LASER & PHOTONICS REVIEWS
www.lpr-journal.org

Extended Learning Robustness for High-Fidelity Human Face Imaging from Spatiotemporally Decorrelated Speckles

Qi Zhao, Huanhao Li, Tianting Zhong, Shengfu Cheng, Haofan Huang, Haoran Li, Jing Yao, Wenzhao Li, Chi Man Woo, Lei Gong, Yuanjin Zheng, Zhipeng Yu,* and Puxiang Lai**

Imaging within or through scattering media has long been a coveted yet challenging pursuit. Researchers have made significant progress in extracting target information from speckles, primarily by characterizing the transmission matrix of the scattering medium or employing neural networks. However, the fidelity of the retrieved images is compromised when the medium's status changes due to intrinsic motion or external perturbations. This variability leads to decorrelation between training and testing data, hindering the practical applications of these frameworks. In this study, we propose a generative adversarial network (GAN)-based framework with extended robustness, which is designed to address the spatiotemporal instabilities of scattering media and the resultant decorrelation between training and testing data. Experiments demonstrate that our GAN can retrieve high-fidelity face images from speckles, even when the scattering medium undergoes unknown changes after training. Notably, our GAN outperforms existing methods by non-holographically retrieving images from unstable scattering media and effectively addressing speckle decorrelation, even after prolonged system inactivity (up to 37 h in experiments, but can be longer if needed). This resilience opens venues for pre-trained networks to maintain effectiveness over time, and can broaden the scope of learning-based methodologies in deep tissue imaging and sensing under extreme environmental conditions.

1. Introduction

Optical imaging stands as a cornerstone in the exploration of the microscopic and macroscopic realms, ranging from cellular structures to whole bodies, this field has continually evolved, offering scalable resolution and expansive fields of view.[1,2] In environments like biological tissues or turbid media such as fog, dusty air, and turbid water, scattering predominantly governs light-matter interactions, curtailing imaging depth and compromising resolution. Coherent light exacerbates this challenge, producing speckles with stark contrasts that obscure clear imaging and interpretation.[3–5] Seminal efforts, particularly the transmission matrix approach, have addressed static scenarios by linearly correlating input and output fields.[6–10] While this enables computational retrieval of images from speckles via the inverse transmission matrix, the linear model's limitations often yield suboptimal imaging

Q. Zhao, H. Li, P. Lai
Bi-optoelectronic-integration and Medical Instrumentation Laboratory
Guangzhou Institute of Technology
Xidian University
Guangzhou, Guangdong 510555, China
E-mail: puxiang.lai@polyu.edu.hk

Q. Zhao, H. Li, T. Zhong, S. Cheng, H. Huang, H. Li, J. Yao, W. Li, C. M. Woo, Z. Yu, P. Lai
Department of Biomedical Engineering
Hong Kong Polytechnic University
Hong Kong 999077, China
E-mail: yu.zh.yu@polyu.edu.hk

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/lpor.202401836

DOI: 10.1002/lpor.202401836

Q. Zhao, H. Li, T. Zhong, S. Cheng, H. Huang, H. Li, J. Yao, W. Li, C. M. Woo, Z. Yu, P. Lai
Shenzhen Research Institute
Hong Kong Polytechnic University
Shenzhen 518063, China

Q. Zhao
Centre for Intelligent Multidimensional Data Analysis
Hong Kong 999077, China

T. Zhong
Department of Electrical and Electronic Engineering
The University of Hong Kong
Hong Kong 999077, China

L. Gong
Department of Optics and Optical Engineering
University of Science and Technology of China
Hefei 230026, China

quality due to the insufficient accuracy of the linear TM model.[11–12]

The emergence of deep learning has catalyzed a paradigm shift, with deep neural networks (DNNs) significantly refining optical system performance.[13] These networks account for the inherent nonlinearity in light-matter interactions and perturbations [14–20] resulting in enhanced imaging fidelity.[21–24] However, despite their success in stationary scattering environments, DNNs show low robustness and face challenges under non-stationary conditions, such as motion, perturbation, vibrations in the scattering medium or optical system. These non-stationary scattering media lead to speckle decorrelation, which deteriorates modulation or imaging performance.[25–30]

Researchers have designed different schemes to address challenges encountered in non-stationary scattering media like natural fog, turbid water, and multimode fibers.[31–35] While promising, existing DNNs primarily operate with simple objects (digits, letters, binary patterns) trained using extensive datasets that reflect different scattering medium statuses (i.e., the relationships between input and output wavefronts). Crucially, these datasets are often collected concurrently, exhibiting strong correlations between the training and testing data. However, practical applications often involve significant time gaps between training and testing, necessitating the use of data acquired under altered medium or system conditions. This temporal gap introduces decorrelation, diminishing the efficacy of pre-trained networks. Despite diverse training data representing various medium statuses, DNN performance degrades when confronted with changed conditions. Consequently, in real-world applications where the scattering medium deviates from previously encountered statuses, pre-trained networks struggle to adapt to decorrelated speckles.

In this study, we introduce a generative adversarial network (GAN) framework capable of training to recover complex face images from a dynamic scattering medium. The framework employs a U-Net-based generator to discern intrinsic speckle features, enabling high-fidelity retrieval of images.[36,37] Simultaneously, a multi-layer convolution-based discriminator assesses the retrieved images, providing critical feedback to refine the generator's outputs.[38,39] Notably, the GAN's generator harnesses speckle features for image retrieval, while the discriminator guides enhancements in the retrieved images. Experimental results affirm that the GAN retains temporal robustness, retrieving face images from speckles even with extended intervals between training and testing datasets, which results in minimal correlation. Further validation occurs through experiments using a non-stationary metasurface as the scattering medium. In this scenario, the network is initially trained with collected data, followed by deactivating the optical system for 37 h (or longer if necessary) and subsequently reactivating it to gather testing data, thereby simulating real-world conditions. To our knowledge, this marks the first instance of a neural network trained to retrieve high-fidelity face images from speckles acquired on different days from the training data. Overall, the proposed GAN demonstrates exceptional temporal robustness, overcoming challenges associated with changing speckles. Consequently, the method opens new venues for non-holographic retrieval of images from decorrelated speckles, signifying a significant advancement in the field.

