photonics

MDPI

Article

Imaging Complex Targets through a Scattering Medium Based on Adaptive Encoding

Enlai Guo †, Yingjie Shi †, Lianfa Bai and Jing Han *

Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing 210094, China; njustgel@njust.edu.cn (E.G.); syj@njust.edu.cn (Y.S.); blf@njust.edu.cn (L.B.)
* Correspondence: eohj@njust.edu.cn
† These authors contributed equally to this work.

Abstract: The scattering of light after passing through a complex medium poses challenges in many fields. Any point in the collected speckle will contain information from the entire target plane because of the randomness of scattering. The detailed information of complex targets is submerged in the aliased signal caused by random scattering, and the aliased signal causes the quality of the recovered target to be degraded. In this paper, a new neural network named Adaptive Encoding Scattering Imaging ConvNet (AESINet) is constructed by analyzing the physical prior of speckle image redundancy to recover complex targets hidden behind the opaque medium. AESINet reduces the redundancy of speckle through adaptive encoding which effectively improves the separability of data; the encoded speckle makes it easier for the network to extract features, and helps restore the detailed information of the target. The necessity for adaptive encoding is analyzed, and the ability of this method to reconstruct complex targets is tested. The peak signal-to-noise ratio (PSNR) of the reconstructed target after adaptive encoding can be improved by 1.8 dB. This paper provides an effective reference for neural networks combined with other physical priors in scattering processes.

Keywords: deep learning; scattering imaging; computational imaging

Citation: Guo, E.; Shi, Y.; Bai, L.; Han, J. Imaging Complex Targets through a Scattering Medium Based on Adaptive Encoding. *Photonics* **2022**, *9*, 467. https://doi.org/10.3390/photonics9070467

Received: 10 June 2022
Accepted: 1 July 2022
Published: 4 July 2022

1. Introduction

Vision is an important way for humans to obtain information. Normally, light travels along a straight line if the medium is homogeneous. Complex media such as fog and biological tissues will cause the light to scatter and the target information obtained by the human eye or camera is highly degraded [1,2]. How to obtain target information through the scattering medium has become a hot research topic. The existing technologies to look through an opaque medium mainly include adaptive optics technology [3,4], optical coherence tomography [5,6], and methods based on point spread function or transmission matrix [7–11]. Methods based on speckle correlation and machine learning are also increasingly being used [12–18]. Physical modeling-based methods have a limitation on optimization and solution capabilities. At present, those methods can successfully restore characters or simple line structure targets, but it is difficult to recover hidden targets with more details such as the human face. The main reason is that the existing physical methods find it difficult to effectively extract target features from the aliased signals after scattering.

Machine learning is more and more widely used in computational imaging, because it can establish connections between data and improve the image quality metrics [19,20]. The nonlinear characteristics of deep learning can well solve the ill-posed problem such as recovering the target hidden behind the diffuser. Yiwei Sun et al. use the Generative Adversarial Network (GAN) to improve the quality of the recovered image through adaptive scattering media [21]. PDSNet proposed by Enlai Guo et al. to look through the diffuser and the field of view (FOV) expended up to 40 times the optical memory effect (OME) [17]. In addition, there are some methods that use machine learning to reconstruct face targets with more details.

Horisak et al. introduced the Support Vector Regression (SVR) to recover the face target, but the reconstructed face is not accurate and lacks detailed information [22]. At the same time, SVR still reconstructs a face learned during the training process when a non-face target appears in the test set. Shuai Li et al. propose IdiffNet to image through the scattering medium, and the quality of reconstructed images influenced by loss function and training set is discussed in detail [23]. This network can reconstruct the face targets which are behind ground glasses with 600 coarser grit. However, the details of the reconstructed target will be lost when the ground glass with the stronger scattering ability of 220 grit is used. Without increasing the amount of training data and system modulation, it is difficult to reconstruct the exact details of complex objects by using deep learning without integrating physical priors.

The separability of the data reflects the difficulty of extracting characteristics from the data to a certain extent. The redundant character of the speckle reduces the separability of data, and it increases the difficulty of neural network optimization without combining any physical feature. Speckles need to be modulated to improve the separability of signals and enhance the ability of the neural network to reconstruct the target. Coding as a good signal modulation method is widely used in scenes such as aliasing signal unmixing, and it has been introduced into the field of scattering research. Tajahuerce et al. proposed a single-pixel-based method to look through scattering media [24]. The encoding in the imaging process is used as the measurement matrix in Compressed Sensing (CS) to recover the encoded hidden target information. Li et al. proposed a method of imaging a target hidden behind scattering media based on the CS theory [25]. However, this method is sensitive to noise and is only suitable for translation-invariant systems. In the reconstruction process of the above two methods, the encoding mask is regarded as a known quantity to recover the encoded object, and the encoding process is used as a modulation means for the reconstruction algorithm to solve the encoded object, rather than as a tool for mining the physical characteristics of speckle or improving the separability of data. Both of them use the fixed encoding mode without considering the difference of the encoded object itself and the lack of effective mining of its structure and noise characteristics, which also limits the reconstruction ability and robustness of the algorithm.

