Supplementary Information for

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

Xin Tong[1,2], Renjun Xu[2], Pengfei Xu[1], Zishuai Zeng[1], Shuxi Liu[1], and Daomu Zhao[1,*]

1 Zhejiang Province Key Laboratory of Quantum Technology and Device, School of Physics, Zhejiang University, Hangzhou 310027, P. R. China
2 Center for Data Science, Zhejiang University, Hangzhou 310027, P. R. China
*E-mail: dmz123@zju.edu.cn

Content

Supplementary Note 1: Architecture of swin-model

Figure S1 distinctly outlines the flow of the swin-model, allowing for a comprehensive understanding of its workings. The architecture is composed of three parts: a preprocessing block, a deep feature extraction module, and a postprocessing block. The deep feature extraction module utilizes six RSTB (Residual Swin Transformer Block) modules based on a global residual architecture, which combines low-frequency and high-frequency information before transmitting to the post-processing module, aiding in the stabilizing training. Each RSTB is composed of three Two-Successive Swin Transformer blocks and incorporates convolution operations and skip connections to form a local residual architecture. This fusion design between Swin Transformer and convolution residual enhances the utilization of prior information while incorporating the inductive bias of the convolution operation into the Transformer-based network, establishing a solid foundation for feature transmission and aggregation across layers. As Swin Transformer decomposes the image into non-overlapping patches with a sequential structure, self-attention mechanisms are employed to capture positional correlations among patches for deep feature extraction. In image restoration tasks, direct output from Swin Transformer often exhibits artifacts and high noise levels, necessitating the inclusion of a postprocessing module comprised solely of convolutional layers. 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. The results section includes ablation experiments, confirming the importance of the postprocessing module.

Supplementary Note 2: Theory of spatial coherence

In our experimental setup, the quasi-monochromatic high-coherence laser enters the diffuser, it is randomly scattered and loses its original amplitude and phase distribution. When the diffuser is rotated, the amplitude and phase of the beam will be changed randomly, which makes the beam become completely incoherent light at the output surface of the diffuser [14]. According to Zernike's theorem, the partially coherent beam will be generated during the subsequent transmission and we use lens 2 (L2) to collimate it. From the rotating diffuser (RD) to the L2 the kernel function $H(r,p)$ is regarded as the response function between the incident surface and the exit surface [45],

$$H(r,\rho)=-\frac{i}{\lambda f}\exp\left[\frac{i\pi}{\lambda f}\left(r^2-2\rho r\right)\right]. \tag{S1}$$

where r and p means the vector coordinates of the input and output surfaces, respectively. For quasi-monochromatic light, the second-order statistical properties of the optical field can be described by the cross-spectral density function,

$$W(\rho_1,\rho_2)=\int I(r)H_0^*(\rho_1,r)H_0(\rho_2,r)d^2r, \tag{S2}$$

where $I(r)$ represents the intensity distribution at the input plane. By employing the above equations, we can derive the cross-spectral density function of the output plane,

$$W(\rho_1,\rho_2)=\int I(r)\exp\left[-\frac{i2\pi r}{\lambda f}(\rho_1-\rho_2)\right]d^2r. \tag{S3}$$

The spectral coherence between two points can be described by the cross-spectral density function,

$$\mu(\rho_1,\rho_2)=\frac{W(\rho_1,\rho_2)}{\sqrt{W(\rho_1,\rho_1)}\sqrt{W(\rho_2,\rho_2)}}. \tag{S4}$$

Therefore, the spatial coherence of the light source depends on the radius of the incident beam on the RD, the larger the radius, the lower the SC of the output light source. We change the radius of the spot by adjusting the distance between lens 1 (L1) and the RD.

We use Michelson interference to capture interference fringes and measure the spatial coherence of the light source by fringe contrast. In order to control the error, we use the average maximum and minimum intensity of the fringes within a certain period to calculate the coherence. The results under different distances between L1 and RD and different turbulences are shown in Table S1.

