

# Harnessing the magic of light: spatial coherence instructed swin transformer for universal holographic imaging

Xin Tong<sup>1,2</sup>, Renjun Xu<sup>2</sup>, Pengfei Xu<sup>1</sup>, Zishuai Zeng<sup>1</sup>, Shuxi Liu<sup>1</sup>, and Daomu Zhao<sup>1,2,\*</sup>

<sup>1</sup>Zhejiang University, School of Physics, Zhejiang Province Key Laboratory of Quantum Technology and Device, Hangzhou, China

<sup>2</sup>Zhejiang University, Center for Data Science, Hangzhou, China

**Abstract.** Holographic imaging poses significant challenges when facing real-time disturbances introduced by dynamic environments. The existing deep-learning methods for holographic imaging often depend solely on the specific condition based on the given data distributions, thus hindering their generalization across multiple scenes. One critical problem is how to guarantee the alignment between any given downstream tasks and pretrained models. We analyze the physical mechanism of image degradation caused by turbulence and innovatively propose a swin transformer-based method, termed train-with-coherence-swin (TWC-Swin) transformer, which uses spatial coherence (SC) as an adaptable physical prior information to precisely align image restoration tasks in the arbitrary turbulent scene. The light-processing system (LPR) we designed enables manipulation of SC and simulation of any turbulence. Qualitative and quantitative evaluations demonstrate that the TWC-Swin method presents superiority over traditional convolution frameworks and realizes image restoration under various turbulences, which suggests its robustness, powerful generalization capabilities, and adaptability to unknown environments. Our research reveals the significance of physical prior information in the optical intersection and provides an effective solution for model-to-tasks alignment schemes, which will help to unlock the full potential of deep learning for all-weather optical imaging across terrestrial, marine, and aerial domains.

Keywords: spatial coherence; holographic imaging; turbulence; image restoration; deep learning.

Received Jul. 10, 2023; revised manuscript received Sep. 5, 2023; accepted for publication Sep. 26, 2023; published online Oct. 25, 2023.

© The Authors. Published by SPIE and CLP under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

[DOI: [10.1117/1.AP.5.6.066003](https://doi.org/10.1117/1.AP.5.6.066003)]

## 1 Introduction

Holographic imaging is an interdisciplinary field that combines optics, computer science, and applied mathematics to generate holographic images using numerical algorithms. Although the concept of using computers to generate holograms can be traced back to the 1960s, it was not until the emergence of digital imaging and processing techniques in the 1990s that computational holography began to develop into a viable technology.<sup>1,2</sup> In the 1990s, digital holography started to gain more attention due to advancements in computer technology and digital image processing.<sup>3</sup> In recent years, holographic imaging has continued to advance, with new research and technology enabling even

more sophisticated holographic imaging capabilities. Researchers have developed increasingly sophisticated numerical algorithms for holographic imaging, such as compressive sensing, sparse coding, and deep-learning techniques.<sup>4-10</sup>

Spatial coherence (SC) is a critical factor that determines the quantity and quality of high-frequency information carried by the light beam in holographic imaging. High-frequency information is crucial for achieving high resolution and capturing fine details in an image. When the SC of the light source is low, the phase relationship of the beam becomes chaotic, causing the interference pattern to be washed out and resulting in insufficient transmission of high-frequency information. As a result, the reconstructed image has a lower resolution and less fine-detail information, as the high-frequency information needed to capture these details has been lost. Therefore, high

\*Address all correspondence to Daomu Zhao, [dmz123@zju.edu.cn](mailto:dmz123@zju.edu.cn)

SC light is preferred for holographic imaging to ensure that sufficient high-frequency information is present in the interference pattern and the hologram, resulting in high-resolution and detailed reconstructed images. However, the SC of light sources is often very low in complex scenes, which leads to image degradation and loss of details. Therefore, how to restore images under low-SC light sources is a challenging issue.<sup>11–15</sup>

Oceanic and atmospheric turbulence may profoundly influence optical imaging, engendering distortions and deterioration in photographs acquired through cameras and alternative optical detection devices. The distortion and degradation of images caused by oceanic turbulence occur because the turbulent motions in the water column cause variations in the refractive index of the water, which in turn leads to variations in the path of light as it travels through the water. Atmospheric turbulence occurs because the Earth's atmosphere is not uniform and contains regions of varying temperature and density, which can cause variations in the refractive index of the air. Whether it is oceanic turbulence or atmospheric turbulence, as the beam passes through these regions of varying refractive index, phase correlation changes, and the SC is distorted, causing the image to become blurred and distorted, or even completely lost. Massive efforts were devoted to finding a solution for imaging in various turbulences.<sup>16–23</sup> There is no denying the fact that it is difficult to use the same methods to simultaneously resolve holographic imaging problems with low-SC scenes and multiple intensities of turbulence. Although low-SC and turbulence may not appear to be correlated at first glance, their influence on computational holography can both be described through the concept of SC. As a result, we can transform the aforementioned issues into the imaging problem of different SCs and leverage the advantages of deep learning to train a generalized model that can achieve image restoration for any turbulence intensity and low SC.

Artificial intelligence for optics has unparalleled advantages, especially in the field of holography. For example, deep learning can address challenging inverse problems in holographic imaging, where the objective is to recover the original scene or object properties from observed images or measurements and enhance the resolution of optical imaging systems beyond their traditional diffraction limit,<sup>24–30</sup> etc. Intersection research of optics and deep learning aims to solve massive tasks with one model, and one important problem is how to guarantee the alignment between the distribution of any given downstream data and tasks with pre-trained models. This means that the same model and weights can only be applied to a specific environment. Our research uses SC as adaptable real-time physical prior information to precisely align any scenes with pretrained models. By combining the most advanced deep-learning algorithms, residual network,<sup>31</sup> and swin transformer,<sup>32</sup> we proposed our deep-learning-based methodology, termed as train-with-coherence-swin (TWC-Swin) method. It can achieve the restoration of computational holographic imaging under any low SC and turbulence.

We summarize the innovations of this paper as follows.

(1) We designed an LPR to simultaneously acquire two outputs: computational holographic imaging results and corresponding interference fringes under different SCs and turbulent scenes [Fig. 1(a)]. We manipulated SC by changing the distance between lens 1 and the rotating diffuser (RD). The SC can be calculated using the obtained interference fringes, while the imaging results serve as training and testing data for the neural network. In our experiment, we employed partially coherent

light as the light source and loaded the turbulence phase using a spatial light modulator (SLM) with a loading frequency of 20 Hz. The original data sets consist of natural images characterized by complex elements and low sparsity, rather than simple symbols or letters.

(2) The core of the TWC-Swin method is a swin adapter and swin-model network [Fig. 1(b)]. The swin adapter can select the optimal model from the model space by obtaining SC. The architecture of the swin model in model space is the same; only the weights are different [Figs. 1(c) and 1(d)]. Table S1 in the [Supplementary Material](#) provides the correspondence between SC and swin-model space. These weights are obtained by network training at different SCs. Our swin model utilizes the swin-transformer algorithm and incorporates local–global convolution residuals to construct its architecture. Additionally, the preprocessing and postprocessing modules are exclusively based on convolutions. The detailed internal architecture of the swin model is shown in Fig. S1 in the [Supplementary Material](#). Compared to the convolutional neural network and pure swin-transformer frameworks, the swin model exhibits superior performance. In addition, the TWC-Swin method takes about 28 ms to process a picture, which can achieve real-time video-level restoration.