This paper proposes a neural network named AESINet which can make effective use of the physical characteristics of speckle information redundancy. AESINet improves the separability of data by adaptively encoding speckle patterns, and better separability of data can help the network to extract more effective features from the training data, which further enhances the ability of detail reconstruction. Targets with similar structure and rich details can be reconstructed even when the industrial CMOS is used, whose sensitivity and resolution are lower than that of the scientific complementary metal-oxide semiconductor (sCMOS). The ability of AESINet to reconstruct the target hidden behind the opaque medium is tested. The PSNR of the reconstructed face target can reach 24 dB and the PSNR can be increased by 1.8 dB at least compared with non-adaptive encoding or random coding methods.

2. Adaptive Encoding Model Design

For a learning-based algorithm, the neural network can map the relationship from speckle to the original target with the help of a large amount of data. Speckle patterns are highly redundant, therefore it is possible to recover target structure information through low-resolution speckles [26]. If a neural network with a reasonable structure is constructed, it is possible to recover the target behind the scattering medium by using the low-resolution speckle. It can reduce the requirements for camera resolution and computing resources.

Current methods based on deep learning to look through scattering media mainly use an end-to-end neural network structure. This structure gives the flexibility of neural network optimization. However, the neural network sometimes converges to a local optimal solution and overfitting occurs because of the high degree of freedom of the network parameter. Solving those problems often requires a lot of training data. It will be

difficult to recover the target with complex structure and rich details if there are not enough data, and a model with good generalization performance is also difficult to build. Although the redundancy of speckles makes it possible to recover the original target through low-resolution speckles, the noise introduced by the industrial camera further reduces the separability of the data, and it increases the difficulty of neural network optimization. Thus, the simple fitting of data by using an end-to-end network will not be conducive to recover the high-quality reconstructed target. A new neural network structure needs to be properly constructed to improve the separability of the pattern by adaptively encoding redundant speckles.

The randomness of scattering causes every pixel on the sensor to receive signals from any area of the target, therefore, the information of the entire target is aliased on the speckle recorded by the sensor. Since the high-frequency part containing the detailed information is weaker than the low-frequency information, the detailed information is easily submerged. In addition, the aliased information makes it more difficult to extract features from speckles and the dataset of speckles is less separable. Inspired by coding modulation, it is expected to find the corresponding coding mask based on speckle signal characteristics, and modulating the original speckle with the optimized coding template. Because the high-dimensional features extracted by neural networks are difficult to express and constrain by existing methods, therefore, a two-stage network is constructed and the unique features of the speckle are extracted through the first stage network and the coding template is generated. The encoded speckles are an input to the second stage neural network for reconstruction. The evaluation of the reconstruction results is used as the constraint of each stage network, and the optimization coding mask and better reconstruction results are sought by the way of gradient descent.

AESINet is composed of two parts that are respectively used for the construction of adaptive encoding and the reconstruction of a hidden target structure. In the first stage, AESINet-E takes the original speckle as input, and it is responsible to build the intrinsic relationship between the original speckle image and its corresponding ideal encoding pattern. After the encoding process, the data separability of the encoded speckle is effectively enhanced compared with the original speckle, which will be proved in the following experiments. AESINet-R as the second stage takes the encoded speckle ES as input, which can better extract features that contain hidden target structure information. ES is calculated by:

$$ES = S .* M, \tag{1}$$

where S is the original speckle, M is the ideal encoding pattern of S, and $.*$ is the dot product operation. Because only the speckles captured by the camera need to be encoded, only the mask needs to be dot to the original speckle, and no additional devices need to be added for system modulation The networks of the first and second stage are jointly optimized; this strategy makes the adaptive encoding process to seek global optimal solutions easier. Finally, the detailed information of the face and other targets is reconstructed.

The U-shaped network commonly used in the optical field is used as the basic structure of AESINet-E and AESINet-R. The two parts of the network have similar structures although the roles of them are different. Speckles with a low-resolution of 256 × 256 are used as the input of AESINet to strike a balance between network computing efficiency and the amount of input information. AESINet uses a combination of the small-scale convolution kernel and dilated convolution to fully extract the features of different dimensions. At the same time, a combination of 1 × 3 and 3 × 1 convolution is used to replace the 3 × 3 convolution to reduce the number of parameters while ensuring the accuracy of the convolution. In the Encoder part, the scale of feature maps is decreasing as the calculation progresses from shallow to deep. The feature information obtained by each layer gradually changes from low-dimensional pixel-level information to high-dimensional semantic-level information. Keeping the size of the convolution kernel constant during this process is equivalent to the expansion of the receptive field. Different sizes of dilated convolution are applied on the 32 × 32 feature layer to extract semantic-level high-dimensional features.

Each convolution output contains information corresponding to a larger range of feature pattern to extract information under different sizes of receptive fields. In addition, to avoid overfitting and vanishing gradients in the training process a dropout strategy with a parameter of 0.1 is also used in AESINet, that is, 10% of the convolution kernel elements are randomly reset to zero during each calculation. At the end of the AESINet-E structure, an additional Sigmoid layer is added. On the one hand, it is used as an activation function and on the other hand, it is to control the grayscale distribution of the output code between 0 and 1. The finally constructed AESINet takes into account the computational efficiency and feature mining capabilities of the network is shown in Figure 1.

