RESEARCH ARTICLE

Reconstructing Images Through Multimode Fibers From the Up-Conversion Speckle Patterns via Deep Learning

YANZHU ZHANG [1,2], **HAO WU** [1], **HE ZHAO** [1], **AND YONG CHEN** [1]

[1] School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110168, China
[2] Science and Technology on Electro-Optical Information Security Control Laboratory, Tianjin 300401, China

Corresponding author: Yanzhu Zhang (syzd710471@163.com)

ABSTRACT The mode mixing and mode dispersion in the multimode fiber (MMF) will produce complex speckle patterns in the distal end of the fiber as an object passes through the MMF, rendering image reconstruction to be a challenging task. In recent years, convolutional neural networks have been successfully applied to image reconstruction from speckles. However, the imaging spectra of these studies are mostly in the visible spectrum range and require complete speckle information for reconstruction. In this paper, researchers build an optical imaging system that employs up-conversion imaging technology to collect speckle patterns generated by infrared light transmitting through a multimode fiber and a frequency-doubling crystal. They propose a speckle restoration network (SRNet) based on a generative adversarial network (GAN) to reconstruct speckle images. The generator of GAN uses ResNest and atrous spatial pyramid pooling (ASPP) to extract multi-level features and multi-scale context information, respectively. The discriminator of GAN significantly improves the quality of the reconstructed image generated by the generator. In addition, researchers adopt a special training method named pre-training generator to avoid gradient disappearance or gradient explosion in the training process. With the designed network, high-quality images were successfully reconstructed even with only a portion of the speckle information.

INDEX TERMS Multimode fiber, neural network, speckle restoration, frequency doubling.

I. INTRODUCTION

Optical fiber has been extensively used in endoscopic imaging [1], [2], [3]. Multimode fiber (MMF) has been demonstrated to transmit a significant number of modes simultaneously [1], which significantly enhances the information transmission efficiency compared to single-mode fiber [4]. However, dispersion, coupling, and other phenomena will occur between various modes during MMF transmission, resulting in unrecognizable speckle patterns in the distal end of the fiber. Some researchers have developed methods for reconstructing images from speckle patterns, such as transmission matrix [5], [6], [7], wavefront shaping [8], and digital phase

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Moinul Hossain [ID].

conjugation [9]. These methods have been proven to be able to reconstruct images from speckles. However, these methods require very precise control and measurement as well as are sensitive to the optical system and imaging environment. Even minor changes in the scattering medium may seriously affect the imaging quality. To address these issues, many researchers have proposed a new imaging method based on deep learning, which overcomes the limitations of traditional imaging methods and reconstructs images more stably and continuously. For example, the widely used convolutional neural network (CNN) has been employed for speckle reconstruction [10], [11], [12], [13]. To train CNN, a large number of speckle image pairs need to be used as training data to realize end-to-end training [14], [15]. In the training process, the neural network learns the corresponding relationship between the input image and the label image, without considering

the structure of the relational optical system and the light propagation process.

The research works mentioned above [10], [11], [12], [13] prove that deep learning can be effectively applied to speckle image reconstruction. However, most of the current research is based on the imaging spectrum in the visible range (wavelength: 400 – 700 nm). When the imaging spectrum is outside of the visible spectrum range, the imaging of MMF will be difficult to obtain [16], [17], such as THZ imaging. In addition, some previous work [18], [19] used complete speckle information for image reconstruction. However, complete speckle fields may not be obtained in practical applications. Therefore, it is of great significance to re-construct the complete original image from the incomplete speckle image. In this paper, we collected the visible speckle images as the infrared laser beam passes through the MMF and subsequently through a frequency-doubling crystal. The deep learning method has been employed to reconstruct the original infrared images from speckles detected by the detector working in the visible range. Specifically, because the infrared spectrum is not in the visible light range, the up-conversion imaging technology (such as frequency doubling) is used in the system to transfer the imaging wavelength from the infrared range to the visible light wavelength, so that the charge-coupled device (CCD) can detect the output speckle pattern. Then we built a network called SRNet (speckle restoration network), which is inspired by Deeplabv3+ [20] and GAN [21]. Similar to most image restoration architectures, our generator is also based on encoder-decoder architecture. The difference is that we do not use the traditional U-shaped structure. In the proposed model, ResNext101 and ASPP [20] modules are embedded in the encoder to extract features, and then the decoder is used to recover the extracted features. To improve the repair quality of image details, we designed a discriminator to monitor the output of the network.