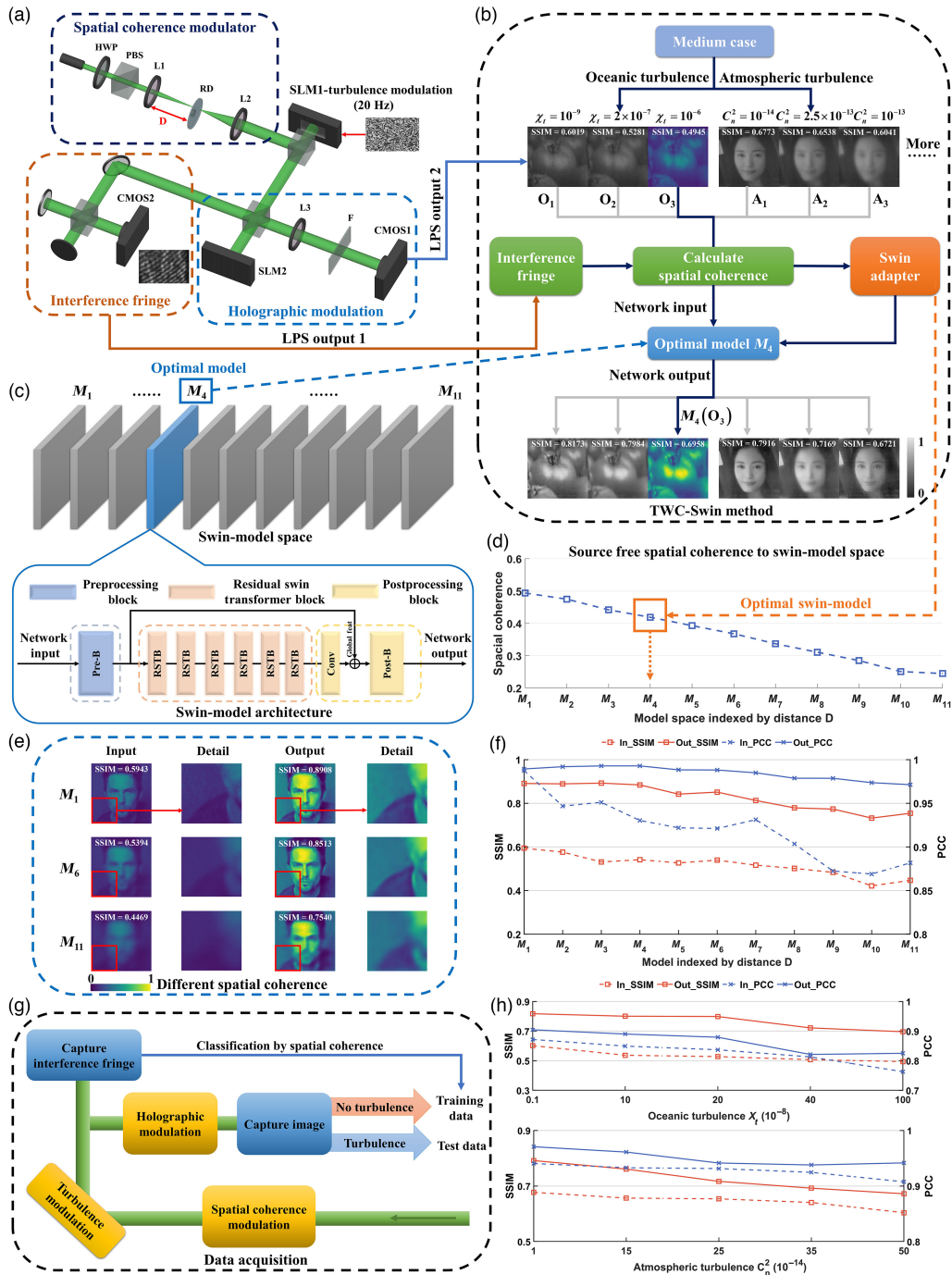
(3) Our model was only trained on different SCs and performed well on turbulence data, despite not being trained on such data. We can effectively restore holographic images under various turbulent scenes using a swin adapter. This suggests that our method has strong generalization capability and learned universal features of image degradation and restoration during training that are applicable beyond the specific conditions of the training data.

(4) Our strategy for training neural networks with SC as the physical prior information is applicable to arbitrary neural networks not limited to the swin model. It showcases the advantage of “transferability,” meaning that it can be transferred or applied effectively across different neural network architectures and tasks. Moreover, we achieve model-to-task alignment using physical principles and demonstrate the strategy's ability to incorporate physical principles into the learning process, which can be particularly beneficial for tasks involving physical data or phenomena. This lends a degree of interpretability to the model, as it leverages known physical principles to inform its predictions.

## 2 Materials and Methods

### 2.1 Scheme of the LPR

Figure 1(a) shows the LPR. The high-coherence light source generated by the solid-state laser (CNI, MLL-FN, 532 nm) is polarized horizontally after passing through a half-wave plate and a polarization beam splitter, allowing it not only to match the modulation mode of the SLM but also to adjust the beam intensity. The RD (DHC, GCL-201) is used to reduce the SC of the light source, with the degree of reduction depending on the radius of the incident beam on the RD—the larger the radius is, the lower the SC of the output light source is (see Note 2 in the [Supplementary Material](#)). In the experiment, we control the incident beam radius by adjusting the distance between lens 1 (L1, 100 mm) and the RD. After being collimated by lens 2 (L2, 100 mm), the beam is incident on the SLM1 (HDSLM80R) loaded with turbulent phase, which is continuously refreshed at a rate of 20 Hz. After passing through the turbulence, the beam is split into two parts by a beam splitter.



**Fig. 1** Principle and performance of TWC-Swin method. (a) LPR. SC modulation can adjust the SC by changing the distance  $D$ . Holographic modulation is used to load the phase hologram. The LPR generates two outputs, one for calculating SC and the other for network input. HWP, half-wave plate; PBS, polarized beam splitter; L, lens; RD, rotating diffuser; SLM, spatial light modulator; F, filter.  $D$ , distance between L1 and RD. (b) The detailed flow of the TWC-Swin method. The swin adapter can select the optimal model from the model space by obtaining SC. The color picture represents a case in progress. (c) Swin-model space and architecture of the swin model. The architecture of  $M_1 - M_{11}$  is the same; only the weights are different. The weights are obtained by network training at different distances. (d) The correspondence between SC and swin-model space. See Table S1 in the [Supplementary Material](#) for detailed data. (e) Inputs and outputs of the swin model with different SCs. (f) SSIM and PCC of swin-model outputs at different SCs. (g) Training and test data acquisition process. The training data did not contain any turbulence. (h) SSIM and PCC of swin-model outputs at different turbulent scenes.

The first part employs Michelson interference to capture interference fringes and measure the SC of the light. The second part is used for holographic imaging, with the phase hologram of the image loaded onto the SLM2 (PLUTO). The high-pass filter is employed to filter out the unmodulated zero-order diffraction pattern, and the final imaging result is captured by the complementary metal–oxide–semiconductor (CMOS, Sony, E3ISPM). In summary, we control the SC of the light source by adjusting the distance between lens L1 and the RD. We simulate a turbulent environment using the SLM1, with the intensity of the turbulence depending on the loaded turbulent phase. If turbulence is not required, the SLM1 can be turned off, and it functions as a mirror equivalent.

## 2.2 Oceanic Turbulence and Atmospheric Turbulence

The turbulence intensity in the experiment is determined by the spatial power spectrum of the turbulence. The function of the spatial power spectrum of the turbulent refractive-index fluctuations used in this paper is based on the assumption that turbulence is homogeneous and isotropic. We use the Nikishov power spectrum to describe oceanic turbulence:<sup>33</sup>

$$\begin{cases} \Phi_n(\kappa) = 0.388 \times 10^{-8} \varepsilon^{-1/3} \kappa^{-11/3} [1 + 2.35(\kappa\eta)^{2/3}] f(\kappa, \omega, \chi_t) \\ f(\kappa, \omega, \chi_t) = \chi_t [\exp(-A_T \delta) + \omega^{-2} \exp(-A_S \delta) \\ \quad - 2\omega^{-1} \exp(-A_{TS} \delta)] \\ \delta = 8.248(\kappa\eta)^{4/3} + 12.978(\kappa\eta)^2 \end{cases}, \quad (1)$$

where  $\kappa$  is the spatial wavenumber of turbulent fluctuations,  $\kappa = \sqrt{\kappa_x^2 + \kappa_y^2 + \kappa_z^2}$ .  $\varepsilon$  is the dissipation rate of turbulent kinetic energy per unit mass.  $\eta = 10^{-3}$  m is the Kolmogorov microscale (inner scale).  $\omega$  is the index of the relative strength of temperature and salinity fluctuations.  $A_T = 1.863 \times 10^{-2}$ ,  $A_S = 1.9 \times 10^{-4}$ , and  $A_{TS} = 9.41 \times 10^{-3}$ .  $\chi_t$  stands for a variate that represents the rate of dissipation of mean-square temperature, which varies from  $10^{-10}$  K<sup>2</sup>/s in deep water to  $10^{-4}$  K<sup>2</sup>/s in surface water. We only changed the oceanic turbulence intensity by adjusting  $\chi_t$ ; the greater the value of  $\chi_t$  is, the stronger the oceanic turbulence is. Detailed parameter settings for the power spectrum of oceanic turbulence can be found in Table S2 in the [Supplementary Material](#).