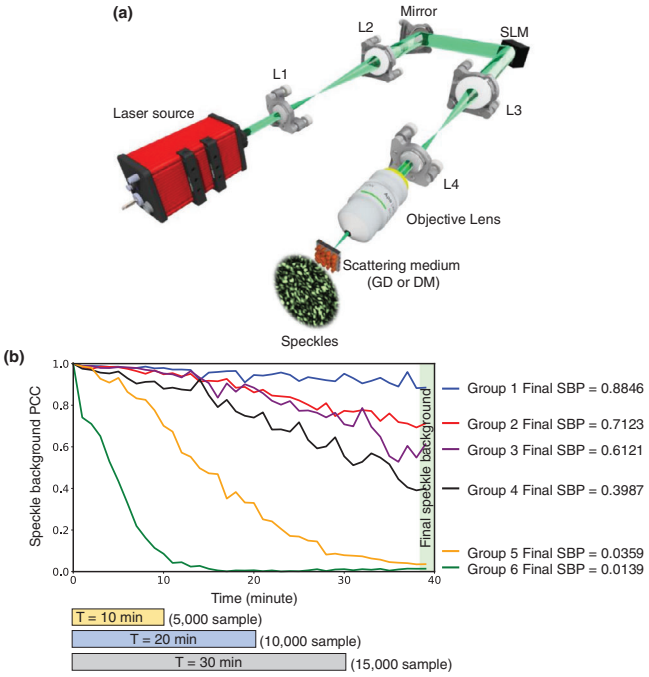

Figure 1. a) Diagram of the optical setup for acquiring speckles. L1 and L2: the first 4-f system to expand the laser beam. SLM: spatial light modulator. L3 and L4: the second 4-f system to shrink the laser beam. Scattering medium: ground glass (GD) or disordered metasurface (DM). b) Speckle background PCC (SBP) during six 40-min ground glass experiments. Lower SBP corresponds to a larger deviation from the initial status and lower stability. Final SBP is the SBP of each data group at the end (marked in green on the right *Y* axis).

2. Methods

2.1. Optical Setups

The optical system for acquiring speckles is depicted in **Figure 1a**. The light source is a continuous-wave 532-nm laser (EXLSR-532-300-CDRH, Single mode, 300 mW, Spectra-Physics, USA). The laser beam from the light source is first expanded by a 4-f system (L1 and L2) to cover the entire aperture of the spatial light modulator (SLM, HOLOEYE PLUTO VIS056 1080p, German). Then the image information displayed on the SLM modulates the wavefront. The information to modulate the input wavefront is the face image from thumbnails in the Flickr Faces High Quality (FFHQ) database.[40] Here, the 128 × 128 thumbnails are upsampled to 640 × 640 and displayed at the center of the SLM. After SLM modulation, the modulated light is shrunk by another

Y. Zheng
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore 639798, Singapore
E-mail: yjzheng@ntu.edu.sg
P. Lai
Photonics Research Institute
Hong Kong Polytechnic University
Hong Kong 999077, China

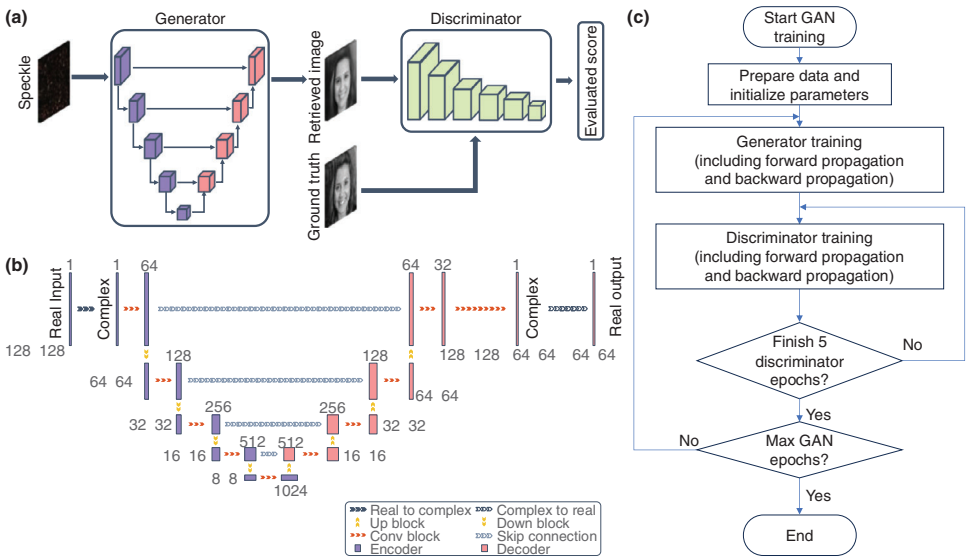

Figure 2. Schematic of the proposed GAN framework. a) GAN structure: the generator is based on U-Net, with speckle as the input and retrieved images as the output; the discriminator is based on six convolutional layers and one linear layer, with retrieved images or ground truth images as input and evaluated loss as output. Ground truth images: Copyright 2010, appnight-122, by Existence Church, Flickr (https://www.flickr.com/photos/sandiegochurch/4379311601/); the original images are cropped and converted to grayscale, under terms of the CC-BY 2.0 license. b) Structures of the generator: the encoders are highlighted in blue, the decoders are highlighted in red. The dimensions of the feature maps are specified next to each block. c) Flow chart of GAN training: during each GAN training epoch, the generator is trained only once, but the discriminator is trained five times to improve convergence and network performance.