We evaluated the performance of SRNet on the collected data set. Specifically, two experiments prove the powerful capability of the proposed network. Firstly, we collected three groups of speckle patterns generated by different incident energy, and the structural similarity (SSIM) of reconstruction results of each group of data can reach above 0.9. Secondly, only local speckle information is used to train the model. The experimental results show that the proposed model can reconstruct high-quality images when only one-quarter of the original speckle image is used.

II. MATERIALS AND METHODS
A. EXPERIMENTAL SETUP AND DATA ACQUISITION

The data acquisition device is shown in Fig. 1. The incident light is emitted by an infrared laser with a wavelength of 1028 nm. The polarizer adjusts the polarization direction of the laser beam to meet the modulated polarization direction of the spatial light modulator (SLM). When the label image is uploaded to SLM, the beam reflected by SLM will
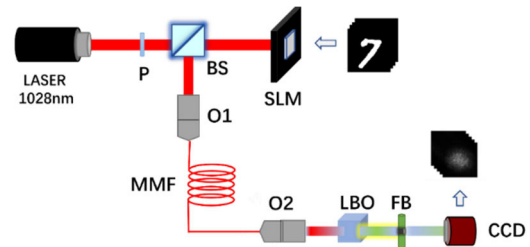

FIGURE 1. Experimental apparatus for collecting data. In the distal end of the MMF, an infrared speckle pattern is generated, and the infrared speckle patterns are upconverted to the visible patterns of wavelength 514nm by second harmonic generation (SHG). P is a polarizer, BS is a beam splitter. SLM is a spatial light modulator, O1 and O2 are two objective lenses, O1: 10×, NA = 0.25, WD = 7.316mm; O2: 20×, NA = 0.4, WD = 1.875mm. MMF is a multimode fiber of length 20 m with a core diameter of 62.5μm, and NA of 0.275. LBO is a frequency-doubling crystal, FB is a bandpass filter, and CCD is a common charge-coupled device.

carry object information and be reflected by BS to objective lens O1. The objective lens O1 couples the laser beam into the MMF. The transmission of light in MMF will lead to dispersion, mode coupling, and other phenomena, which will lead to the formation of infrared speckles. The infrared speckle output from MMF is collected by objective lens O2 and is up-converted to visible speckle by an LBO frequency doubling crystal. The remaining infrared light can be filtered out by using a band-pass filter. The visible speckle pattern is collected by CCD.

We use the device shown in Fig. 1 to get our dataset, which is based on the open-source dataset MNIST [22]. The specific method is to extract 10000 pictures from MNIST and upload them to the spatial light modulator (SLM), after passing through the device of Fig. 1, the corresponding visible speckle pattern is collected by the CCD. The dataset is further divided into three groups by modulating the incident light power, i.e. the incident energy is 1000 nJ, 2000 nJ, and 4000 nJ. There are 10000 pairs of images in each group. The data set is divided into the training set and the test set by the ratio of 8:2.

B. NETWORK STRUCTURE

The architecture of the SRNet is shown in Fig. 2. Inspired by the GAN [21] network, the structure is divided into a generator and a discriminator. The generator is responsible for generating the reconstructed speckle image, and the discriminator is responsible for judging the difference between the image generated by the generator and the real image. The discriminator will promote the quality of the image generated by the generator. The generator and the discriminator play games with each other until they reach a balance. In this situation, the discriminator cannot recognize the true and false pictures from the generator.