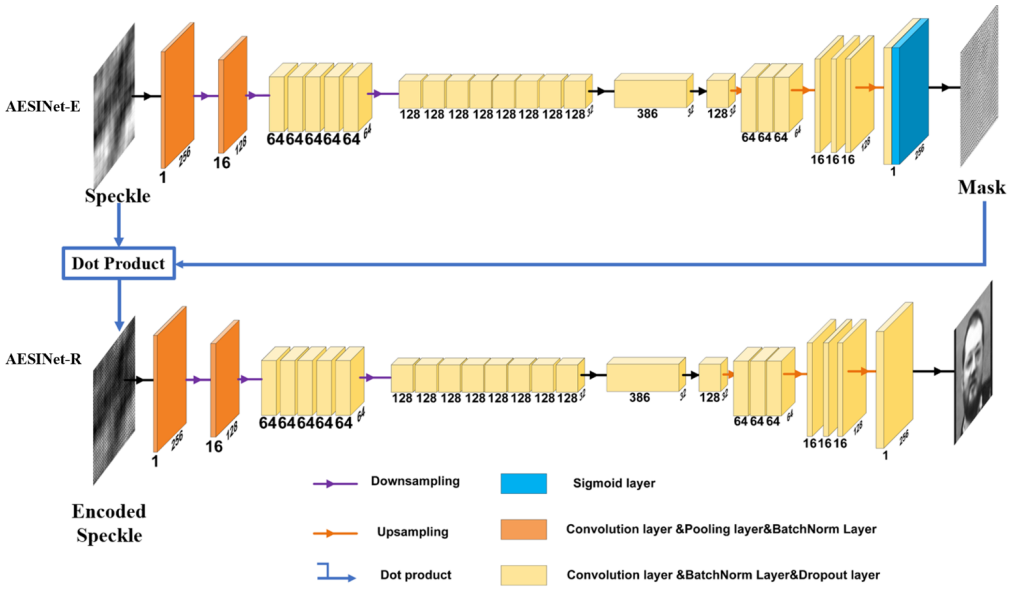

Figure 1. Network structure diagram of AESINet.

The Mean Square Error (MSE) is used as the loss function in the training process of AESINet. The MSE is formulated as:

$$MSE = \frac{1}{H * W} \sum_{i=1}^{H} \sum_{j=1}^{W} [I'(i,j) - I(i,j)]^2, \quad (2)$$

where I' is the reconstructed image, I is the original image contained targets. H and W are the height and width of those images, respectively. The mini-batch strategy is adopted, that is, the entire training set is randomly divided into several parts, and the samples in a subset are trained each time. This training strategy can help the network to jump out of the local minimum and can effectively speed up the convergence of the network. The original low-resolution speckle image is first sent to AESINet-E for forwarding propagation calculation. After getting the adaptive encoding template under the current network parameters, the encoding template is multiplied to the original speckle image and sent to AESINet-R. The reconstructed target and the Ground Truth (GT) in the training set are sent to the loss function to calculate. Then the backpropagation algorithm is used to optimize the network parameters.

3. Experiment

3.1. Experimental Verification of Speckle Redundancy

The optical system set up in this paper is shown in Figure 2. A projector based on an LED light source (Acer, K631i, 1280 × 820 DPI, Xinbei, China) was used to project the target. The light signal carrying the target information was modulated by the ground glass (Edmund, #47-953, 220 grit, Barrington, NJ, USA), then the beam passed through the pupil (Thorlabs, ID25SS/M, Diameter = 12 mm, Nuneaton, NJ, USA) and bandpass filter.

Finally, the speckle was recorded by the camera. The projector was working at the shortest projection distance in the process of data acquisition, and the distance between the focal plane and the projector was 40 cm. The distance between the focal plane and the scattering medium was 170 cm and the distance between the scattering medium and the camera was 25 cm.