For atmospheric turbulence, we use the non-Kolmogorov power spectrum,<sup>34</sup>

$$\begin{cases} \Phi_n(\kappa) = A(\alpha) C_n^2 \frac{\exp(-\kappa^2/\kappa_m^2)}{(\kappa^2 + \kappa_0^2)^{\alpha/2}} \quad (0 < \kappa < \infty, 3 < \alpha < 4) \\ A(\alpha) = \frac{1}{4\pi} \Gamma(\alpha - 1) \cos\left(\frac{\pi\alpha}{2}\right) \\ c(\alpha) = \left[\frac{2\pi}{3} \Gamma\left(5 - \frac{\alpha}{2}\right) A(\alpha)\right]^{1/\alpha-5} \\ \kappa_m = \frac{c(\alpha)}{l_0} \\ \kappa_0 = \frac{2\pi}{L_0} \end{cases}, \quad (2)$$

where  $\alpha$  is the refractive index power spectral density power law.  $l_0$  and  $L_0$  represent inner and outer scales, respectively.  $C_n^2$  denotes the refractive index structure constant. We only changed the atmospheric turbulence intensity by adjusting  $C_n^2$ ; the greater the value of  $C_n^2$  is, the stronger the atmospheric turbulence is. Detailed parameter settings for the power spectrum

of atmospheric turbulence can be found in Table S2 in the [Supplementary Material](#). After setting reasonable parameters and returning to the space domain through the inverse Fourier transform, the turbulent phase can be obtained, which will be input into SLM1 to simulate the turbulent scene.

## 2.3 Data Acquisition

Low SC and turbulence are different physical scenarios, but the influence of these scenarios on holographic imaging can be described through SC. Based on the above method, we only use the data obtained under different SCs for model training, and any other data are used for testing [Fig. 1(g)]. The process of data acquisition is as follows.

(1) For low-SC data sets, we do not consider turbulence, turn off SLM1, and only adjust the SC of the light source by changing the distance between lens L1 and the RD. The initial distance is the focal length of lens L1 and the rotation speed of the diffuser is 200 r/min.

(2) The original image generates phase holograms through the iterative Gerchberg–Saxton algorithm<sup>35</sup> and loads them into the SLM2 sequentially. The imaging results are captured by CMOS1, and CMOS2 captures the interference fringes and calculates the degree of coherence.

(3) After all the data are captured, we increase the distance between L1 and RD. Each increment is 0.1 times the focal length of L1, and we keep the rotation speed of the diffuser constant and then repeat step 2.

(4) For the turbulence data set, the random turbulence phase generated by the simulation was loaded into SLM1 at a frequency of 20 Hz, and the distance between L1 and RD was kept as the focal length of L1. We input the phase hologram into SLM2. CMOS2 captures the interference fringes and CMOS1 captures 500 images. The final imaging result can be obtained by weighting and averaging these images.

(5) We change the intensity and type of turbulence loaded on SLM1 and then repeat step 4.

Our original images consist of public data sets, such as the Berkeley segmentation data set (BSD),<sup>36</sup> Celebfaces attributes high-quality data set (CelebA),<sup>37</sup> Flickr data set (Flickr),<sup>38</sup> Webvision data set (WED),<sup>39</sup> and DIV2k data set (DIV).<sup>40</sup> The training set is only composed of images captured by CMOS1 in steps 2 and 3.

In the training phase, we divide the training data into 11 groups based on SC and send them to the network for training in turn. Therefore we can obtain a model space containing swin models with different weights. In the testing phase, the swin adapter is a program that needs to receive the SC information of the light source and selects the optimal model in model space to achieve the image restoration task. Here we set to distance priority mode, and the swin adapter will select the weight parameter closest to the measured SC. The test set comes from the images generated in steps 4 and 5. Note that none of the test sets have been trained; they are blinded to the network. Our model was implemented using PyTorch; the detailed architecture can be found in Note 1 in the [Supplementary Material](#). We use adaptive moment estimation with weight decay (AdamW) as optimizer,<sup>41</sup> which is utilized to update the weights with initial learning rates of 0.0005 with a 50% drop every 10 epochs. The total epoch is 100. Mean-squared error (MSE) is the loss function of the network. All training and testing stages are

placed on the NVIDIA GTX3080Ti graphics card, and a full training period takes about 12 h. To effectively verify the performance of our method, a series of credible image quality assessment measures were applied. The full-reference measures include peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and Pearson correlation coefficient (PCC), which are used to provide an assessment of a single image in relation to perceived visual quality. See Note 4 in the [Supplementary Material](#) for descriptions of evaluation indices.

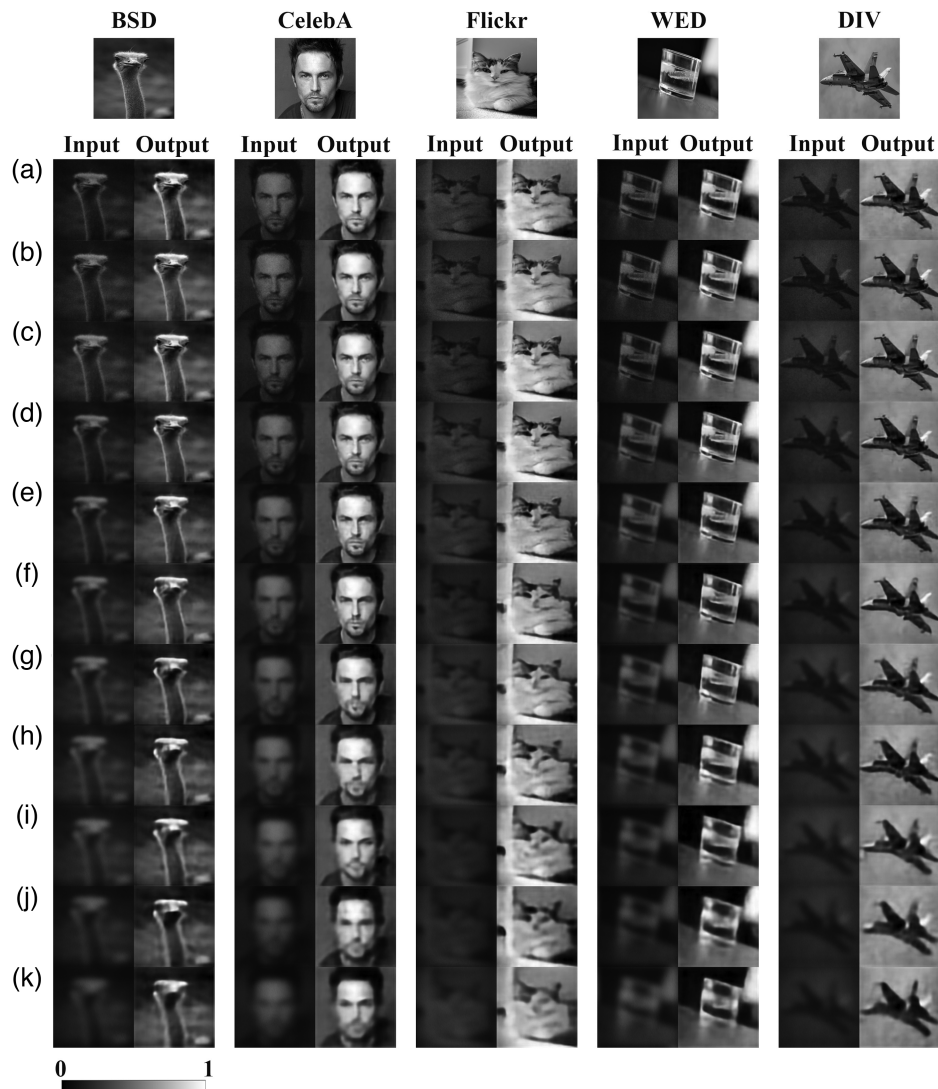
### 3 Results and Discussion

This section primarily showcases the performance of our method under various SCs and turbulent scenes. We simulated different strengths of oceanic and atmospheric turbulence,

enhancing the diversity of turbulence intensities and types. Additionally, we conducted comparative analyses with traditional convolutional-residual networks and performed ablation studies to reinforce the validity and efficiency of our proposed method. It is important to emphasize that our training data exclusively consisted of holographic imaging results obtained under different SC conditions, with none of the test data used during the training phase.