Fig. 2 (a) and Fig. 2 (b) are the encoder and decoder of the generator, respectively. In the encoder, the feature extraction backbone network is Resnest-101, which is different from Resnet-101 [23] in Deeplabv3+ [20]. We think that the split-attention in Resnest can perform better in extracting the features from speckle images, which will help improve the
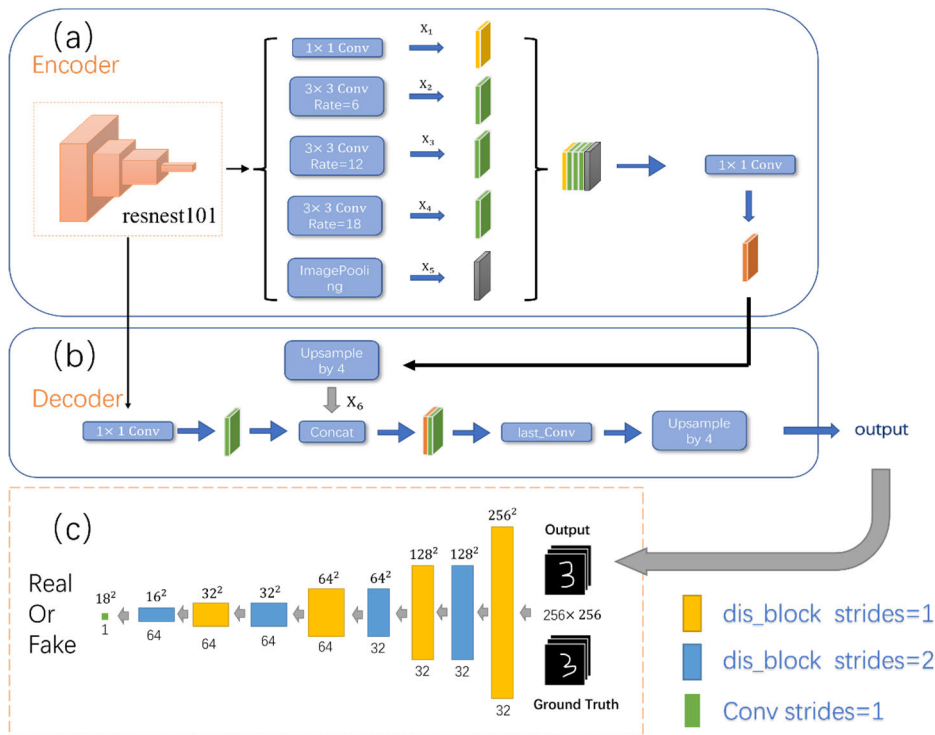
FIGURE 2. SRNet network structure (a) the encoder (b) the decoder c) the discriminator.

quality of reconstructed images. The backbone network outputs two feature maps of different sizes low_level_features and X. X is first sent into the atrous spatial pyramid pooling (ASPP) module, which contains five operations: a 1 × 1 ordinary convolution, three convolution kernels with a size of 3 × 3 with different expansion rates, and a global average pooling [24]. Concatenate the five feature maps output by the above five operations, and then the output of ASPP is gotten. The output of ASPP is connected to the decoder through a jump connection, in which we use 1 × 1 convolution and up-sampling to adjust the dimensions of the feature map.

Another output of the feature extraction backbone network, low_level_features, is directly input into the decoder. In the decoder, we first apply 1 × 1 convolution to low_level_features to reduce its channel number, so that low_level_features can balance with X6 in dimensions. After the connection, several 3 × 3 convolutions are used to refine the features, and then another bilinear upsampling is used to obtain the final output.

The predicted picture from the generator and the ground truth is input to the discriminator and then converted to an 18 × 18 tensor by a series of convolution operations, as shown in Fig. 2 (c). Each dis_block module is composed of Conv2D, BatchNormalization, and LeakyRelu. The convolution kernel size in the yellow dis_block is 3 × 3 and the stride is 1, the convolution kernel size in the blue dis_block is 3 × 3 and the stride is 2. The discriminator calculates the loss using all

1 tensors of the ground truth value image and all zero tensors of the generator output, respectively. This strategy attempts to classify the input image as true or false [25].

C. LOSS FUNCTION

The objective function in this paper is a combination of global loss and adversarial loss. Such multiple losses are conducive to the convolution network to avoid the local optimum [26]. The expression is

$$loss = \alpha V + global\_Loss \tag{1}$$

where V is the adversarial loss and global_ Loss is the global loss, the α is the weight of the adversarial loss. In (1), the global_ Loss is responsible for repairing the main content of the image, while the adversarial can enrich the image details and make the reconstructed image more realistic.

The form of adversarial loss of the model proposed in this paper is given by

$$\min_{G} \max_{D} V(D, G) = E_{x \sim P_{data}(x)}[logD(x)]$$
$$+ E_{z \sim P_z(z)}[log(1 - D(G(z)))] \tag{2}$$

where X is the label image, Z is the speckle image. G(z) and D(x) represent the output of the generator and discriminator. Equation (2) can be expressed as maximizing the discriminator loss and minimizing the generator loss. D(x) represents the probability that an image is a real image. When the discriminator thinks that the probability of an image being