4-f system (L3 and L4) and then focused by an objective lens (RMS20X, Olympus, Japan). The focused laser travels through a scattering medium, transforming into random optical speckles, which are captured by a CMOS camera (FL3-U3-32S2M-CS, PointGrey, Canada). The scattering medium used in experiments includes a ground glass diffuser and a disordered metasurface. During data collection, ambient perturbations (e.g., moving people, other ongoing experiments on the same optical table, running instruments in neighboring rooms, *etc.*) could mechanically shift the scattering medium, and light source stability depends on the working status of the laser (not intentionally controlled), leading to decorrelated speckles.

2.2. Data Acquisition and Speckle Instability

Experimentally, to generate data at different levels of instability, a scattering medium (i.e., ground glass) is constantly disturbed by surrounding perturbations, including air flow and platform vibrations *etc.* The corresponding instability is characterized by variations of background speckles generated by loading a uniform phase pattern every minute (i.e., every 500 captured speckles), in order to facilitate comparisons with related research (details in Figure S1, Supporting Information). In the ground glass experiments, the duration for the collection of each data group is up to 40 min. Every captured background speckle is compared with the initial background speckle for the calculation of the Pearson correlation coefficient (PCC), which is termed the speckle background PCC (SBP). As shown in Figure 1b, the SBP of six groups of data continuously decays with time in experiments due to environmental perturbations (including air flow, platform vibrations, *etc.*), a lower SBP corresponds to a higher deviation from the initial status and hence lower stability. With monotonic variations, we can mark SBP at the end of each group (i.e., Final SBP) to distinguish different groups. For example, for Group 1, Final SBP is 0.8846 (the highest among six groups), indicating relatively stationary conditions. For Group 6, SBP drops down more quickly. Final SBP is 0.0139, which is less than $1/e \approx 0.3678$, indicating speckles have become totally decorrelated and the dataset includes significantly perturbed information.[23,31] These datasets are available at Optical Speckle Datasets 2025 in PolyU Research Data Repository.[41] In the following experiments, these six groups of datasets are utilized to train and test the proposed GAN framework.

2.3. Neural Network Design

The proposed GAN includes a generator and a discriminator, as shown in **Figure 2**a, the related Python code is available on GitHub (https://github.com/863zq/863zq.github.io/blob/main/Code/Main_complex_GAN.py). The generator is based on an U-Net extensively employed in speckle imaging.[7,18,24,28] Compared with traditional U-Net,[36] the primary difference is that convolution layers here are all based on complex algebra, i.e., inputs, outputs, and parameters in convolutions are all complex-valued to more accurately mimic the random scattering process as modeled by the transmission matrix theory.[42] Here, the input of the U-Net-based generator is the speckle with 128×128 pixels, which is transferred into the complex domain with zero phase. Then, the input speckles can be processed by the generator's encoders (four down-sampling paths, blue filters in Figure b) and decoders (four up-sampling paths, red filters in Figure b). After that, the final layer converts the complex feature map into real numbers and

outputs the retrieved images with 64 × 64 pixels. In addition to the generator, a discriminator based on six convolutional layers is designed to evaluate the retrieved images from the generator. With the retrieved image or the ground truth as the input, the discriminator's output is the evaluated loss to discriminate between the retrieved images and the corresponding ground truth images, the generator tries to cheat the discriminator so that the retrieved image is refined to approach the corresponding ground truth image. After network training, the so-called Nash equilibrium is eventually reached, and the generator then successfully retrieves images from speckles with high fidelity.[38,39]

2.4. Network Training

The loss function of the generator L_G combines the image loss $L_{G,image}$ and the adversarial loss $L_{G,adv}$. Here, the image loss combines the mean square error MSE and the Pearson correlation coefficient PCC to reveal differences between the retrieved image $G(x)$ (i.e., the generator output when x is the input speckle) and the ground truth y. The adversarial loss $L_{G,adv}$ is MSE between the discriminator output with the generator output as the input, and the discriminator output with the corresponding ground truth as the input. Then, the final loss function for the generator L_G is a weighted average of the adversarial loss $L_{G,adv}$ and the image loss $L_{G,image}$. The weights in L_G here are empirically tuned to improve the performance of the generator:

$$L_{G,image} = MSE\left[G(x),\ y\right] - PCC\left[G(x),\ y\right] \tag{1}$$

$$L_{G,adv} = MSE\left[D(G(x)),\ D(y)\right] \tag{2}$$

$$L_G = 0.2 \times L_{G,adv} + 0.8 \times L_{G,image} \tag{3}$$

$$MSE\ (x, y) = (x - y)^2 \tag{4}$$

$$PCC\ (x, y) = \frac{\langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle}{\sigma_x \sigma_y} \tag{5}$$

In Equations (4)–(5), $\langle\rangle$ and σ denote the average operation and standard deviation, respectively. The proposed discriminator is designed to assess the generator output, the discriminator output is expected to be the image loss. The proposed discriminator is different from discriminators in traditional GAN to generate new images, which discriminate images and noises. In the proposed discriminator, the loss function L_D is the MSE between the predicted output of the discriminator, i.e., $((x))$, the real image loss $L_{G,image}(G(x),\ y)$:

$$L_D = MSE\left(D(G(x)),\ L_{G,image}(G(x),\ y)\right) \tag{6}$$

In the following experiments, each group in Figure 1b contains 20 000 image-speckle pairs, and six neural networks are individually trained with only one group of data. Human face images are collected from thumbnails in the FFHQ dataset,[40] from which 128 × 128 thumbnails are down-sampled to 64 × 64 as the ground truth images. The dimensions of the speckles fed into the generator are 128 × 128. As for training, GAN is trained for 20 epochs using Adam optimizers with batch size = 32, the initial learning rate is 0.0001 with cosine annealing. Furthermore, in each training epoch, the discriminator is trained five times (as illustrated in Figure 2c), while the generator is trained once to enhance the discrimination between retrieved images and ground truth images.[43]