#### 3.1 Performance on Low SC

Figures 2 and 1(e) show the original images captured by CMOS1 and restored images processed by the TWC-Swin method under different SCs. We present 11 groups of test



**Fig. 2** Qualitative analysis of our method's performance at the different SCs. Input, raw image captured by CMOS1. Output, image processed by the network. (a)–(k) Different SCs: (a)  $D = f_1$ , SC is 0.494; (b)  $D = 1.1f_1$ , SC is 0.475; (c)  $D = 1.2f_1$ , SC is 0.442; (d)  $D = 1.3f_1$ , SC is 0.419; (e)  $D = 1.4f_1$ , SC is 0.393; (f)  $D = 1.5f_1$ , SC is 0.368; (g)  $D = 1.6f_1$ , SC is 0.337; (h)  $D = 1.7f_1$ , SC is 0.311; (i)  $D = 1.8f_1$ , SC is 0.285; (j)  $D = 1.9f_1$ , SC is 0.25; and (k)  $D = 2f_1$ , SC is 0.245.  $D$  means the distance between L1 and RD in the LPR and  $f_1$  is the focal length of L1. Our method can achieve improved image quality under low SC ([Video 1](#), MP4, 1.5 MB [URL: <https://doi.org/10.1117/1.AP.5.6.066003.s1>]).

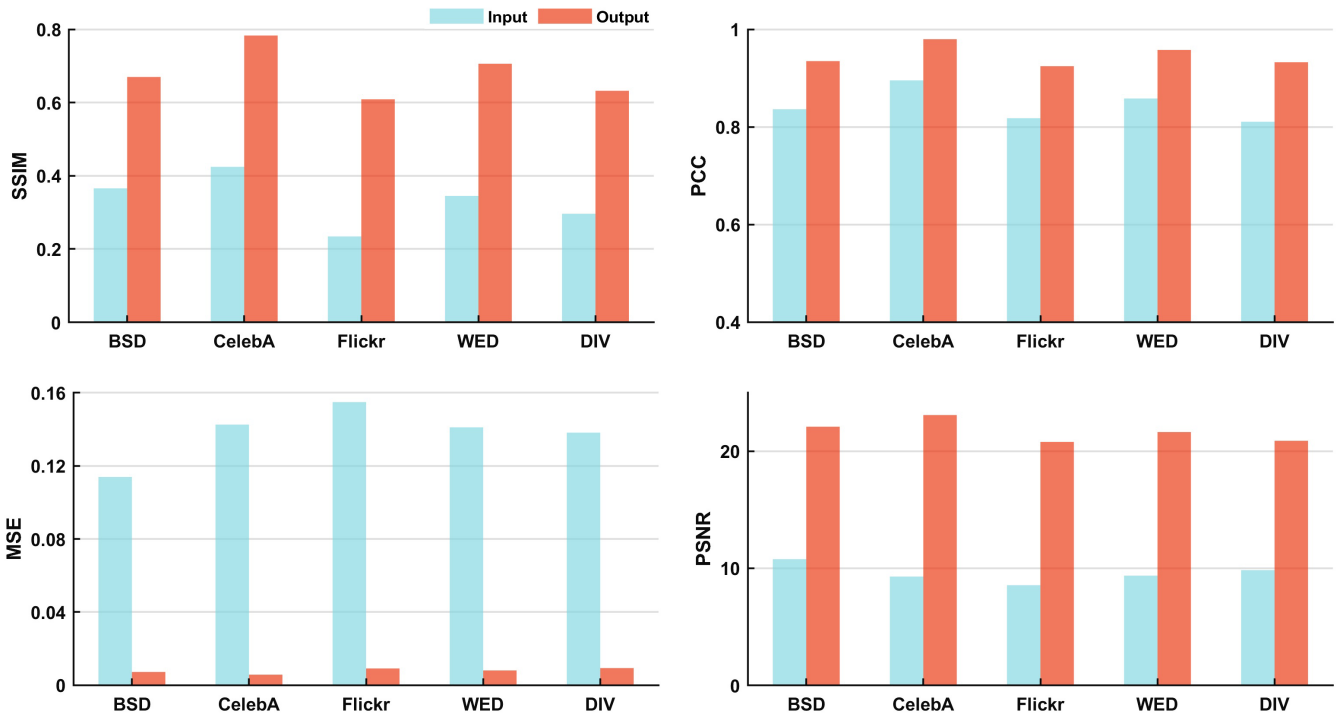
**Table 1** Quantitative analysis of evaluation indices (SSIM and PCC) at different SCs and test samples<sup>a</sup>.  $f_1$  is the focal length of L1. SC means spatial coherence of the light source.

SC	SSIM					PCC				
	BSD	CelebA	Flickr	WED	DIV	BSD	CelebA	Flickr	WED	DIV
Input_ $f_1$ , SC = 0.494	0.5893	0.5943	0.4296	0.6155	0.4625	0.9368	0.9575	0.9210	0.9146	0.8753
Output_ $f_1$	<b>0.8984</b>	<b>0.8908</b>	<b>0.8523</b>	<b>0.9019</b>	<b>0.8940</b>	<b>0.9807</b>	<b>0.9893</b>	<b>0.9848</b>	<b>0.9930</b>	<b>0.9819</b>
Input_ $1.3f_1$ , SC = 0.419	0.5775	0.5415	0.3917	0.6245	0.4184	0.8953	0.9303	0.8588	0.9149	0.8043
Output_ $1.3f_1$	<b>0.9189</b>	<b>0.8842</b>	<b>0.8676</b>	<b>0.8997</b>	<b>0.8918</b>	<b>0.9843</b>	<b>0.9928</b>	<b>0.9880</b>	<b>0.9928</b>	<b>0.9827</b>
Input_ $1.5f_1$ , SC = 0.368	0.6178	0.5394	0.2777	0.5677	0.3892	0.8957	0.9211	0.8396	0.8961	0.8144
Output_ $1.5f_1$	<b>0.8906</b>	<b>0.8513</b>	<b>0.8171</b>	<b>0.8541</b>	<b>0.8622</b>	<b>0.9691</b>	<b>0.9881</b>	<b>0.9783</b>	<b>0.9869</b>	<b>0.9680</b>
Input_ $1.7f_1$ , SC = 0.311	0.6040	0.5017	0.3183	0.5510	0.4136	0.8303	0.9035	0.8511	0.8568	0.7979
Output_ $1.7f_1$	<b>0.8624</b>	<b>0.7791</b>	<b>0.7483</b>	<b>0.8013</b>	<b>0.8038</b>	<b>0.9644</b>	<b>0.9787</b>	<b>0.9702</b>	<b>0.9759</b>	<b>0.9583</b>
Input_ $2f_1$ , SC = 0.245	0.4881	0.4469	0.3073	0.5271	0.3643	0.8072	0.8817	0.7557	0.8326	0.7196
Output_ $2f_1$	<b>0.8146</b>	<b>0.7540</b>	<b>0.6962</b>	<b>0.7722</b>	<b>0.7572</b>	<b>0.9431</b>	<b>0.9713</b>	<b>0.9505</b>	<b>0.9631</b>	<b>0.9341</b>
Ground truth	1	1	1	1	1	1	1	1	1	1

<sup>a</sup>Bold values indicate the better index.

results, each representing a different SC level and containing samples from five distinct data sets. As described in Sec. 2, the SC of the light source can be altered by adjusting the distance between RD and L1. It is evident that as the SC decreases, the quality of holographic imaging deteriorates significantly, exhibiting high levels of noise and blurriness. Simultaneously, the decrease in SC corresponds to a reduction in light efficiency, resulting in darker images that ultimately become indiscernible. After processing through the trained network, these degraded