a real image is relatively high, D(x) approaches 1, otherwise, D(x) approaches 0. In the training process, G is first fixed and D is trained. Since x is a real label image and z is a speckle image, D(X) is expected to be larger and D(G(x)) smaller, that is max_D. Then fix D and train G. G is expected to generate a realistic enough picture to fool the discriminator, so D(G(z)) is expected to be larger, that is min_G. Through this continuous alternating training, the generator and discriminator are gradually optimized. Finally, a balance is reached, so that D(G(z)) is close to 0.5. For the global loss in (1), i.e. the generator loss, the binary cross entropy loss is adopted, and the formula is

$$l\left(\hat{y}_i, y_i\right) = -\frac{1}{N} \sum_i^N \left(y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)\right) \quad (3)$$

where $\hat{y}_i$ is the prediction, whose size is between 0 and 1, and y_i is the ground truth, which is equal to 0 or 1. N is the total number of pixels.

D. NETWORK TRAINING
Experiments were performed using the data collected as described in subsection A, each group of data is composed of 10000 image pairs, with 80% being the training set and 20% being the test set.

1) DETERMINE THE WEIGHT OF THE LOSS FUNCTION
The total loss function in (1) is in the form of weighted global loss and adversarial loss. The weight of global loss is set to 1. To prevent the loss of the discriminator feed-back to the generator from being too large, resulting in model divergence, the value α should be carefully chosen. After several attempts by selecting different values between 0 and 0.01, the value is selected as 0.00004.

FIGURE 3. The output of the generator after 1 epoch of training. (a) is the ground truth and (b) is the generated image.

2) TRAINING METHOD
In the training process, the generator is used to reconstruct the target image, and the discriminator is used to judge whether the generated image is true or false. In theory, the discriminator should be abandoned after the two reach Nash equilibrium. But considering the poor performance of the generator at the initial stage of training and the difference between the true graph and the false graph is very obvious, the discriminator can easily have strong discrimination ability and feedback large losses to the generator, which makes it difficult for the generator to converge. Therefore, the generator is trained for n (number of epochs that the generator has

been pre-trained) epoch firstly to ensure that the generator has a certain ability, and then the discriminator is introduced. Experiments show that the generator has enough performance to confuse the discriminator when n is 1. The generator output after one epoch of training is shown in Fig. 3. It can be seen from Fig. 3 that the generator has formed the corresponding contour of the image after one round of training. After a round of training, the discriminator is introduced to constrain the generator.

The optimizer selects Adam. The learning rate is finally determined to be 0.00003 through continuous adjustment, batch_size = 16, and all data was trained for 60 epochs. The detailed parameters of training are shown in Table 1.

TABLE 1. The training parameters of the network.

Learning rate	0.00003
Optimizer	Adam
Batch-size	16
Epoch	60
Pytorch Version	1.11
GPU	NVIDIA GeForce RTX 3060 Laptop GPU

III. RESULT
A. EVALUATING INDICATOR
To evaluate the quality of the generated image, we introduced the structural similarity index (SSIM) [27] and Pearson correlation coefficient (PCC) [28]. Structural similarity evaluates the image quality from three aspects of brightness, contrast, and structure, which is more consistent with the visual effect observed by human vision, given two signals x and y, the calculation formula is given by

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (4)$$

where μ and σ represent the mean and covariance, respectively, and c_1, c_2, c_3 are constants. Pearson correlation coefficient can measure the correlation between two images, and its calculation formula is:

$$\text{PCC} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\left(A\left(i,j\right) - \bar{A}\right)\left(B\left(i,j\right) - \bar{B}\right)}{\sqrt{\sum_{i=1}^{a}\sum_{j=1}^{b}\left(A\left(i,j\right) - \bar{A}\right)^2}\sqrt{\sum_{i=1}^{a}\sum_{j=1}^{b}\left(B\left(i,j\right) - \bar{B}\right)^2}} \quad (5)$$