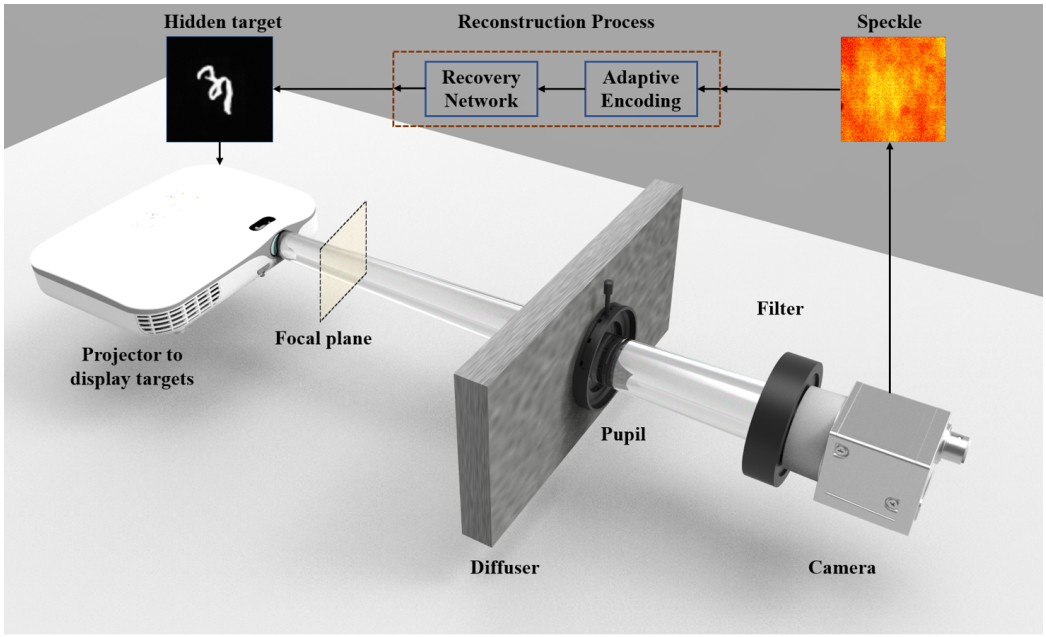

Figure 2. The optical system setup and the process of adaptive encoding and reconstruction of speckles.

The current learning-based methods usually use speckle signals with a spectral bandwidth of about 1 nm as the network input. Either a narrow-band light source such as a laser is used, or a narrow-band filter is used in front of the camera. The use of narrow-band light sources will increase the cost of the optical system, and other devices need to be introduced for coordination. If a narrow-band filter is used, it is required that the light source has sufficient intensity in the gated band. Otherwise, it is necessary to increase the exposure time of the camera to ensure that enough light signals are collected, which will introduce more detector noise into the collected image. Both of these methods are not conducive to the application of this technology in actual scenarios. Thus, a bandpass filter (Thorlabs, FB500-10, Nuneaton, NJ, USA) with a half-height width of 10nm was used in front of the camera. Compared with the narrowband speckle signal of 1nm, the contrast of the speckle signal input to AESINet was significantly reduced [13], and it was more difficult to construct the mapping relationship between the speckle and the original target. The industrial camera (Balser, acA1920-155um, Ahrensburg, Germany) was used to collect 8-bit low-resolution speckle signals. Compared with scientific CMOS, the hardware price is significantly reduced. Low-resolution speckles also place higher requirements on the ability of the network to extract data features. AESINet needs to be able to reconstruct the original target from the speckle where part of the information is submerged by noise.

In order to intuitively verify the characteristics of speckle redundancy, the original speckle image with a resolution of 1920 × 1200 pixel was collected, and five areas with a size of 400 × 400 pixel were randomly cropped on the image as R1 to R5, as shown in Figure 3a. A0 is the autocorrelation of GT, and A1 to A5 is the autocorrelation of R1 to R5 respectively. As shown in Figure 3b, to put the effective autocorrelation scale close to A0 for comparison, the autocorrelations of subspeckle are the 65 × 65 areas of the center original image. The five curves in the left image of Figure 3c are the intensity curves of A1 to A5 at the position L1, and the right image of Figure 3c is the intensity curve corresponding to autocorrelations at the position L2. Figure 3d–f correspond to three different independent

experiments. The FOV in the experiments was within the constraints of OME, therefore, the autocorrelation of the speckle image in the experiments should be consistent with the original target.

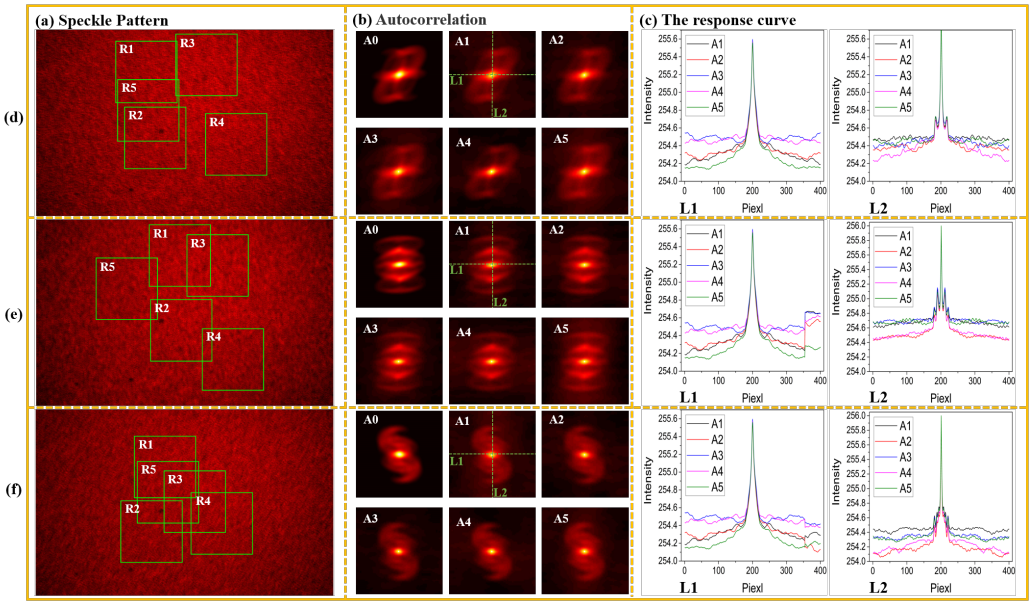

Figure 3. Comparison of speckle redundancy. (**a**) Original speckle pattern, R1 to R5 are the position of five subspeckle. (**b**) A0 is the autocorrelation of the GT corresponding to (**a**), A1 to A5 is the autocorrelation of speckles at the corresponding positions at R1 to R5. (**c**) The left and right curves are the gray values at direction of L1 and L2 on A1 to A5 respectively. (**d**–**f**) shows speckles corresponding to three different targets

Although the positions of the five subspeckles in each group of experiments were randomly selected, it can still be found that the autocorrelations of these subspeckles were highly consistent with the autocorrelation of GT corresponding to the complete speckles. It shows that the local speckle was modulated with all the information of the original hidden target. This experimental phenomenon verifies that the speckle has the physical characteristic of redundancy.