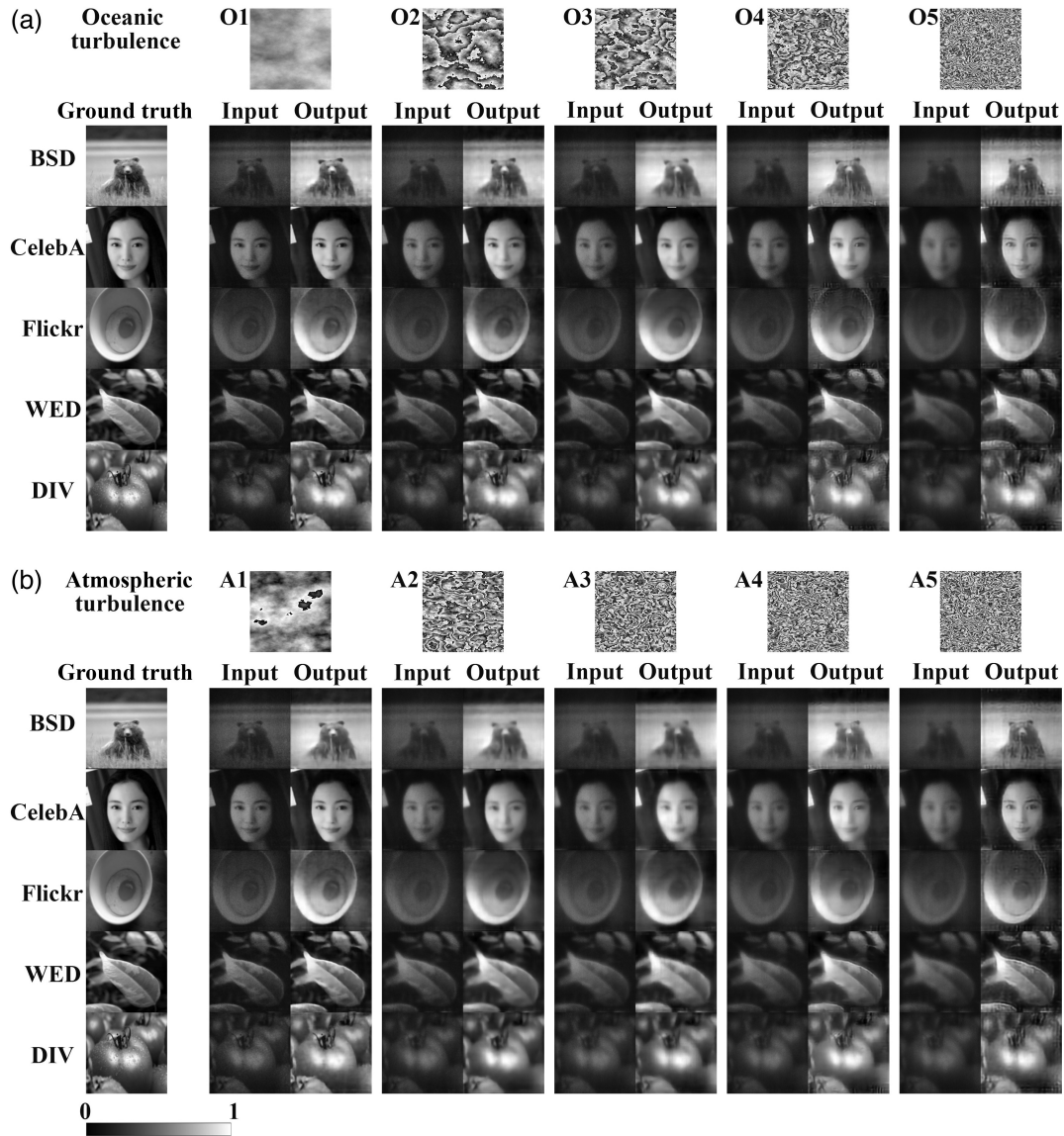
images become smoother, with improved sharpness, enhanced details, and reduced noise. Remarkably, even in low SC conditions where the original images captured by the CMOS1 sensor lack any discernible details, our network successfully reconstructs a significant portion of the elements. To accurately evaluate the effectiveness of image restoration, we present the evaluation indices (SSIM and PCC), comparing the original and reconstructed images with respect to the ground truth for different SCs [Fig. 1(f) and Table 1]. Other indices are provided in



**Fig. 3** Average results of the evaluation indices for each test data set. The coherence is 0.368. Results of other coherences are provided in Fig. S2 in the [Supplementary Material](#). All evaluation indices demonstrate that our method possesses strong image restoration ability under low SC.

Table S3 in the [Supplementary Material](#). The quantitative results further validate the significant improvement achieved in various indicators of the reconstructed images compared to the original ones, approaching the ground truth. Figure 3 illustrates the average evaluation indices for each test set. Here only partial results are shown; more detailed results are included in Fig. S2 in the [Supplementary Material](#). It can be seen that each evaluation index of images has risen significantly compared to the

original images after being processed by the TWC-Swin method, indicating a substantial improvement in the image quality. Moreover, the network demonstrates its robust generalization capability by performing image restoration on multiple test sets, which are beyond the scope of the training set. This implies that our method has effectively learned the underlying patterns in the data during training and can apply these patterns to unseen data, resulting in successful image restoration.



**Fig. 4** Qualitative analysis of our method's performance across varying intensities of (a) oceanic and (b) atmospheric turbulence. The network trained with coherence as physical prior information can effectively overcome the impact of turbulence on imaging and improve image quality. (O1)–(O5) mean oceanic turbulence phase and (A1)–(A5) mean atmospheric turbulence phase. (O1)  $\chi_t = 10^{-9}$  K<sup>2</sup>/s, coherence is 0.491. (O2)  $\chi_t = 10^{-7}$  K<sup>2</sup>/s, coherence is 0.482. (O3)  $\chi_t = 2 \times 10^{-7}$  K<sup>2</sup>/s, coherence is 0.447. (O4)  $\chi_t = 4 \times 10^{-7}$  K<sup>2</sup>/s, coherence is 0.404. (O5)  $\chi_t = 10^{-6}$  K<sup>2</sup>/s, coherence is 0.373. (A1)  $C_n^2 = 10^{-14}$  m<sup>3- $\alpha$</sup> , coherence is 0.507. (A2)  $C_n^2 = 1.5 \times 10^{-13}$  m<sup>3- $\alpha$</sup> , coherence is 0.459. (A3)  $C_n^2 = 2.5 \times 10^{-13}$  m<sup>3- $\alpha$</sup> , coherence is 0.43. (A4)  $C_n^2 = 3.5 \times 10^{-13}$  m<sup>3- $\alpha$</sup> , coherence is 0.403. (A5)  $C_n^2 = 5 \times 10^{-13}$  m<sup>3- $\alpha$</sup> , coherence is 0.378. Other parameter settings of the turbulent power spectrum function can be found in Table S2 in the [Supplementary Material](#) (Video 2, MP4, 36.4 MB [URL: <https://doi.org/10.1117/1.AP.5.6.066003.s2>]).

### 3.2 Performance on Oceanic Turbulence and Atmospheric Turbulence

Owing to the stochastic variations of the refractive index within oceanic and atmospheric turbulence, the phase information of light beams becomes distorted, thereby reducing SC and degrading the quality of computational holography images. This issue can be effectively addressed using the TWC-Swin method. It should be mentioned that all images captured under turbulent scenes were never trained by the network. Figure 4 demonstrates the remarkable image restoration capability of TWC-Swin method under varying intensities of oceanic and atmospheric turbulence. As discussed in Sec. 2, the turbulence intensity depends on certain variates of the power spectrum function, where stronger turbulence presents more complex simulated turbulence phases, as shown in Figs. 4(A5) and 4(O5). We carried

out experiments under five distinct intensities of both oceanic and atmospheric turbulence, and simultaneously measured the SC of the light source for selecting the optimal model. It should be noted that the turbulence phase loaded on the SLM is continuously refreshed (20 Hz). To provide stronger evidence, we present the evaluation indices (SSIM and PCC) for oceanic and atmospheric turbulence in Tables 2 and 3 and Fig. 1(h), whereas additional indices (MSE and PSNR) can be found in Tables S4 and S5 in the [Supplementary Material](#). Our analysis concluded that as the turbulence intensity increases, the SC experiences a downturn, which subsequently degrades image quality. Nevertheless, our proposed method is capable of overcoming these adverse effects and effectively improving the image quality regardless of the turbulence intensity. Our model learns the universal features of image degradation and restoration that depend on SC. This further demonstrates the strong generalization

**Table 2** Quantitative analysis of evaluation indices (SSIM and PCC) at different oceanic turbulence intensities<sup>a</sup>.

Oceanic turbulence	SSIM					PCC				
	BSD	CelebA	Flickr	WED	DIV	BSD	CelebA	Flickr	WED	DIV
Input (O1)	0.5331	0.6773	0.6810	0.6016	0.7018	0.8978	0.9404	0.8876	0.9096	0.8718
Output (O1)	<b>0.8088</b>	<b>0.7916</b>	<b>0.8368</b>	<b>0.8077</b>	<b>0.8172</b>	<b>0.9303</b>	<b>0.9707</b>	<b>0.9334</b>	<b>0.9560</b>	<b>0.9044</b>
Input (O2)	0.5098	0.6566	0.6690	0.5716	0.5371	0.8855	0.9329	0.8786	0.8970	0.8494
Output (O2)	<b>0.7823</b>	<b>0.7609</b>	<b>0.8015</b>	<b>0.7819</b>	<b>0.8005</b>	<b>0.9211</b>	<b>0.9611</b>	<b>0.9209</b>	<b>0.9448</b>	<b>0.8901</b>
Input (O3)	0.4950	0.6538	0.6575	0.5455	0.5281	0.8764	0.9313	0.8585	0.8916	0.8371
Output (O3)	<b>0.7191</b>	<b>0.7169</b>	<b>0.8434</b>	<b>0.7378</b>	<b>0.7984</b>	<b>0.8896</b>	<b>0.9413</b>	<b>0.8871</b>	<b>0.9344</b>	<b>0.8793</b>
Input (O4)	0.4796	0.6408	0.6474	0.5034	0.5074	0.8774	0.9245	0.8576	0.8664	0.8130
Output (O4)	<b>0.7060</b>	<b>0.6932</b>	<b>0.7287</b>	<b>0.6718</b>	<b>0.7217</b>	<b>0.8847</b>	<b>0.9379</b>	<b>0.8835</b>	<b>0.8892</b>	<b>0.8213</b>
Input (O5)	0.4519	0.6041	0.6202	0.4446	0.4945	0.8456	0.9075	0.8287	0.8281	0.7631
Output (O5)	<b>0.6899</b>	<b>0.6721</b>	<b>0.7225</b>	<b>0.6286</b>	<b>0.6958</b>	<b>0.8909</b>	<b>0.9415</b>	<b>0.8888</b>	<b>0.8839</b>	<b>0.8152</b>
Ground truth	1	1	1	1	1	1	1	1	1	1