B. RESULTS ON THE ORIGINAL SPECKLE IMAGE
The size of the speckle image we collected from the speckle field is 256×256, an image of this size can contain almost all the useful information in the speckle field. Fig. 4 shows the loss change curve during data training corresponding to different incident energy. Fig. 4 (a) is the loss curve of the first epoch of training for the generator. It can be seen that the first epoch of training loss converges quickly. In this
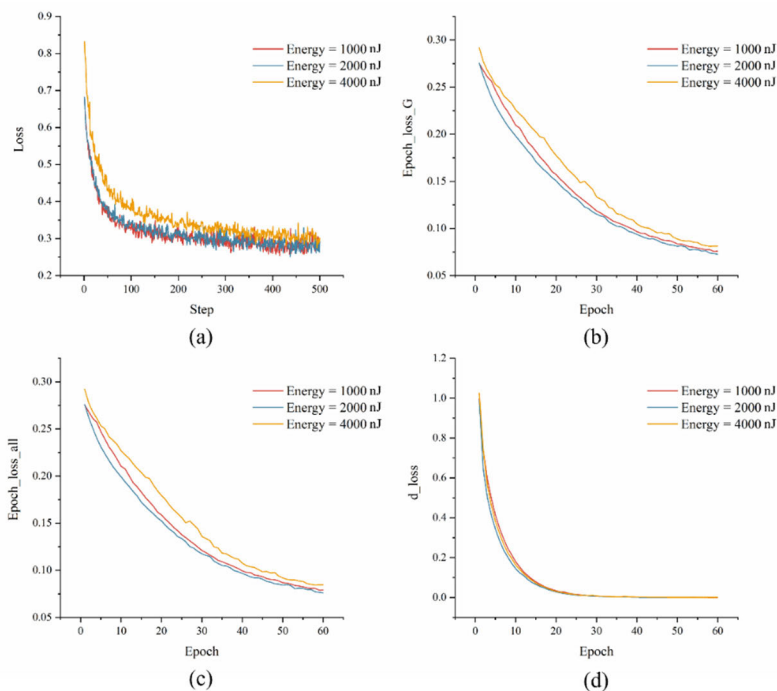
FIGURE 4. The loss curve during the training process. (a) The loss curve of one round of pre-training by the generator. (b) The loss change of the generator after adding the discriminator. (c) The change curve of the total loss during the training process. (d) The convergence process of the discriminator.

stage, the picture output by the generator has been somewhat confusing to the discriminator. Fig. 4 (b) shows the loss change curve of the generator after adding the discriminator, the network can converge well.

In each group of data, 2000 pairs of images are used as the test set to verify the training effect of the model. Fig. 5 shows the results of each model on its corresponding test set. The upper left corner of the image is the SSIM value. The results show that the visual effect of the reconstructed digital image is very similar to that of the label image. It seems that the reconstruction quality is slightly declining with the increasing energy. This is related to the physical mechanism of the frequency doubling. We employ frequency doubling to realize the up-conversion imaging. The intensity of the frequency doubling is proportional to the square of the intensity of the incident fundamental frequency, this is $I(2\omega) \propto I(\omega)^2$. From this point, the speckle pattern of the up-conversion imaging generally is different from that of the initial speckle pattern (the fundamental frequency $I(\omega)$), as is shown in Fig. 6, and the increasing of the intensity will result in a larger difference deviated from the initial speckle pattern (the fundamental frequency $I(\omega)$)), leading to the decreasing of the reconstruction quality.

The average SSIM and PCC of the test set are given in Fig.7. The results show that our network can reconstruct high-fidelity images from handwritten digital data sets. It is worth mentioning that, because our loss function contains adversarial loss, the network repairs the details of the image very well, making the recovered image have a higher SSIM.

C. RECONSTRUCTION OF SPECKLE IMAGE USING ONLY LOCAL INFORMATION

In some practical applications, it is difficult to detect the full speckle pattern. Therefore, it is meaningful to verify whether partial information in the speckle field can be used to reconstruct images. Based on the speckle image obtained, we intercept the local information of different sizes and explore how to use only a part of the speckle information to reconstruct the complete speckle image. This task is challenging because a portion of the speckle pattern may not contain enough information about the original image.

Specifically, we intercept parts of different sizes at the center of the speckle field, as shown in Fig. 8. The intercepted size is 128×128 and 64×64. This means that only a small part of speckle information is used to reconstruct a complete image.

To achieve our goal, we use the proposed network to train on data of different sizes, we only use speckles with an incident energy of 1000nJ as an example in this work. Fig. 9 shows the loss of generators in each group of data training, each group of data can converge well.