3.2. Dual-Character Target and Face Dataset Experiment

Dual-characters and human faces were used as hidden targets to build two datasets respectively, and the image quality of reconstructed targets with different complexities was tested and analyzed to verify the effectiveness of the AESINet. The dual-character dataset was generated by the MNIST handwritten character set. Each time, two characters were randomly selected from MNIST and combined into a new target to increase the complexity of the target structure. The dual-characters generated by these random characters were used as the target, and the corresponding gray speckle image with a resolution of 256 × 256 was collected in the optical system. Seven thousand five hundred groups of data pairs consisting of speckle and GT were randomly selected as the training set, and the remaining 500 groups were used as the test set. The ability of the AESINet was tested by recovering the unseen objects in the test set. The recovered target of the test set is shown in Figure 4. However, part of the structure in the reconstructed targets had some errors, such as the first reconstruction result which was composed of the number 8 and the number 1; the continuous part of the number 8 was recovered into a discontinuous structure. As shown in Table 1, the average SSIM of the reconstructed image is 0.9060, which also proves that the image reconstructed by AESINet can contain the overall structure of the original target.

Table 1. Objective evaluation indexes of AESINet reconstruction results under test sets.

Dataset	MAE	SSIM	PSNR (dB)
Untrained character target	0.0167	0.9060	23.9437
Untrained face target	0.0402	0.7680	24.1698

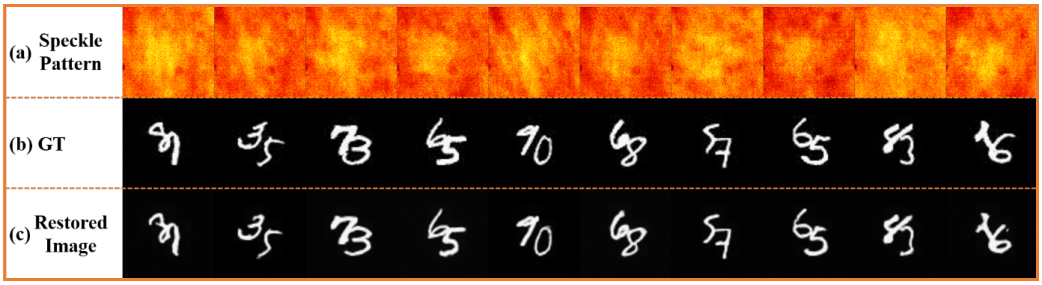

Figure 4. Dual-character experiment results. (**a**) The speckle pattern. (**b**) The Ground Truth (GT) of the hidden target. (**c**) The reconstruction results by using adaptive coding.

A face dataset was used for targets in the following experiments to verify that the network can recover targets with more details. The FEI face dataset was used as the hidden target which contains faces with different expressions and shooting angles of 200 different people. The optical system in Figure 2 was used to collect speckles corresponding to 20 different expressions of 25 people. Among the collected data, 18 group datasets were randomly selected to become 450 groups of training data, and the remaining 50 groups were used as the test set. After the training process converges, the recovered test set is shown in Figure 5 The recovered image is relatively close to the real face distribution in GT, and the global structure of the face is restored well. There are obvious differences in the face details of different restored people, and even the details of the character's hairstyle have been reconstructed. The network has realized the reconstruction of the face with rich details. However, it can also be seen from the reconstruction results that many structures of the image are blurred, which is consistent with the changes in evaluation indicators such as the average MAE, and SSIM drops by 0.138 compared with the reconstruction results of the dual-character target. As shown in Table 1, the ambiguity in the reconstruction result is basically since the features extracted by the network from the speckle signal are not sufficient to fully characterize all the details of the target structure.

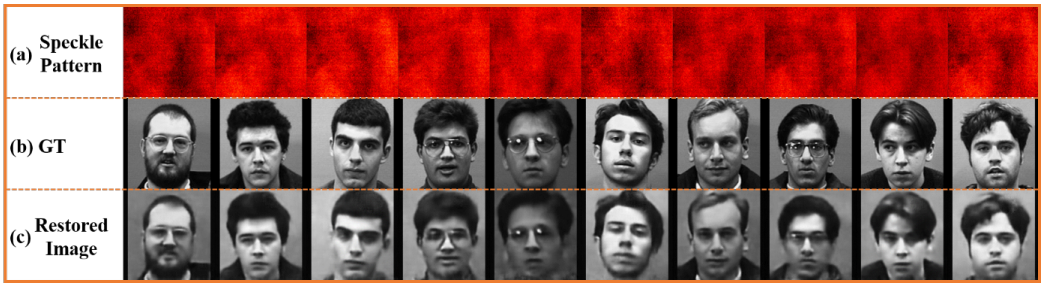

Figure 5. The reconstruction result of the face data set by AESINet. (**a**) The Speckle pattern. (**b**) The Ground Truth (GT) of the hidden face. (**c**) The reconstruction results by using adaptive coding.