Notably, speckles used during network testing are sampled after the training dataset collection is finished. As for different experiments, the acquisition time intervals between training and testing datasets vary, in order to test the network performance at different scattering medium statuses. As for the software framework, we utilize PyTorch 2.0.1 with CUDA 12.1 and Python 3.11.4, which is implemented on a Dell Precision Tower 5810 with Intel Xeon E5-1650 V3 CPU, 64 GB RAM, and a Nvidia GeForce RTX 3090 GPU. During network training, one epoch takes ≈20 min for a training dataset with 15,000 samples; the entire training process with 20 epochs takes ≈6 h.

3. Results

3.1. Imaging Through a Non-Stationary Diffuser

3.1.1. Qualitative Analysis

In order to evaluate the applicability of the proposed GAN in practical applications, we separate training and testing datasets with different time intervals, so that training and testing data acquisition windows do not overlap in time; the scattering medium statuses corresponding to the testing dataset (red points in **Figure 3a**, including 32 speckles) vary and decorrelate from those for the training dataset (corresponding to the training dataset durations T in Figure 3a). Three representative sets, Groups 1, 4, 6, are analyzed here, as shown in Figure 3a. The entire dataset duration (T) is 40 min, the data that cover the first 1/4 (T = 10 min), 1/2 (T = 20 min), 3/4 (T = 30 min) of the full dataset are respectively used as the training dataset, whose training sample amounts are 5,000, 10,000, 15,000 speckle-image pairs, respectively. Then, the testing dataset is sampled with different time intervals (Δt) after the training dataset (red points in Figure 3a), such as 1, 5, 10 min, etc.

Performance with T = 30 min (15 000 training samples) is illustrated in Figure 3b. As for Group 1 (Final SBP = 0.8846), PCC between the retrieved image and the corresponding ground truth in the testing dataset (called imPCC for simplicity) can reach 0.9665 when Δt = 1 min, fine features, such as eyes, eyebrows, noses, ears, mouths, cheeks, can be clearly seen with high fidelity. As Δt increases to 5 and 10 min, the visualized results are still retrieved with high fidelity (imPCC is 0.9526 and 0.9372, respectively), although SBP drops from 0.9493 (Δt = 1 min) to 0.9295 (Δt = 5 min) and 0.9009 (Δt = 10 min). With stronger decorrelation, i.e., Group 4 with Final SBP = 0.3987, the retrieval performance (Group 4 in Figure 3b) is comparable to Group 1, while the retrieved image PCC of Group 4 is just slightly lower than that of Group 1 due to non-stationary medium statuses. Under highly non-stationary conditions, i.e., Group 6 with Final SBP = 0.0139, the performance deteriorates with larger time intervals, imPCC is significantly lower than that in Groups 1 and 4.

With shorter training durations, e.g., T = 20 min with 10 000 training samples, although visually discernable, the overall imPCCs for Δt = 1, 5, 10 min in Figure 3c are systematically lower than their counterparts in Figure 3b. As for Figure 3c IV and V,