Fig. 10 shows the comparison of reconstruction results of speckle images with different sizes. The reconstruction
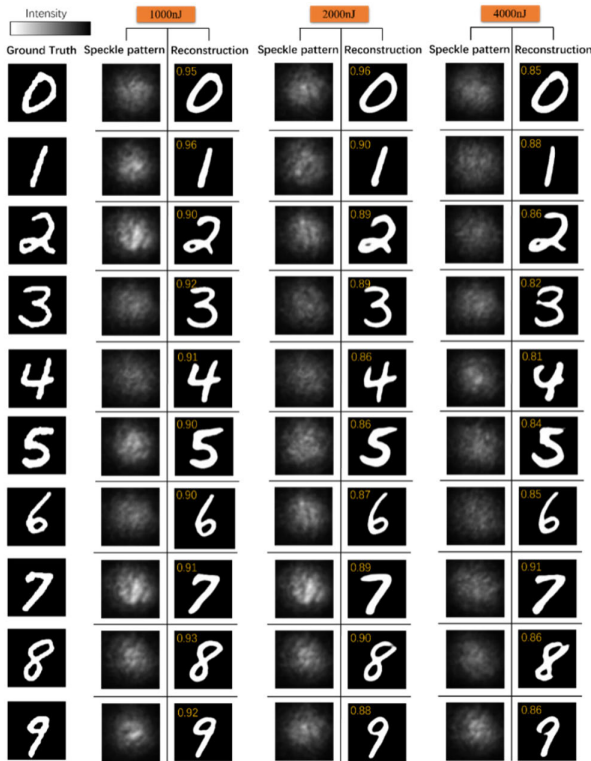
FIGURE 5. Object reconstruction with speckles collected at different incident energy using SRNet. The upper left corner of the reconstructed image is the corresponding SSIM value.
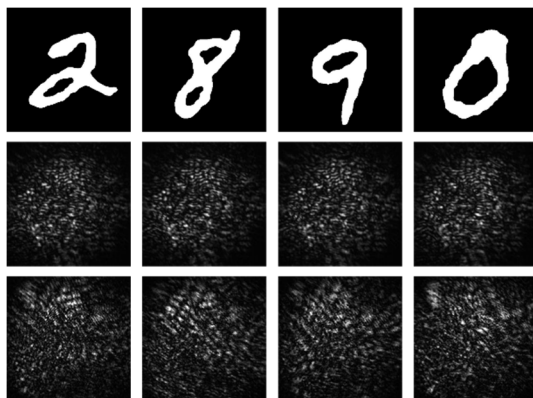

FIGURE 6. The second row is the speckles without frequency doubling (*i.e.*, the speckle in the wavelength 1028 nm), and the third row is the speckles after frequency doubling (*i.e.*, the speckle in the wavelength 514 nm).

results are different in different sizes. Specifically, the result obtained on the original speckle image is the best. With the decrease in the speckle image size, the quality of the reconstructed image decreases slightly. However, according to the results obtained still has a good appearance and high SSIM. Table 2 shows the index comparison of speckle image reconstruction at different scales.

D. ABLATION STUDY

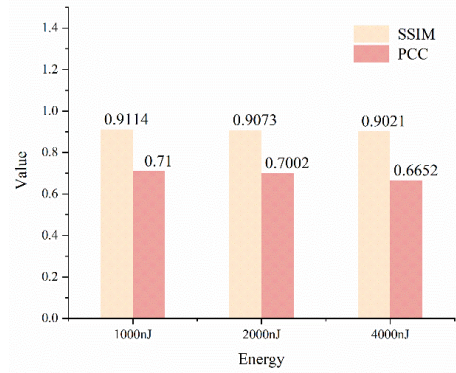To analyze the effectiveness of the feature extraction backbone network and discriminator, ablation experiments were

FIGURE 7. The column diagram shows the speckle reconstruction results with an incident energy of 1000nJ, 2000nJ, and 3000nJ.
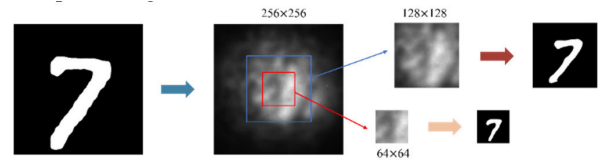

FIGURE 8. Complete speckle image restoration using local speckle information.
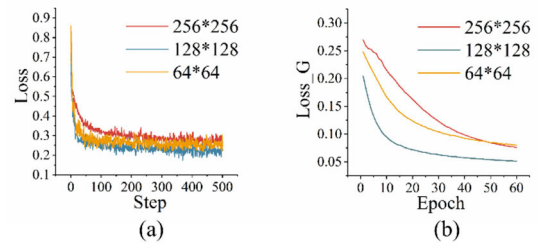

FIGURE 9. Generator loss variation curve for different size training. (a) Loss of one epoch of pre-training by the generator. (b) Generator loss for 60 training epochs.

TABLE 2. Reconstruction results of speckle images with different scales.