From the experimental results of the two different targets in this section, it can be seen that AESINet is suitable for target reconstruction tasks of different complexity. AESINet can extract the features needed to reconstruct hidden targets from low-resolution speckles, especially for the reconstruction task with rich details, and the network can recover most of the detailed information of the face structure.

3.3. Improved Data Separability by Adaptive Encoding

In this section, experiments were conducted to verify that the adaptive encoding process can improve the separability of data. The face targets with the corresponding speckle were used as the experimental dataset, and the low-resolution speckle signal was used as the input to train the complete AESINet. At this time, the input of AESINet-R was the speckle image after adaptive encoding, and then the training model was saved. The same training set was directly input into the AESINet-R without adaptive encoding as a contrast experiment. Six untrained speckle images from the test set were input into the above two AESINet-R network models, and the feature maps at each layer were saved. The feature maps output by each layer of the network were mapped from high-dimensional space to low-dimensional space. When different speckles were used as input, the distance of the feature map in the low-dimensional space was compared to analyze the impact of adaptive encoding on the data separability.

The used data dimensionality reduction visualization method is Uniform Manifold Approximation and Projection (UMAP). This algorithm is a non-linear dimensionality reduction algorithm based on local manifold approximation. Low-dimensional projection of the data is performed by searching for the closest equivalent fuzzy topological structure, which can better reflect the distance between the high-dimensional global structure and the local structure.

Figure 6 shows the visualization results of the first layer, the sixth layer, and the 21st layer. Each point in the figure represents a feature map output by the corresponding layer, and each color represents several feature maps corresponding to a test speckle image. In the figure, a total of six colors correspond to the feature maps of six test speckle images. It can be seen from Figure 6 that the positions of the various color points are staggered if the original speckle image is directly input into AESINet-R without adaptive encoding. It is shown that the divergence between features of different test speckle images is small whether in the shallow feature extraction process or the target reconstruction process. As a comparison, the distribution of the same color points is relatively concentrated whether in the shallow or deep layers after adaptive encoding, and the divergence between the features represented by different color points increases. The separability of the data input to the reconstruction network is effectively enhanced after adaptive encoding.

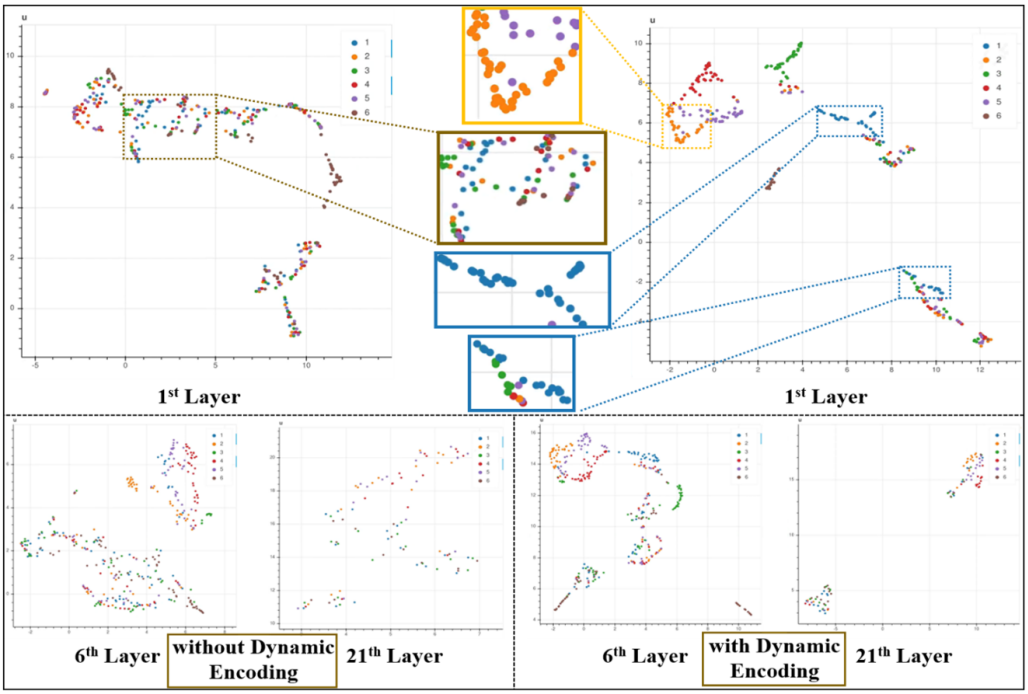

Figure 6. Comparison of data separability before and after adaptive encoding.

The experiments in this section prove that the adaptive encoding process can effectively improve the data separability of the input signal while similar structural features are also extracted through the original low-resolution speckle. The adaptive encoding provides a good data basis for the subsequent reconstruction of the network AESINet-R to recover detailed hidden target images.