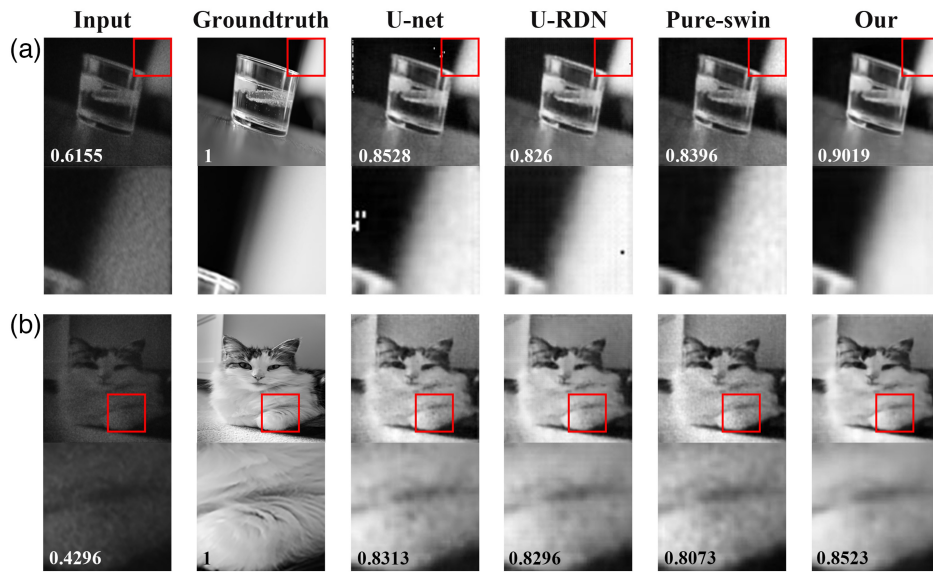
<sup>a</sup>Bold values indicate the better index.

**Table 3** Quantitative analysis of evaluation indices (SSIM and PCC) at different atmospheric turbulence intensities<sup>a</sup>.

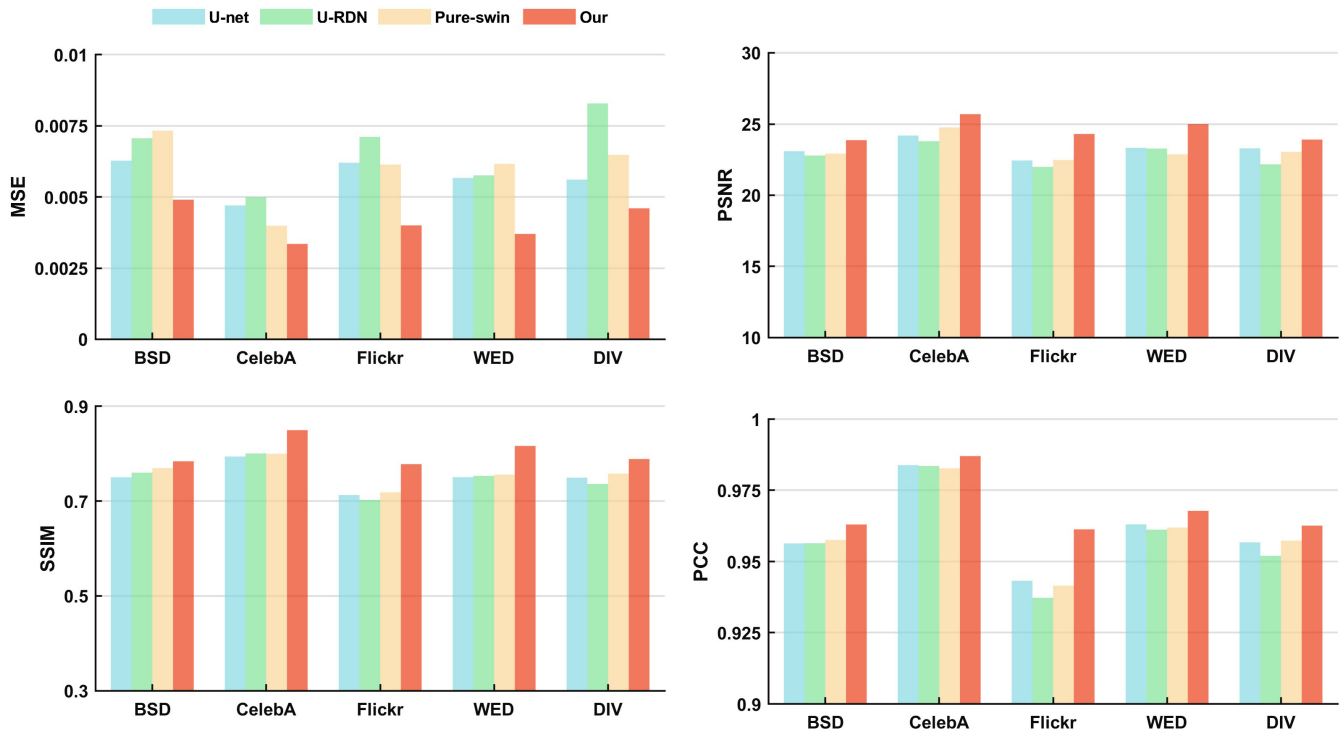
Atmospheric turbulence	SSIM					PCC				
	BSD	CelebA	Flickr	WED	DIV	BSD	CelebA	Flickr	WED	DIV
Input (A1)	0.5738	0.6821	0.6988	0.6495	0.6338	0.9014	0.9404	0.8929	0.9160	0.9766
Output (A1)	<b>0.7798</b>	<b>0.7741</b>	<b>0.8337</b>	<b>0.8161</b>	<b>0.8231</b>	<b>0.9361</b>	<b>0.9564</b>	<b>0.9215</b>	<b>0.9574</b>	<b>0.9116</b>
Input (A2)	0.5311	0.6513	0.6727	0.5743	0.5701	0.8797	0.9264	0.8676	0.8896	0.8279
Output (A2)	<b>0.7312</b>	<b>0.6938</b>	<b>0.7699</b>	<b>0.6960</b>	<b>0.7581</b>	<b>0.8920</b>	<b>0.9353</b>	<b>0.8924</b>	<b>0.9141</b>	<b>0.8643</b>
Input (A3)	0.5083	0.6383	0.6785	0.5348	0.5720	0.8688	0.9202	0.8493	0.8747	0.8081
Output (A3)	<b>0.6615</b>	<b>0.6797</b>	<b>0.7427</b>	<b>0.6362</b>	<b>0.7369</b>	<b>0.8843</b>	<b>0.9392</b>	<b>0.8708</b>	<b>0.8919</b>	<b>0.8418</b>
Input (A4)	0.4965	0.6264	0.6635	0.5202	0.5575	0.8590	0.9161	0.8364	0.8673	0.8040
Output (A4)	<b>0.6915</b>	<b>0.6751</b>	<b>0.7287</b>	<b>0.6336</b>	<b>0.7273</b>	<b>0.8789</b>	<b>0.9308</b>	<b>0.8705</b>	<b>0.8855</b>	<b>0.8331</b>
Input (A5)	0.4959	0.6153	0.6595	0.4840	0.5407	0.8524	0.9080	0.8263	0.8493	0.7862
Output (A5)	<b>0.6761</b>	<b>0.6893</b>	<b>0.7201</b>	<b>0.6127</b>	<b>0.6802</b>	<b>0.8719</b>	<b>0.9465</b>	<b>0.8875</b>	<b>0.8749</b>	<b>0.8255</b>
Ground truth	1	1	1	1	1	1	1	1	1	1

<sup>a</sup>Bold values indicate the better index.