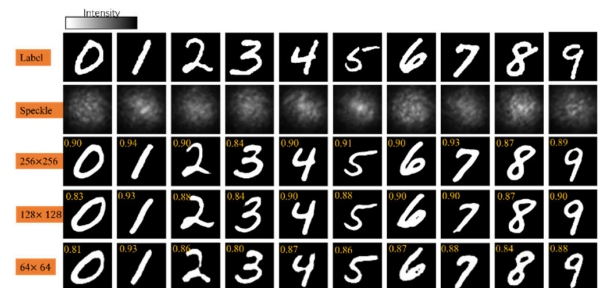| Index	256×256	128×128	64×64
SSIM	0.9114	0.9029	0.8767
PCC	0.7100	0.6662	0.6263

FIGURE 10. Reconstruction results using local speckle information of different sizes.

performed on the data set with an incident energy of 1000 nJ. The experimental results are given in Table 3.

First of all, compare the effects of Resnest-101 and Resnet-101 on the reconstruction quality without

TABLE 3. Effects of different combinations of Resnest-101, Resnet-101, and discriminator on reconstruction quality.

Resnest-101	Resnet-101	Discriminator	SSIM	PCC
✓			0.8932	0.6524
	✓		0.8868	0.6442
✓		✓	0.9114	0.7100
	✓	✓	0.9087	0.7002

adding a discriminator (Row 2 and Row 3 in Table 3). When Resnest-101 is used instead of Resnet-101, both SSIM and PCC are improved. Then we add the discriminator (Row 4 and Row 5 in Table 3), and Resnest-101 still performs better. In addition, no matter who is chosen as the feature extraction backbone network, the addition of a discriminator will also improve the SSIM and PCC, this proves the validity of the designed discriminator.

IV. CONCLUSION

In this paper, we combine upconversion imaging technology with deep learning to study the imaging problem of infrared objects through multi-mode fiber. Firstly, we convert infrared speckles into visible speckles by using a frequency-doubling crystal. Then, to get useful information, we set up a neural network to reconstruct the speckle. The proposed network is named SRNet, which uses Resnest101 as the backbone of the generator to extract multi-level features. A discriminator is introduced to improve the quality of the reconstructed image of the generator. Moreover, the method of pre-training generator is used to avoid the problem of model divergence that may be caused by introducing a discriminator.

The results show that the trained SRNet can reconstruct high-quality images by using the partial speckle image, and the network has good generalization for different incident energies. Considering that it is difficult to detect complete speckle images in some scenes of practical applications, our work is very meaningful. In addition, our work also proves that the combination of upconversion imaging technology and deep learning technology is significant for expanding the imaging wavelength.

ACKNOWLEDGMENT

REFERENCES

[1] J. P. Dumas, M. A. Lodhi, B. A. Taki, W. U. Bajwa, and M. C. Pierce, "Computational endoscopy-a framework for improving spatial resolution in fiber bundle imaging," *Opt. Lett.*, vol. 44, no. 16, pp. 3968–3971, 2019.

[2] M. Hughes, T. P. Chang, and G.-Z. Yang, "Fiber bundle endocytoscopy," *Biomed. Opt. Exp.*, vol. 4, no. 12, pp. 2781–2794, 2013.

[3] B. Rahmani, I. Oguz, U. Tegin, J.-L. Hsieh, D. Psaltis, and C. Moser, "Learning to image and compute with multimode optical fibers," *Nanophotonics*, vol. 11, no. 6, pp. 1071–1082, Mar. 2022.

[4] J. Zhao, X. Ji, M. Zhang, X. Wang, Z. Chen, Y. Zhang, and J. Pu, "High-fidelity imaging through multimode fibers via deep learning," *J. Phys., Photon.*, vol. 3, no. 1, Jan. 2021, Art. no. 015003.

[5] S. M. Popoff, G. Lerosey, R. Carminati, M. Fink, A. C. Boccara, and S. Gigan, "Measuring the transmission matrix in optics: An approach to the study and control of light propagation in disordered media," *Phys. Rev. Lett.*, vol. 104, no. 10, Mar. 2010, Art. no. 100601.

[6] Y. Choi, C. Yoon, M. Kim, T. D. Yang, C. Fang-Yen, R. R. Dasari, K. J. Lee, and W. Choi, "Scanner-free and wide-field endoscopic imaging by using a single multimode optical fiber," *Phys. Rev. Lett.*, vol. 109, no. 20, Nov. 2012, Art. no. 203901.

