou, and W. Chen, "Experimental demonstration of ghost-imaging-based authentication in scattering media," *Opt. Express*, vol. 27, no. 15, pp. 20558–20566, 2019.