**Fig. 5** Visualization of performance of different methods. The SSIM is shown in the bottom left corner. Our method presents the best performance, which is shown by smoother images with lower noise. (a) Sample selected with the WED data set and magnified insets of the red bounding region. (b) Sample selected with Flickr data set and magnified insets of the red bounding region. The pure swin model can be obtained by removing the postprocessing block of the swin model (Video 3, MP4, 0.6 MB [URL: <https://doi.org/10.1117/1.AP.5.6.066003.s3>]).

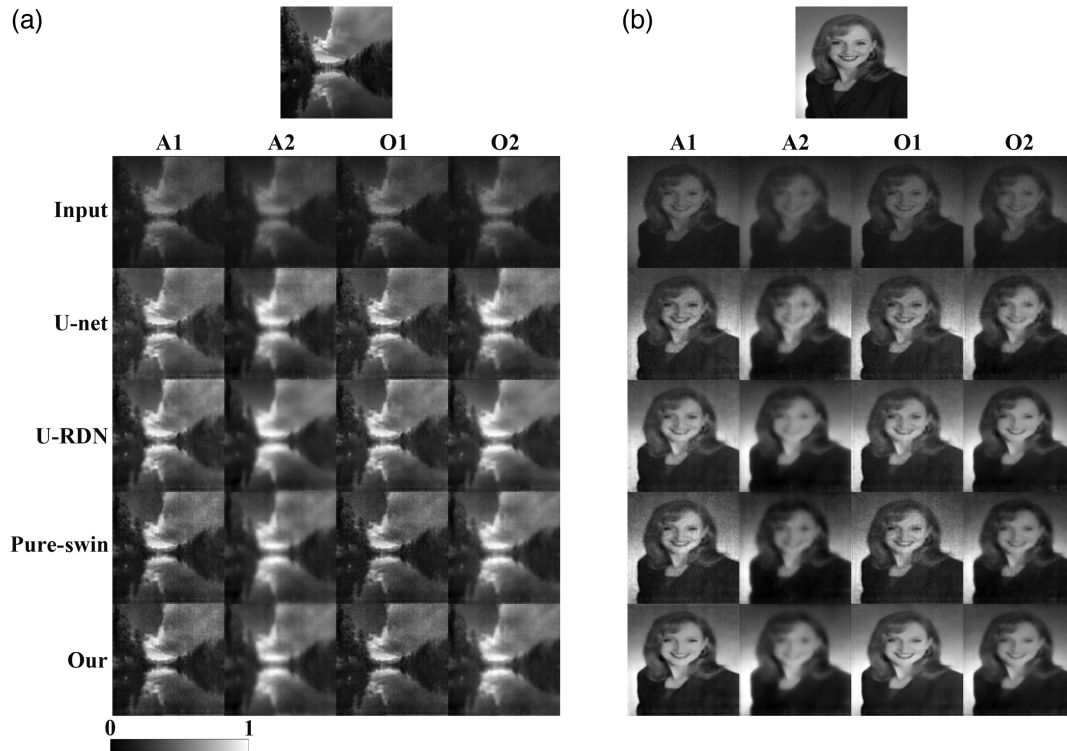


**Fig. 6** Performance between different methods on various data sets with SC being 0.494. Our model outperforms other methods across various data sets and indices.

capability of the network trained with SC as physical prior information and the ability to apply learned knowledge from the training set to new, unseen scenes. This versatility is a desirable trait in a neural network, as it suggests the method’s potential for broad application.

### 3.3 Comparison between Different Methods and Ablation Study

In this section, we conduct a comprehensive comparative study of different methodologies, assessing their performance and



**Fig. 7** (a), (b) Performance comparison between different methods at various turbulent scenes. (A1)  $C_n^2 = 10^{-14} \text{ m}^{3-\alpha}$ , coherence is 0.506. (A2)  $C_n^2 = 1.5 \times 10^{-13} \text{ m}^{3-\alpha}$ , coherence is 0.459. (O1)  $\chi_t = 10^{-9} \text{ K}^2/\text{s}$ , coherence is 0.491. (O2)  $\chi_t = 10^{-7} \text{ K}^2/\text{s}$ , coherence is 0.482. Note that all methods are trained with coherence as physical prior information and improve image quality under turbulence conditions. This demonstrates that incorporating appropriate physical prior information can help the network cope with multiscale tasks.

efficacy in restoring images under challenging conditions of low SC and turbulent scenes. Traditional convolution-fusion framework methods, U-net,<sup>42</sup> and U-RDN<sup>13</sup> were compared to demonstrate the power of the proposed swin model.

In our network architecture, the swin transformer serves as a robust backbone module, responsible for extracting high-level features from input. The special working mechanism gives it powerful hierarchical representation and global perception capabilities. However, direct output from the swin transformer often exhibits artifacts and high-noise levels in image restoration tasks. Therefore, it is necessary to add lightweight convolutional layers as postprocessing blocks. Convolution layers capture local features of the image through local receptive fields, aiding in a better understanding of image details and textures while facilitating mapping from high-dimensional to low-dimensional spaces, resulting in high-quality output. To validate the effectiveness of the postprocessing block in the swin model, we conduct an ablation study. In the ablation study, we created a control group named pure swin, which was obtained by removing the postprocessing block from the swin model. The training processes and data sets of all methods are consistent. Figure 5 shows detailed comparisons of images processed by various methods. Figure 6 illustrates the quantitative results between different methods on various data sets. More qualitative results are provided in Figs. S3 and S4 in the [Supplementary Material](#). Comparing the visual output results of pure swin and the swin model, we found that the output results of the pure swin framework will produce black spots, resulting in blurred perception;

the SSIM is 0.8396, a 7% reduction. This is because the swin transformer lacks the ability to sense local features and dimensional mapping. Convolutional layers can fill this gap by offering a mechanism to refine and enhance local features past the swin transformer blocks. The ablation study (compared with pure swin) validates that the postprocessing module is indispensable for the swin model.

We tested the performance of other networks under the same conditions. Our proposed network outperforms other methods by presenting the lowest noise and best evaluation index. Tables S6 and S7 in the [Supplementary Material](#) provide a detailed quantitative comparison of the performance across different models and different SCs. In the task of image restoration under low SC, our proposed methodology exhibits superior performance across all evaluative indices when juxtaposed with alternative approaches. Figure 7 shows the comparative performance of various methods when faced with image degradation due to various turbulence types and intensities. We observed that all networks trained with SC exhibit the ability to significantly improve the image quality under turbulent scenes and not just the swin model. This is an exciting result, as it signifies the successful integration of physical prior information into network training, enabling the networks to be applied to multiple tasks and scenarios.

## 4 Conclusions

By leveraging the SC as physical prior information and harnessing advanced deep-learning algorithms, we proposed a methodology,

TWC-Swin, which demonstrates exceptional capabilities in simultaneously restoring images in low SC and random turbulent scenes. Our multicoherence and multiturbulence holographic imaging data sets, consisting of natural images, are created by the LPR, which can simulate different SCs and turbulence scenes (see Sec. 2). Though the swin model used in the tests was trained solely on the multicoherence data set, it can achieve promising results on both low SC, oceanic turbulence and atmospheric turbulence scenes. The key is that we capture the common physical property in these scenes, SC, and use it as physical prior information to generate a training set, so that the TWC-Swin method exhibits remarkable generalization capabilities, effectively restoring images from unseen scenes beyond the training set. Furthermore, through a series of rigorous experiments and comparisons, we have established the superiority of the swin model over traditional convolutional frameworks and alternative methods in terms of image restoration from qualitative and quantitative analysis (see Sec. 3). The integration of SC as a fundamental guiding principle in network training has proven to be a powerful strategy in aligning downstream tasks with pretrained models.

Our research findings offer guidance not only for the domain of optical imaging but also for the integration with the segment anything model (SAM),<sup>43</sup> extending its applicability to multiphysics scenarios. For instance, in turbulent scenes, our methodology can be implemented for preliminary image processing, enabling the utilization of unresolved images for precise image recognition and segmentation tasks facilitated by SAM. Moreover, our experimental scheme also provides a simple idea for turbulence detection. Our research contributes valuable insights into the use of deep-learning algorithms for addressing image degradation problems in multiple scenes and highlights the importance of incorporating physical principles into network training. It is foreseeable that our research can serve as a successful case for the combination of deep learning and holographic imaging in the future, which facilitates the synergistic advancement of the fields of optics and computer science.