[7] W. Fan, Z. Chen, L. Chen, L. Wu, X. Ji, and J. Pu, "Polarization transmission matrix for completely polarization control of focal spots in speckle field of multimode fiber," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 4, pp. 1–5, Jul. 2020.

[8] L. Wan, Z. Chen, H. Huang, and J. Pu, "Focusing light into desired patterns through turbid media by feedback-based wavefront shaping," *Appl. Phys. B, Lasers Opt.*, vol. 122, no. 7, pp. 1–7, Jul. 2016.

[9] A. Gover, C. Lee, and A. Yariv, "Direct transmission of pictorial information in multimode optical fibers," *JOSA*, vol. 66, no. 4, pp. 306–311, 1976.

[10] X. Lai, Q. Li, X. Wu, G. Liu, Z. Chen, and J. Pu, "Mutual transfer learning of reconstructing images through a multimode fiber or a scattering medium," *IEEE Access*, vol. 9, pp. 68387–68395, 2021.

[11] H. He, Y. Guan, and J. Zhou, "Image restoration through thin turbid layers by correlation with a known object," *Opt Exp.*, vol. 21, no. 10, pp. 12539–12545, 2013.

[12] J. Bertolotti, E. G. van Putten, C. Blum, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Non-invasive imaging through opaque scattering layers," *Nature*, vol. 491, no. 7423, pp. 232–234, Nov. 2012.

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

[14] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019.

[15] L. Aziz, Md. S. B. Haji Salam, U. U. Sheikh, and S. Ayub, "Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review," *IEEE Access*, vol. 8, pp. 170461–170495, 2020.

[16] X. Qiu, F. Li, W. Zhang, Z. Zhu, and L. Chen, "Spiral phase contrast imaging in nonlinear optics: Seeing phase objects using invisible illumination," *Optica*, vol. 5, no. 2, pp. 208–212, 2018.

[17] F. Lin, X. Qiu, W. Zhang, and L. Chen, "Seeing infrared optical vortex arrays with a nonlinear spiral phase filter," *Opt. Lett.*, vol. 44, no. 9, pp. 2298–2301, 2019.

[18] C. Zhu, E. A. Chan, Y. Wang, W. Peng, R. Guo, B. Zhang, C. Soci, and Y. Chong, "Image reconstruction through a multimode fiber with a simple neural network architecture," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, Jan. 2021.

[19] Y. Li, Z. Yu, Y. Chen, T. He, J. Zhang, R. Zhao, and K. Xu, "Image reconstruction using pre-trained autoencoder on multimode fiber imaging system," *IEEE Photon. Technol. Lett.*, vol. 32, no. 13, pp. 779–782, Jul. 1, 2020.

[20] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Munich, Germany: Springer, 2018, pp. 801–818.

[21] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018.

[22] L. Deng, "The MNIST database of handwritten digit images for machine learning research [best of the web]," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 141–142, Nov. 2012.

[23] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, Y. Sun, T. He, J. Mueller, R. Manmatha, M. Li, and A. Smola, "ResNeSt: Split-attention networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 2736–2746.

[24] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.

[25] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[26] C. Xu, C. Lu, X. Liang, J. Gao, W. Zheng, T. Wang, and S. Yan, "Multi-loss regularized deep neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 12, pp. 2273–2283, Dec. 2016.

[27] 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.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[28] P. Sedgwick, "Pearson's correlation coefficient," *Bmj*, vol. 345, pp. 1–2, Jul. 2012.

HE ZHAO was born in Heilongjiang, China, in 1999. He is currently pursuing the master's degree with Shenyang Ligong University. His research interests include image reconstruction and multiple scattering.

YANZHU ZHANG received the Ph.D. degree in engineering from Northeastern University, Shenyang, China, in 2008. In 2012, she stand out from the Computer Science Postdoctoral Research Station, Northeastern University. From 2015 to 2016, she was a Visiting Scholar with the University of California. She is currently an Associate Professor with the School of Automation and Electrical Engineering, Shenyang Ligong University, Liaoning, China. Her research interests include fractional control, image processing and recognition, data-driven control, and scattering imaging.

HAO WU was born in Liaoning, China, in 1999. He is currently pursuing the master's degree in control engineering with Shenyang Ligong University, China. His research interests include deep learning and the imaging of scattered media.

YONG CHEN is currently a Research Fellow with the School of Automation and Electrical Engineering, Shenyang Ligong University, China. His research interest includes scattering imaging.

• • •