Supplementary Note 3: Parameters of the turbulence power spectrum

Oceanic turbulence power spectrum[3],

$$\begin{cases}\Phi_n(\kappa)=0.388\times10^{-8}\varepsilon^{-1/3}\kappa^{-11/3}\left[1+2.35(\kappa\eta)^{2/3}\right]\times f(\kappa,\omega,\chi_t)\\ f(\kappa,\omega,\chi_t)=\chi_t\left[\exp(-A_T\delta)+\omega^{-2}\exp(-A_S\delta)-2\omega^{-1}\exp(-A_{TS}\delta)\right],\\ \delta=8.248(\kappa\eta)^{4/3}+12.978(\kappa\eta)^2\end{cases} \tag{S5}$$

Atmosphere turbulence power spectrum[4],

$$\begin{cases}\Phi_n(\kappa)=A(\alpha)C_n^{\ 2}\dfrac{\exp\left(-\kappa^2/\kappa_m^2\right)}{\left(\kappa^2+\kappa_0^2\right)^{\alpha/2}},(0<\kappa<\infty,3<\alpha<4)\\ A(\alpha)=\dfrac{1}{4\pi}\Gamma(\alpha-1)\cos\left(\dfrac{\pi\alpha}{2}\right)\\ c(\alpha)=\left[\dfrac{2\pi}{3}\Gamma\left(5-\dfrac{\alpha}{2}\right)A(\alpha)\right]^{1/\alpha-5}\\ \kappa_m=\dfrac{c(\alpha)}{l_0}\\ \kappa_0=\dfrac{2\pi}{L_0}\end{cases}. \tag{S6}$$

Table S2 lists the meanings and values of all symbols. For oceanic turbulence, we adjust χ_t to change the turbulent intensity. For atmospheric turbulence, we adjust C_n^2 to change the turbulent intensity.

Supplementary Note 4: Evaluation index

We use mean squared error (MSE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and Pearson correlation coefficient (PCC) to effectively verify the performances of TWC-Swin method. Their expressions are as follows,

$$\begin{cases} \mathrm{MSE}(O,G)=\dfrac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(O(i,j)-G(i.j)\right)^2 \\ \mathrm{PSNR}(O,G)=10\log_{10}\left(\dfrac{\left(2^p-1\right)^2}{\mathrm{MSE}(O,G)}\right) \\ \mathrm{SSIM}(O,G)=\dfrac{\left(2\mu_O\mu_G+C_1\right)\left(2\sigma_{OG}+C_2\right)}{\left(\mu_O{}^2+\mu_G{}^2+C_1\right)\left(\sigma_O{}^2+\sigma_G{}^2+C_2\right)} \\ \mathrm{PCC}(O,G)=\dfrac{\sum_{\mathrm{i}=1}^{m}\sum_{j=1}^{n}\left(O(i,j)-\tilde{O}\right)\left(G(i,j)-\tilde{G}\right)}{\sqrt{\sum_{\mathrm{i}=1}^{m}\sum_{j=1}^{n}\left(O(i,j)-\tilde{O}\right)^2}\sqrt{\sum_{\mathrm{i}=1}^{m}\sum_{j=1}^{n}\left(G(i,j)-\tilde{G}\right)^2}} \end{cases} . \tag{S7}$$

Where m and n are the width and height of the image. O means the output of the network, G denotes the ground truth. $\tilde{O}$ and $\tilde{G}$ are the mean value of output and ground truth, respectively. p denotes the bit depth of the image. μ_O is the mean of O, $\sigma_O{}^2$ and σ_{OG} are the variance of O and the covariance of O and G, respectively. $C_1=(0.01L)^2$, $C_2=(0.03L)^2$, where L is the dynamic range of the image.