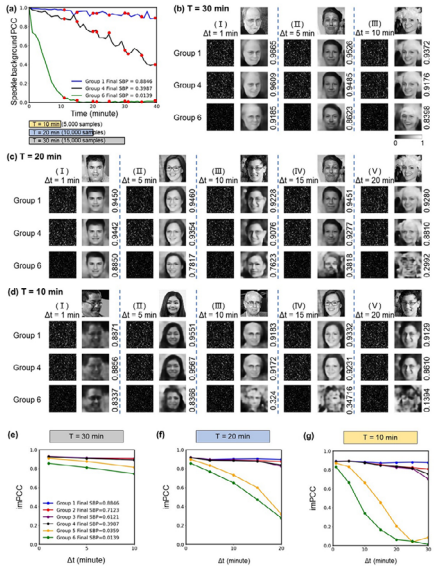

Figure 3. Retrieved images from speckles with different training durations T and different time intervals Δt between training and testing datasets. a) Datasets: Group 1, Group 4, Group 6 are divided according to training dataset durations T, including 10, 20, 30 min. The speckle background PCC of the testing datasets is marked as red points. b–d) Retrieved images from speckles: the top row of each column is the ground truth image; the right columns on other rows represent the corresponding retrieved images by inputting the speckles in the left columns into the generator of the GAN; the numbers next to retrieved images are PCCs between the retrieved images and the corresponding ground truth images; the ground truth images are selected from FFHQ dataset.[40] e–g) imPCC (average PCC between the retrieved images and the corresponding ground truths) versus different time intervals for training dataset duration T = 30, 20, 10 min, respectively. Ground truth image in (b-I): Copyright 2009, 100_1861, by Daniel Taylor, Flickr (https://www.flickr.com/photos/dtaylor404/3920892082/); the original images are cropped and converted to grayscale, under terms of the CC-BY 2.0 license. Ground truth images (b-II) and (c-IV): Copyright 2010, Living Room Comedy 101 052810, by Alex Erde, Flickr (https://www.flickr.com/photos/alexerde/4682485372/); the original images are cropped and converted to grayscale, under terms of the CC-BY 2.0 license. Ground truth images (b-III) and (c-V): Copyright 2015, Naughty Snowball IV-128, by PJ Rey, Flickr (https://www.flickr.com/photos/pjrey/8261366051/); the original images are cropped and converted to grayscale, under terms of the CC-BY 2.0 license. Ground truth images (c-I): Copyright 2016, Fr15, by Európa Pont, Flickr (https://www.flickr.com/photos/europapont/26666567983/); the original images are cropped and converted to grayscale, under terms of the CC-BY 2.0 license. Ground truth images (c-II) and (d-IV): Copyright 2017, Henriette Sørbøe Andreassen, by Senterpartiet (Sp) (https://www.flickr.com/photos/senterpartiet/32951636095/); the original images are cropped and converted to grayscale, under terms of the CC-BY 2.0 license. Ground truth images (c-III) and (d-V): Copyright 2009, them again, by Josh, Flickr (https://www.flickr.com/photos/iheartphoto/204002570/); the original images are cropped and converted to grayscale, under terms of the CC-BY 2.0 license. Ground truth image (d-I): Copyright 2014, Demand14 Speakers, by Cision Global, Flickr (https://www.flickr.com/photos/vocus/14439031135/); the original images are cropped and converted to gray-scale, under terms of the CC-BY 2.0 license. Ground truth image (d-II): Copyright 2018, by sirokame000, Flickr (https://www.flickr.com/photos/95949782@N03/27719339539/); the original images are cropped and converted to gray-scale, under terms of the CC-BY 2.0 license. Ground truth image (d-III): Copyright 2018, Mosman Faces launch, by Mosman Library, Flickr (https://www.flickr.com/photos/mosmanlibrary/5760074593/); the original images are cropped and converted to gray-scale, under terms of the CC-BY 2.0 license.

when the time intervals are respectively extended to 15 min and 20 min, the testing datasets decorrelate more from the training datasets. As a result, the retrieved images appear blurrier than before. Especially for Group 6 in Figure 3c IV and V, imPCC is as low as 0.3818 and 0.2992, and important facial features (such as eyes and cheeks) become more obscure. When the training duration is even shortened to 10 min with merely 5,000 training samples, as shown in Figure 3d, the proposed GAN for Groups 1 and 4 can mostly retain imPCCs above 0.86. While for Group 6, imPCC significantly drops below 0.35 when $\Delta t \geq 10$ min (SBP reduces to 0.0064), the testing dataset from Group 6 can be regarded to be totally decorrelated from the training dataset.

3.1.2. Quantitative Analysis

Quantitatively, performances from six groups of experiments (i.e., imPCC) with different time-varied divisions based on the training data duration T are further investigated to assess temporal generalization of the proposed GAN, as shown in Figure 3e–g. Generally, a larger time interval between training and testing datasets (Δt) corresponds to a lower instant SBP as the background speckles change continuously for each group. For Groups 1–4 with Final SBPs above 0.3987, the proposed GAN is able to retrieve images from speckles with high fidelity. When $\Delta t \leq 10$ min, for T = 30 min with 15 000 image-speckle pairs (Figure 3e), imPCCs are relatively stable (above 0.8910) and can hardly be differentiated. Similar observations can also be discovered in Figure 3f (T = 20 min with 10 000 training samples) and Figure 3g (T = 10 min with 5,000 training samples), though their overall imPCCs are lower than their counterparts in Figure 3e. With larger time intervals, i.e., $\Delta t \geq 10$ min, mild descending trends are seen in Figure 3f,g, imPCCs for Groups 3 and 4 decrease more rapidly than those for Groups 1 and 2. For more severely decorrelating datasets, i.e., Groups 5 and 6 (Final SBP = 0.0359 and 0.0139, respectively), imPCCs drop much more apparently than other groups.

Overall, the robustness of the trained neural network is improved by learning more scattering medium statuses, leading to better retrieved image qualities after speckle decorrelation. However, more scattering medium statuses also correspond to higher decorrelation in the speckles, which can affect the retrieved images (details in Figure S2, Supporting Information). The proposed GAN is designed to improve the robustness of the trained neural network to balance the scattering medium statuses and the retrieved image qualities, resulting in better robustness and improved retrieved image qualities.

3.2. Imaging Through a Disordered Metasurface

In the ground glass-based experiments, the proposed framework can overcome moderate decorrelation of the scattering medium, with the longest time interval (Δt) between training and testing datasets being up to 30 min. Next, we will test whether the proposed framework is able to generalize its ability for even longer periods. A disordered metasurface with random phase profiles is used as the scattering medium,[44–46] the ground truth images are human face images selected from the Large-scale Celeb

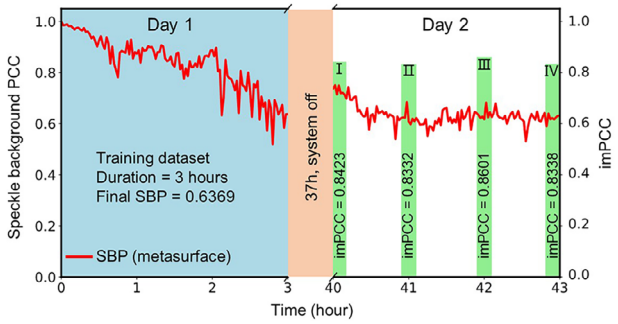

Figure 4. Metasurface experimental results. The speckle background PCC (SBP, the red solid curves) on Day 1 and Day 2, between which the optical system is turned off for 37 h. The network is trained based on the data acquired during the first 3 h on Day 1, with Final SBP = 0.6369 (marked in light blue), containing 60 000 speckle-image pairs, leading to 24 h of training time. The testing dataset is acquired on Day 2, with the representative imPCCs at $\Delta t = 37, 38, 39, 40$ h being represented by the green bars. Note: The ground truth images (human faces) and retrieved images cannot be published due to copyright licensing restrictions of the dataset used. Please refer to the data availability statement.