3.4. Comparison Experiment of Reconstruction Effect under Different Encoding Methods

In this section, AESINet-R was used as the reconstruction network, and the face speckle data in Section 3.2 was used as the dataset. A comparative experiment with three modulation modes as variables is designed.

(1) Adaptive encoding modulation: AESINet-E is used to adaptively encode the original low-resolution speckle, and the modulated speckle signal is used as the input of AESINet-R to reconstruct the target as shown in Figure 7c;
(2) No encoding modulation: The original speckles are directly used for the training of AESINet-R, and the targets AESINet-R recovered are shown in Figure 7d;
(3) Random encoding modulation: The randomly generated gaussian encoding mask is employed to modulate the original speckle, and the modulated signal is used to train the AESINet-R. To avoid the influence of randomness on the experimental conclusions, three different sets of random masks are generated and tested. The reconstruction results are shown in Figure 7e,f, respectively.

Figure 7. Comparison of AESINet-R reconstruction results with different encoding methods. (**a**) The Speckle pattern. (**b**) The Ground Truth (GT) of the hidden target. (**c**) AESINet-R test set reconstruction results after adaptive encoding by AESINet-E. (**d**) The AESINet-R reconstruction result when the speckles without adaptive encoding as input. (**e**–**g**) The AESINet-R reconstruction result after speckle is modulated by randomly generated gaussian encoding mask.

A comprehensive comparison of the reconstruction results of several modulation methods in Figure 7 shows that adaptive encoding can help the network to reconstruct the resulting image with more accurate details. The most representative ones are the eyes and eyebrows of the face in Figure 7. Under the experiments of unmodulated and random encoded modulation, the eyes and eyebrows are integrated into a group of shadows in the restored image. With the help of adaptive encoding, the structure of eyes and eyebrows can be distinguished in the reconstructed results, as shown in Figure 7c.

The average MAE, SSIM, and PSNR of the recovered results of the test set under the three modulation modes are shown in Table 2. The results with adaptive encoding are better than the results of the non-encoding and random encoding. Compared with the method of directly using the original speckle as the input of the reconstruction network, the PSNR of the recovered target is increased 1.8 dB and the SSIM is increased 0.107 by the adaptive encoding method. Those results prove that the improvement of the data separability by the adaptive encoding technology does help the reconstruction network to extract effective features. The result in Table 2 fully explains the instability of random encoding, and also shows that this modulation method can not improve the quality of reconstructed images. In the three repeated experiments, the first random mask obtains the best indicators, but the quality of the reconstructed image is only roughly equal to that of the method without encoding modulation, and even the visual effect of some images in Figure 7e is reduced compared with that in Figure 7d.

Table 2. AESINet-R reconstruction results with different modulation methods.

Modulation Method	MAE	SSIM	PSNR (dB)
AESINet-E with adaptive encoding	0.0402	0.7680	24.1698
Without encoding	0.0513	0.6610	22.3664
1st random encoding	0.0555	0.6672	21.8219
2nd random encoding	0.0937	0.5862	18.2702
3rd random encoding	0.0847	0.6278	19.2093

The comparison results in Figure 7 and Table 2 fully show that the improvement of data separability by adaptive encoding can help the network obtain better reconstruction results.

4. Conclusions

This paper proposes an adaptive encoding method based on the physical characteristics of speckle redundancy. Experiments show that the proposed method can be restored with different complexity targets by using the 10nm bandwidth scattering signal. Adaptive encoding can not only improve the separability of the feature between different speckles, but also be more conducive to the reconstruction of the network to recover targets with more details. In addition, the proposed method does not require additional devices to modulate the system, which improves the reconstruction accuracy of target details without increasing the system complexity. In the future, we will further explore the influence of adaptive encoding on different speckle patterns and combine other physical characteristics of speckles to further optimize the neural network.

Author Contributions: Conceptualization, E.G.; methodology, E.G.; software, Y.S.; validation, E.G. and Y.S.; formal analysis, Y.S.; investigation, J.H.; resources, E.G., L.B. and J.H.; data curation, Y.S.; writing—original draft preparation, Y.S.; writing—review and editing, E.G. and J.H.; visualization, Y.S.; supervision, L.B. and J.H.; project administration, L.B.; funding acquisition, E.G., L.B. and J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Natural Science Foundation of China (62031018, 61971227, 62101255), China Postdoctoral Science Foundation (2021M701721), Fundamental Research Funds for the Central Universities (30920031101).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the compounds are not available from the authors.