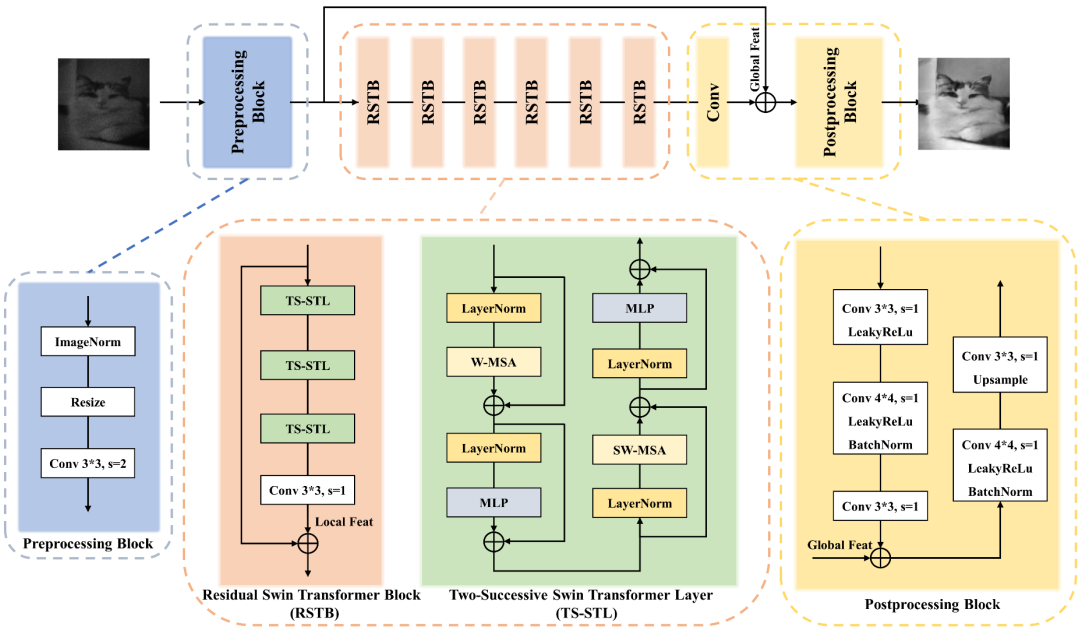

Figure S1. Overall architecture of the swin-model. The network combines with swin transformer and convolution to improve performance. The plus symbol denotes layer concatenation. Conv, convolution; S, stride; SW-MSA, shifted windows multi-head self-attention; MLP, multilayer perceptron.

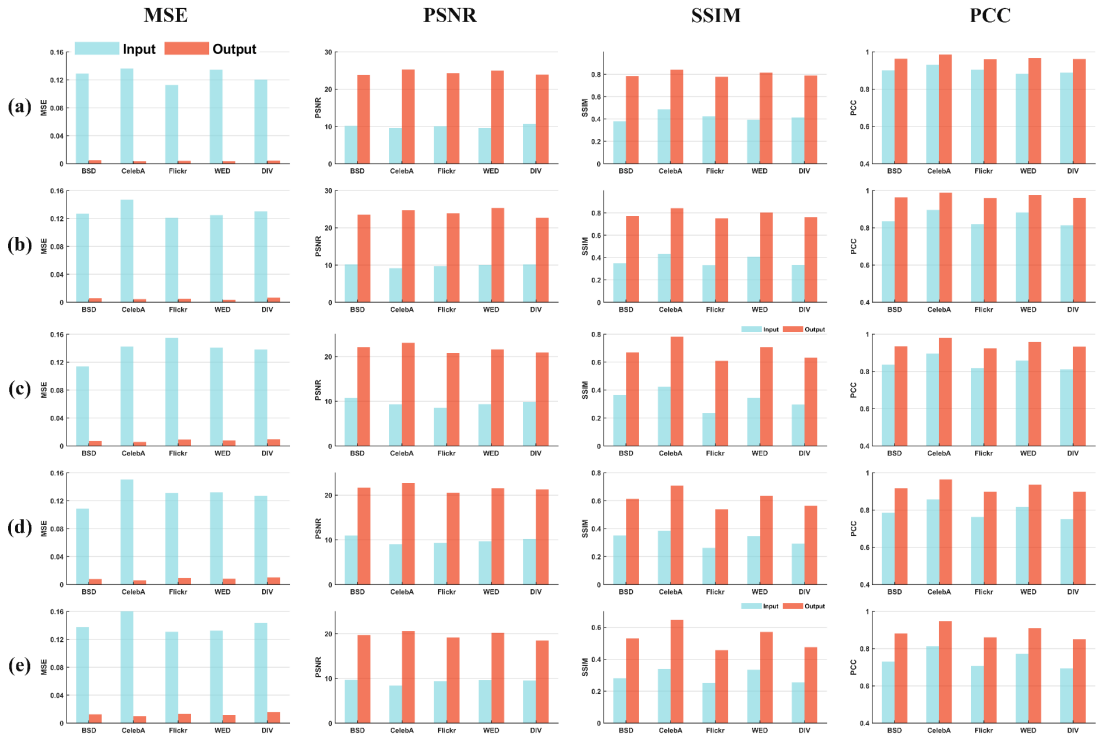

Figure S2. Average results of the evaluation indexes under different spatial coherence. (a)-(e) different spatial coherence (a) spatial coherence is 0.494. (b) spatial coherence is 0.419. (c) spatial coherence is 0.368. (d) spatial coherence is 0.311. (e) spatial coherence is 0.245.