Faces Attributes (CelebA) Dataset.[47] The fabrication process of the metasurface involves a combination of electron beam lithography (EBL) and reactive ion etching (RIE). The detailed fabrication process is provided in Figure S3 (Supporting Information).

The results of the metasurface experiments are illustrated in **Figure 4**. As seen, although SBP on Day 1 drops down to 0.6 after 3 h, the overall stability (SBP) fluctuates from 0.6 to 0.7 on Day 2, even if the optical system (including the laser source and all electronic devices) is shut off for 37 h. During the inactive period, the optical system has experienced complex perturbations, including but not limited to air flow, vibrations of setups, vibrations of the building, and other optical experiments. Subsequently, the optical system is reactivated to gather testing data. It is worth noting that the network is trained based on the data acquired on Day 1, but the proposed framework can still effectively retrieve the face images from speckles obtained on Day 2. All this indicates the temporal generalization capability of the proposed framework.

4. Discussions

Although the diffuser remains unchanged during each group of data collection, environmental perturbations vary across different days and moments, which is a common challenge in optical experiments. To address this, we have demonstrated that the proposed framework possesses sufficient spatiotemporal robustness to adapt to scattering medium statuses that are not present in the training data. This feature is enabled by training the network with a diverse set of statuses from the scattering medium, allowing it to extract the underlying speckle features.

In experiments, we measure the quantified perturbation of a non-stationary scattering medium based on the speckle background PCC (SBP). Here, SBP is derived from speckles resulting from loading a uniform phase pattern on the SLM and can be used to evaluate the influence of scattering medium instabilities. Then, complex and dynamic relations between speckle intensities and input images are learned by the proposed GAN, so that the generator extracts inherent features from speckles through complex-valued convolutions, which is able to mimic the random scattering process accurately as modeled by the transmission matrix theory (performances of real-valued GAN and complex-valued GAN are compared in Figure S4, Supporting Information). The discriminator evaluates the retrieved images to help the generator further improve outputs, so that the retrieved images are continually refined to approach the corresponding ground truth image during network training.

Notably, after the metasurface experiments (fluctuating SBP on Day 2 in Figure 3a), we can conclude that the proposed GAN successfully overcomes speckle decorrelation even for a time interval $\Delta t > 37$ h (or longer, if needed). To the best of our knowledge, this is the first research in speckle imaging of complex objects (e.g., grayscale human face images with detailed features) whose network training and testing are performed on various days with significantly different or unpredictable system statuses. This capability opens new venues for applications where networks can be trained in advance but maintain their validity for data acquired later, regardless of the status of the medium. This is essential for extending the impact of the learning-based approaches in the field.

Additionally, due to its robustness across different scattering media (ground glass and metasurface in experiments), the proposed GAN can be further extended to imaging through deep biological tissues with strong volume scattering. This is a long-desired yet considered challenge, since the optical field decorrelates rapidly on the order of milliseconds due to physiological motions such as breathing, heartbeat, and blood flow.[20,48] With the proposed GAN-based framework, one can train the network with datasets that cover more statuses of the sample.

Furthermore, the performance of the proposed GAN can be enhanced by incorporating additional encoded speckle information during the training stage or by adopting an optical neural network for faster processing.[49] Retrieving high-quality images from speckles remains challenging, particularly when speckles significantly decorrelate from the training data. To address this, including more training speckle samples with varying scattering medium statuses generally leads to better-retrieved images.

Additionally, speckle data with a larger region of interest (ROI), containing more speckle grains, provides more access to additional encoded information. However, this improvement comes at the cost of longer training time (details in Figure S5, Supporting Information). Alternatively, for less complicated target information, such as binary digits and letters,[50,51] using a smaller training dataset and a lighter network structure can reduce the time and computing resources required for network training. For instance, **Figure 5** demonstrates that even with T = 10 min and $\Delta t = 20$ min, digits can be retrieved from speckles with high fidelity. Moreover, exploiting the spatial redundancy of target information encoded in the speckle field allows down-sampling[19] or using a partial field of view of the speckles.[32] These could significantly reduce the size of the training data, thereby accelerating the learning process of the network.

These enhancements hold significant promise for practical applications of speckle imaging, particularly when utilizing a disordered metasurface as the scattering medium. For example, by training a GAN with speckles from different system conditions, we can create a pre-trained GAN capable of retrieving information from optical speckles acquired later—even when the

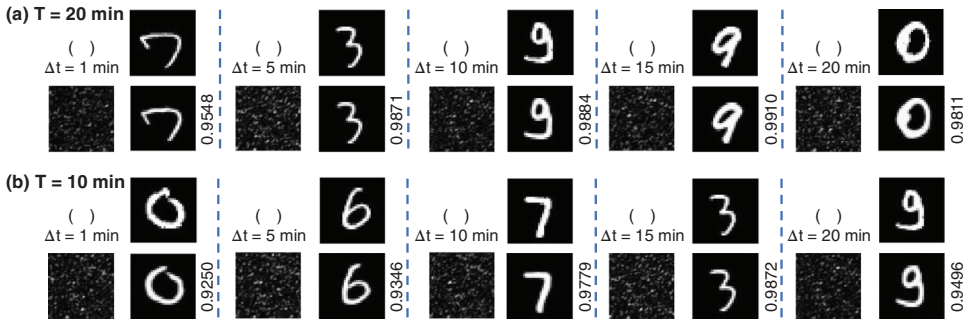

Figure 5. Digit-induced speckle experiments with different training durations (T = 20 and 10 min) and different time intervals (Δt = 1, 5, 10, 15, 20 min) between training and testing datasets: the top row of each column is the ground truth image; the right columns on other rows represent the corresponding retrieved images by inputting the speckles in the left columns into the generator of the GAN; the numbers next to the retrieved images are PCCs between the retrieved images and the corresponding ground truth images.

medium's statuses are unknown to the pre-trained model. The robustness of the network against spatiotemporal decorrelation could greatly enhance the accuracy of perception systems, allowing them to adapt to various scenarios.