References

1. Goodman, J.W. *Speckle Phenomena in Optics: Theory and Applications*; Roberts and Company Publishers: Greenwood Village, CO, USA, 2007.
2. Roggemann, M.; Welsh, B.; Hunt, B. *Imaging Through Turbulence, 320*; CRC Press: New York, NY, USA, 1996.
3. Wang, K.; Sun, W.; Richie, C.T.; Harvey, B.K.; Betzig, E.; Ji, N. Direct wavefront sensing for high-resolution in vivo imaging in scattering tissue. *Nat. Commun.* **2015**, *6*, 7276. [CrossRef] [PubMed]
4. Nixon, M.; Katz, O.; Small, E.; Bromberg, Y.; Friesem, A.A.; Silberberg, Y.; Davidson, N. Real-time wavefront shaping through scattering media by all-optical feedback. *Nat. Photonics* **2013**, *7*, 919–924. [CrossRef]
5. Mao, Y.; Flueraru, C.; Chang, S.; Popescu, D.P.; Sowa, M.G. High-quality tissue imaging using a catheter-based swept-source optical coherence tomography systems with an integrated semiconductor optical amplifier. *IEEE Trans. Instrum. Meas.* **2011**, *60*, 3376–3383.
6. Huang, D.; Swanson, E.A.; Lin, C.P.; Schuman, J.S.; Stinson, W.G.; Chang, W.; Hee, M.R.; Flotte, T.; Gregory, K.; Puliafito, C.A.; et al. Optical coherence tomography. *Science* **1991**, *254*, 1178–1181. [CrossRef]
7. Lu, D.; Liao, M.; He, W.; Cai, Z.; Peng, X. Imaging dynamic objects hidden behind scattering medium by retrieving the point spread function. In Proceedings of the Speckle 2018: VII International Conference on Speckle Metrology. International Society for Optics and Photonics, Janow Podlaski, Poland, 10–12 September 2018; Volume 10834, p. 1083428.
8. He, H.; Xie, X.; Liu, Y.; Liang, H.; Zhou, J. Exploiting the point spread function for optical imaging through a scattering medium based on deconvolution method. *J. Innov. Opt. Health Sci.* **2019**, *12*, 1930005. [CrossRef]
9. Drémeau, A.; Liutkus, A.; Martina, D.; Katz, O.; Schülke, C.; Krzakala, F.; Gigan, S.; Daudet, L. Reference-less measurement of the transmission matrix of a highly scattering material using a DMD and phase retrieval techniques. *Opt. Express* **2015**, *23*, 11898–11911. [CrossRef]
10. Popoff, S.M.; Lerosey, G.; Carminati, R.; Fink, M.; Boccara, A.C.; Gigan, S. Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media. *Phys. Rev. Lett.* **2010**, *104*, 100601. [CrossRef]
11. Kim, M.; Choi, W.; Choi, Y.; Yoon, C.; Choi, W. Transmission matrix of a scattering medium and its applications in biophotonics. *Opt. Express* **2015**, *23*, 12648–12668. [CrossRef]
12. Bertolotti, J.; Van Putten, E.G.; Blum, C.; Lagendijk, A.; Vos, W.L.; Mosk, A.P. Non-invasive imaging through opaque scattering layers. *Nature* **2012**, *491*, 232–234. [CrossRef]
13. Katz, O.; Heidmann, P.; Fink, M.; Gigan, S. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nat. Photonics* **2014**, *8*, 784–790. [CrossRef]
14. Zhu, S.; Guo, E.; Gu, J.; Bai, L.; Han, J. Imaging through unknown scattering media based on physics-informed learning. *Photonics Res.* **2021**, *9*, B210–B219. [CrossRef]
15. Li, Y.; Cheng, S.; Xue, Y.; Tian, L. Displacement-agnostic coherent imaging through scatter with an interpretable deep neural network. *Opt. Express* **2021**, *29*, 2244–2257. [CrossRef]
16. Borhani, N.; Kakkava, E.; Moser, C.; Psaltis, D. Learning to see through multimode fibers. *Optica* **2018**, *5*, 960–966. [CrossRef]
17. Guo, E.; Zhu, S.; Sun, Y.; Bai, L.; Zuo, C.; Han, J. Learning-based method to reconstruct complex targets through scattering medium beyond the memory effect. *Opt. express* **2020**, *28*, 2433–2446. [CrossRef]
18. Guo, E.; Sun, Y.; Zhu, S.; Zheng, D.; Zuo, C.; Bai, L.; Han, J. Single-shot color object reconstruction through scattering medium based on neural network. *Opt. Lasers Eng.* **2021**, *136*, 106310. [CrossRef]
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
20. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, UK, 2016; Volume 1.
21. Sun, Y.; Shi, J.; Sun, L.; Fan, J.; Zeng, G. Image reconstruction through dynamic scattering media based on deep learning. *Opt. Express* **2019**, *27*, 16032–16046. [CrossRef]
22. Horisaki, R.; Takagi, R.; Tanida, J. Learning-based imaging through scattering media. *Opt. Express* **2016**, *24*, 13738–13743. [CrossRef]
23. Li, S.; Deng, M.; Lee, J.; Sinha, A.; Barbastathis, G. Imaging through glass diffusers using densely connected convolutional networks. *Optica* **2018**, *5*, 803–813. [CrossRef]
24. Durán, V.; Soldevila, F.; Irles, E.; Clemente, P.; Tajahuerce, E.; Andrés, P.; Lancis, J. Compressive imaging in scattering media. *Opt. Express* **2015**, *23*, 14424–14433. [CrossRef]
25. Li, X.; Stevens, A.; Greenberg, J.A.; Gehm, M.E. Single-shot memory-effect video. *Sci. Rep.* **2018**, *8*, 13402. [CrossRef]
26. Lyu, M.; Wang, H.; Li, G.; Zheng, S.; Situ, G. Learning-based lensless imaging through optically thick scattering media. *Adv. Photonics* **2019**, *1*, 036002. [CrossRef]