## Code and Data Availability

The codes of the TWC-Swin method, trained model, as well as some example images for testing, are publicly available at <https://github.com/tongxinoptica/TWC-Swin>. All relevant data that support the findings of this work are available from the corresponding author upon reasonable request. The parameter settings of TWC-Swin used in synthesizing the training and evaluation data sets will be publicly available along with this paper.

## Acknowledgments

This work was mainly supported by the National Natural Science Foundation of China (Grants Nos. 12174338 and 11874321) received by D.M.Z. All authors contributed to the discussions and preparation of the manuscript.<sup>44</sup> The authors declare no competing interests.

## References

1. L. B. Lesem, P. M. Hirsch, and J. A. Jordan, "Scientific applications: computer synthesis of holograms for 3D display," *Commun. ACM* **11**, 661–674 (1968).
2. M. Lurie, "Fourier-transform holograms with partially coherent light: holographic measurement of spatial coherence," *J. Opt. Soc. Am.* **58**, 614–619 (1968).
3. U. Schnars and W. Jüptner, "Direct recording of holograms by a CCD target and numerical reconstruction," *Appl. Opt.* **33**, 179–181 (1994).
4. R. Horisaki et al., "Compressive propagation with coherence," *Opt. Lett.* **47**, 613–616 (2022).
5. D. Blinder et al., "Signal processing challenges for digital holographic video display systems," *Signal Process. Image Commun.* **70**, 114–130 (2019).
6. H. Ko and H. Y. Kim, "Deep learning-based compression for phase-only hologram," *IEEE Access* **9**, 79735–79751 (2021).
7. L. Shi et al., "Towards real-time photorealistic 3D holography with deep neural networks," *Nature* **591**, 234–239 (2021).
8. C. Lee et al., "Deep learning based on parameterized physical forward model for adaptive holographic imaging with unpaired data," *Nat. Mach. Intell.* **5**, 35–45 (2023).
9. X. Guo et al., "Stokes meta-hologram toward optical cryptography," *Nat. Commun.* **13**, 6687 (2022).
10. H. Yang et al., "Angular momentum holography via a minimalist metasurface for optical nested encryption," *Light Sci. Appl.* **12**, 79 (2023).
11. R. Fiolka, K. Si, and M. Cui, "Complex wavefront corrections for deep tissue focusing using low coherence backscattered light," *Opt. Express* **20**, 16532–16543 (2012).
12. S. Lim et al., "Optimal spatial coherence of a light-emitting diode in a digital holographic display," *Appl. Sci.* **12**, 4176 (2022).
13. Y. Deng and D. Chu, "Coherence properties of different light sources and their effect on the image sharpness and speckle of holographic displays," *Sci. Rep.* **7**, 5893 (2017).
14. X. Tong et al., "A deep-learning approach for low-spatial-coherence imaging in computer-generated holography," *Adv. Photonics Res.* **4**, 2200264 (2023).
15. Y. Peng et al., "Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration," *Sci. Adv.* **7**, 5040 (2021).
16. F. Wang et al., "Propagation of coherence-OAM matrix of an optical beam in vacuum and turbulence," *Opt. Express* **31**, 20796–20811 (2023).
17. D. Jin et al., "Neutralizing the impact of atmospheric turbulence on complex scene imaging via deep learning," *Nat. Mach. Intell.* **3**, 876–884 (2021).
18. Q. Zhang et al., "Effect of oceanic turbulence on the visibility of underwater ghost imaging," *J. Opt. Soc. Am. A* **36**, 397–402 (2019).
19. K. Wang et al., "Deep learning wavefront sensing and aberration correction in atmospheric turbulence," *Photonix* **2**, 8 (2021).
20. Y. Chen et al., "A wavelet based deep learning method for underwater image super resolution reconstruction," *IEEE Access* **8**, 117759–117769 (2020).
21. L. Zhang et al., "Restoration of single pixel imaging in atmospheric turbulence by Fourier filter and CGAN," *Appl. Phys. B* **127**, 45 (2021).
22. Y. Baykal, Y. Ata, and M. C. Gökçe, "Underwater turbulence, its effects on optical wireless communication and imaging: a review," *Opt. Laser Technol.* **156**, 108624 (2022).
23. J. Bertolotti and O. Katz, "Imaging in complex media," *Nat. Phys.* **18**, 1008–1017 (2022).
24. T. Zeng, Y. Zhu, and E. Y. Lam, "Deep learning for digital holography: a review," *Opt. Express* **29**, 40572–40593 (2021).
25. A. Khan et al., "GAN-Holo: generative adversarial networks-based generated holography using deep learning," *Complexity* **2021**, 6662161 (2021).
26. M. Liao et al., "Scattering imaging as a noise removal in digital holography by using deep learning," *New J. Phys.* **24**, 083014 (2022).
27. T. Shimobaba et al., "Deep-learning computational holography: a review," *Front. Photonics* **3**, 854391 (2022).

28. Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning in holography and coherent imaging," *Light Sci. Appl.* **8**, 85 (2019).
29. Z. Chen et al., "Physics-driven deep learning enables temporal compressive coherent diffraction imaging," *Optica* **9**, 677 (2022).
30. Y. Jo et al., "Holographic deep learning for rapid optical screening of anthrax spores," *Sci. Adv.* **3**, e1700606 (2023).
31. K. He et al., "Deep residual learning for image recognition," in *IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, pp. 770–778 (2016).
32. Z. Liu et al., "Swin transformer: hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 10012–10022 (2021).
33. V. V. Nikishov, "Spectrum of turbulent fluctuations of the sea-water refraction index," *Int. J. Fluid Mech. Res.* **27**, 82–98 (2000).
34. B. E. Stribling, B. M. Welsh, and M. C. Roggemann, "Optical propagation in non-Kolmogorov atmospheric turbulence," *Proc. SPIE* **2471**, 181–195 (1995).
35. R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik* **35**, 237–246 (1972).
36. D. Martin et al., "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Eighth IEEE Int. Conf. Comput. Vis.*, Vol. 2, pp. 416–423 (2001).
37. Z. Liu et al., "Deep learning face attributes in the wild," in *IEEE Int. Conf. Comput. Vis.*, pp. 3730–3738 (2015).
38. P. Young et al., "From image descriptions to visual denotations: new similarity metrics for semantic inference over event descriptions," *Trans. Assoc. Comput. Linguist.* **2**, 67–78 (2014).
39. W. Li et al., "WebVision database: visual learning and understanding from web data," arXiv:1708.02862 (2017).
40. E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: dataset and study," in *IEEE Conf. Comput. Vis. and Pattern Recognit. Workshops (CVPRW)*, pp. 1122–1131 (2017).
41. I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," arXiv:1711.05101 (2017).
42. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
43. A. Kirillov et al., "Segment anything," arXiv:2304.02643 (2023).
44. F. Gori and M. Santarsiero, "Devising genuine spatial correlation functions," *Opt. Lett.* **32**, 3531–3533 (2007).

**Xin Tong** is a PhD student at the School of Physics, Zhejiang University, Hangzhou, China. He received his BS degree in physics from Zhejiang University of Science and Technology, Hangzhou, China. His current research interests include holographic imaging, deep learning, computational imaging, and partial coherence theory.

**Renjun Xu** received his PhD from the University of California, Davis, California, United States. He is a ZJU100 Young Professor and a PhD supervisor at the Center for Data Science, Zhejiang University, Hangzhou, China. He was the senior director of data and artificial intelligence at VISA Inc. His research interests include machine learning, alignment techniques for large-scale pretrained models, transfer learning, space editing, transformation, generation, and the interdisciplinarity of physics and mathematics.

**Pengfei Xu** is a PhD student at the School of Physics, Zhejiang University, Hangzhou, China. He received his BS degree in physics from Zhejiang University, Hangzhou, China, in 2017. His current research interests include computational holographic imaging, partially coherent structured light field, and vortex beam manipulation techniques.

**Zishuai Zeng** is a PhD student at the School of Physics, Zhejiang University, Hangzhou, China. He received his BS degree in 2019 from the School of Information Optoelectronic Science and Engineering at South China Normal University. His current research interests include computer-generated holography, as well as beam propagation transformation and computational imaging.

**Shuxi Liu** is a PhD student at the School of Physics, Zhejiang University, China. He received his BS degree in physics from Zhejiang University in 2022. His current research interests include catastrophe optics, optical vortex, and computational imaging.

**Daomu Zhao** received his PhD from Zhejiang University, Hangzhou, China. Since 2003, he has been as a professor of the School of Physics at Zhejiang University. Now, he is the director of the Institute of Optoelectronic Physics at Zhejiang University. He has broad research interests in beam transmission, coherence and polarization theory, diffraction optics, holographic imaging, and deep learning.