5. Conclusion

In summary, we propose a GAN-based learning framework designed for non-holographic image retrieval from spatiotemporally decorrelated speckles acquired in a non-stationary scattering medium. Our framework's extended robustness is convincingly demonstrated through experiments using ground glass as the scattering medium. Despite the inherent non-stationarity of the medium and the separation of data into distinct training periods, our approach successfully retrieves high-fidelity face images from speckles. Furthermore, when employing a disordered metasurface as the scattering medium, our framework remains effective even when tested with images acquired on the second day of network training—after a 37-h optical system downtime. These results underscore the exceptional spatiotemporal decorrelation robustness of our proposed GAN-based framework. It enables high-fidelity retrieval of complex information (e.g., human face images) from speckles acquired under varying media or system statuses. Importantly, our framework's ability to train in advance and adapt to input data acquired later holds significant promise. It opens avenues for learning-based approaches in diverse speckle imaging scenarios, including non-holographic imaging through biological tissues and target sensing under dynamic environmental conditions.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Q.Z., H.L., and T.Z. contribute equally to this work. The authors would like to thank the Photonics Research Institute of Hong Kong Polytechnic University for facilities and technical support. This work is supported by National Natural Science Foundation of China (NSFC) (81930048), Hong Kong Research Grant Council (C7074-21GF, 15125724), Guangdong Science and Technology Commission (2019BT02×105), Shenzhen Science and Technology Innovation Commission (JCYJ20220818100202005), Hong Kong Polytechnic University (P0045680, P0043485, P0045762, P0049101), Postdoctoral Fellowship Program of China Postdoctoral Science Foundation (CPSF) (No. GZB20240593) and the Fundamental Research Funds for the Central Universities (QTZX25121).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The speckle data in Figures 1 and 3 are openly available in Optical Speckle Datasets 2025 at https://doi.org/10.60933/PRDR/FY7EDJ, reference number [41]. Other data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

deep learning, diffused light, high-fidelity imaging, speckle decorrelation, wavefront shaping

Received: October 30, 2024
Revised: March 10, 2025
Published online:

[1] S. H. Yun, S. J. Kwok, *Nat. Biomed. Eng.* **2017**, *1*, 0008.
[2] L. Zhu, F. Soldevila, C. Moretti, A. d'Arco, A. Boniface, X. Shao, H. B. de Aguiar, S. Gigan, *Nat. Commun.* **2022**, *13*, 1447.
[3] Y. Luo, S. Yan, H. Li, P. Lai, Y. Zheng, *Photonics Res.* **2021**, *9*, 262.
[4] Q. Zhao, C. M. Woo, H. Li, T. Zhong, Z. Yu, P. Lai, *Opt. Lett.* **2021**, *46*, 2880.
[5] H. Li, Z. Yu, T. Zhong, S. Cheng, P. Lai, *Adv. Photonics* **2023**, *5*, 020502.
[6] S. Turtaev, I. T. Leite, T. Altwegg-Boussac, J. M. P. Pakan, N. L. Rochefort, T. Čižmár, *Light Sci. Appl.* **2018**, *7*, 92.
[7] L. Wang, T. Qi, Z. Liu, Y. Meng, D. Li, P. Yan, M. Gong, Q. Xiao, *APL Photon* **2022**, *7*, 106104.
[8] D. S. Wiersma, *Nat. Photon.* **2013**, *7*, 188.

[9] S. M. Popoff, G. Lerosey, R. Carminati, M. Fink, A. C. Boccara, S. Gigan, *Phys. Rev. Lett.* **2010**, *104*, 100601.
[10] Y. Huang, W. Zhao, A. Zhai, D. Wang, *Laser Photonics Rev.* **2024**, *18*, 2301020.
[11] H. Lee, S. Yoon, P. Loohuis, J. H. Hong, S. Kang, W. Choi, *Light Sci. Appl.* **2022**, *11*, 16.
[12] H. Li, S. Zheng, Z. Wang, T. Zhong, C. M. Woo, S. Cheng, M. He, M. Lei, Z. Qiu, P. Lai, *Opt. Laser Technol.* **2025**, *180*, 111423.
[13] X. Zhang, J. Gao, Y. Gan, C. Song, D. Zhang, S. Zhuang, S. Han, P. Lai, H. Liu, *PhotoniX* **2023**, *4*, 10.
[14] L. Tian, B. Hunt, M. A. L. Bell, J. Yi, J. T. Smith, M. Ochoa, X. Intes, N. J. Durr, *Lasers. Surg. Med.* **2021**, *53*, 748.
[15] A. Turpin, I. Vishniakou, J. D. Seelig, *Opt. Express* **2018**, *26*, 30911.
[16] A. Doronin, I. Meglinski, *Biomed. Opt. Express* **2011**, *2*, 2461.
[17] R. Horisaki, R. Takagi, J. Tanida, *Opt. Express* **2016**, *24*, 13738.
[18] P. Tang, K. Zheng, W. Yuan, T. Pan, Y. Xu, S. Fu, Y. Wang, Y. Qin, *Appl. Phys. Lett.* **2022**, *121*, 081107.
[19] H. Li, Z. Yu, Q. Zhao, Y. Luo, S. Cheng, T. Zhong, C. M. Woo, H. Liu, L. V. Wang, Y. Zheng, P. Lai, *Photonics Res.* **2023**, *11*, 631.
[20] S. Cheng, H. Li, Y. Luo, Y. Zheng, P. Lai, *J. Innov. Opt. Health Sci.* **2019**, *12*, 1930006.
[21] Z. Yu, H. Li, T. Zhong, J.-H. Park, S. Cheng, C. M. Woo, Q. Zhao, J. Yao, Y. Zhou, X. Huang, W. Pang, H. Yoon, Y. Shen, H. Liu, Y. Zheng, Y. Park, L. V. Wang, P. Lai, *The Innovation* **2022**, *3*, 623.
[22] A. d'Arco, F. Xia, A. Boniface, J. Dong, S. Gigan, *Opt. Express* **2022**, *30*, 30845.
[23] Y. Li, Y. Xue, L. Tian, *Optica* **2018**, *5*, 1181.
[24] Q. Zhao, H. Li, Z. Yu, C. M. Woo, T. Zhong, S. Cheng, Y. Zheng, H. Liu, J. Tian, P. Lai, *Adv. Sci.* **2022**, *9*, 2202407.
[25] A. M. Caravaca-Aguirre, R. Piestun, *Opt. Express* **2017**, *25*, 1656.
[26] S. A. Vasquez-Lopez, R. Turcotte, V. Koren, M. Plöschner, Z. Padamsey, M. J. Booth, T. Čižmár, N. J. Emptage, *Light Sci. Appl.* **2018**, *7*, 110.
[27] H. Li, C. M. Woo, T. Zhong, Z. Yu, Y. Luo, Y. Zheng, X. Yang, H. Hui, P. Lai, *Photonics Res.* **2021**, *9*, 202.
[28] G. Wu, Y. Sun, L. Yin, Z. Song, W. Yu, *Opt. Lett.* **2023**, *48*, 2764.
[29] M. Lyu, H. Wang, G. Li, S. Zheng, G. Situ, *Adv. Photonics* **2019**, *1*, 036002.
[30] D. Wang, S. K. Sahoo, X. Zhu, G. Adamo, C. Dang, *Nat. Commun.* **2021**, *12*, 3150.
[31] S. Zheng, H. Wang, S. Dong, F. Wang, G. Situ, *Photonics Res.* **2021**, *9*, 220.
[32] S. Resisi, S. M. Popoff, Y. Bromberg, *Laser Photonics Rev.* **2021**, *15*, 2000553.
[33] Y. Bian, F. Wang, Y. Wang, Z. Fu, H. Liu, H. Yuan, G. Situ, *Photonics Res.* **2023**, *12*, 134.
[34] H. Liu, F. Wang, Y. Jin, X. Z. Ma, S. T. Li, Y. M. Bian, G. Situ, *Light Sci. Appl.* **2024**, *13*, 194.
[35] Z. Li, W. Zhou, Z. Zhou, S. Zhang, J. Shi, C. Shen, J. Zhang, N. Chi, Q. Dai, *Nat. Commun.* **2024**, *15*, 1498.
[36] O. Ronneberger, P. Fischer, T. Brox, *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, Springer, Cham, **2015**, 234.
[37] H. Huang, Q. Zhao, H. Li, Y. Zheng, Z. Yu, T. Zhong, S. Cheng, C. M. Woo, Y. Gao, H. Liu, Y. Zheng, J. Tian, P. Lai, *Adv. Intell. Syst.* **2024**, *6*, 2400150.
[38] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, *Commun. ACM* **2020**, *63*, 139.
[39] A. Odena, C. Olah, J. Shlens, *Proceed. 34th Int. Conf. Mach. Learn.* **2017**, *70*, 2642.
[40] T. Karras, S. Laine, T. Aila, *Proceed. IEEE/CVF Conf. Computer Vis. Pattern Recognit.* **2019**, 4401.
[41] Q. Zhao, H. Li, T. Zhong, Z. Yu, P. Lai;, Optical Speckle Datasets **2025**; PolyU Research Data Repository; Version: V1, https://doi.org/10.60933/PRDR/FY7EDJ.
[42] P. Caramazza, O. Moran, R. Murray-Smith, D. Faccio, *Nat. Commun.* **2019**, *10*, 2029.
[43] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, *Adv. Neural Inform. Process. Syst.* **2014**, *27*, 1.
[44] Z. Yu, X. Gao, J. Yao, H. Li, Y. Shi, B. Li, Z. Xie, X. Yuan, P. Lai, Q. Song, *Adv. Mater.* **2024**, *37*, 2415142.
[45] Z. Yu, H. Li, W. Zhao, P. S. Huang, Y. T. Lin, J. Yao, W. Li, Q. Zhao, P. C. Wu, P. Genevet, Q. Song, P. Lai, *Nat. Commun.* **2024**, *15*, 2607.
[46] Z. Yu, X. Gao, J. Yao, Z. Wang, T. Zhong, Y. Shi, B. Li, P. Lai, X. Li, Q. Song, *eLight* **2024**, *4*, 21.
[47] Z. W. Liu, P. Luo, X. G. Wang, X. O. Tang, *IEEE I. Conf. Comp. Vis.* **2015**, 3730.
[48] S. Zheng, W. Li, W. Pang, T. Zhong, P. Lai, *Adv. Photonics* **2024**, *6*, 060504.
[49] H. Li, Z. Yu, Q. Zhao, T. Zhong, P. Lai, *The Innov.* **2022**, *3*, 100252.
[50] P. Fan, M. Ruddlesden, Y. Wang, L. Zhao, C. Lu, L. Su, *Laser Photonics Rev.* **2021**, *15*, 2000348.
[51] C. Y. Zhu, E. A. Chan, Y. Wang, W. N. Peng, R. X. Guo, B. L. Zhang, C. Soci, Y. D. Chong, *Sci. Rep.* **2021**, *11*, 896.