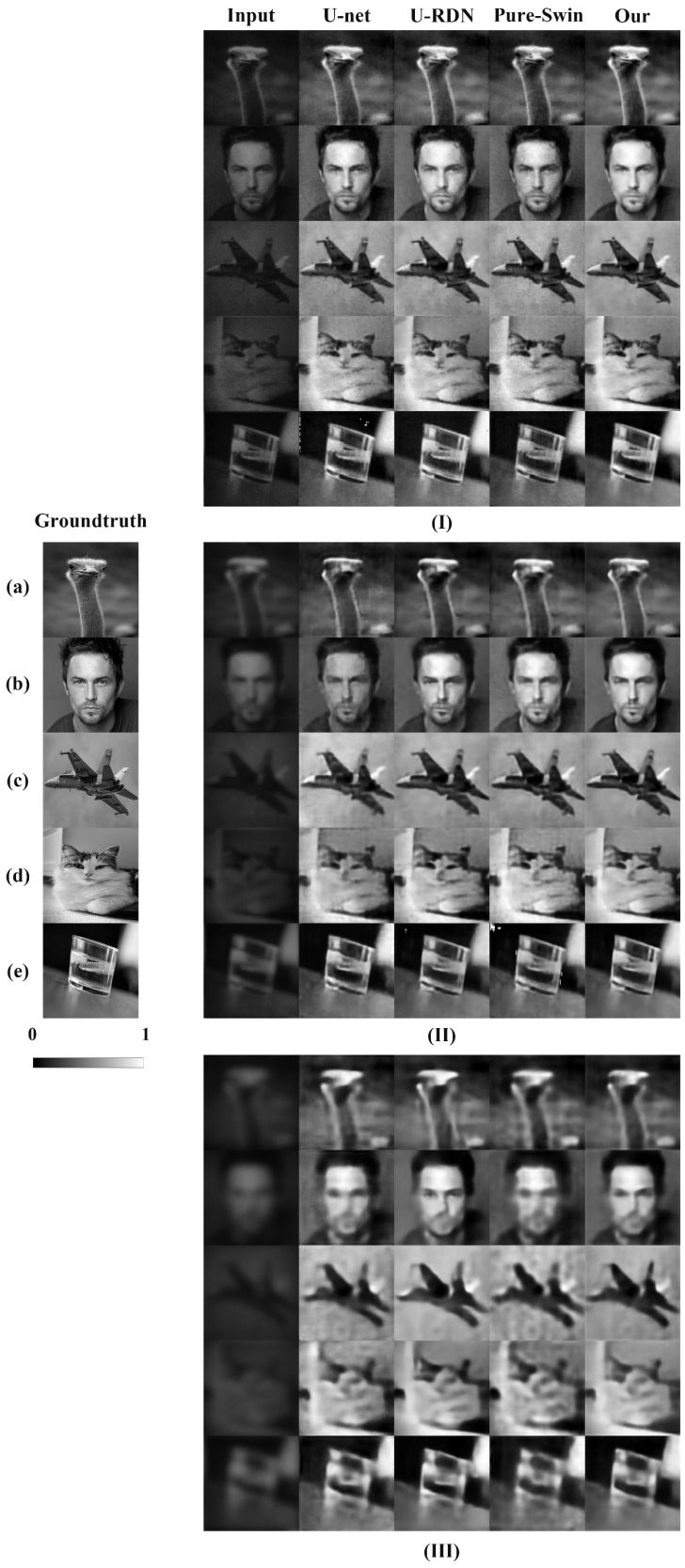

Figure S3. Visualization of performance between different methods. (a) BSD dataset. (b) CelebA dataset. (c) DIV dataset. (d) Flickr dataset. (e) WED dataset. (I) spatial coherence is 0.494. (II) spatial coherence is 0.368. (III) spatial coherence is 0.245. The Pure-Swin model can be obtained by removing the

postprocessing block of swin-model.

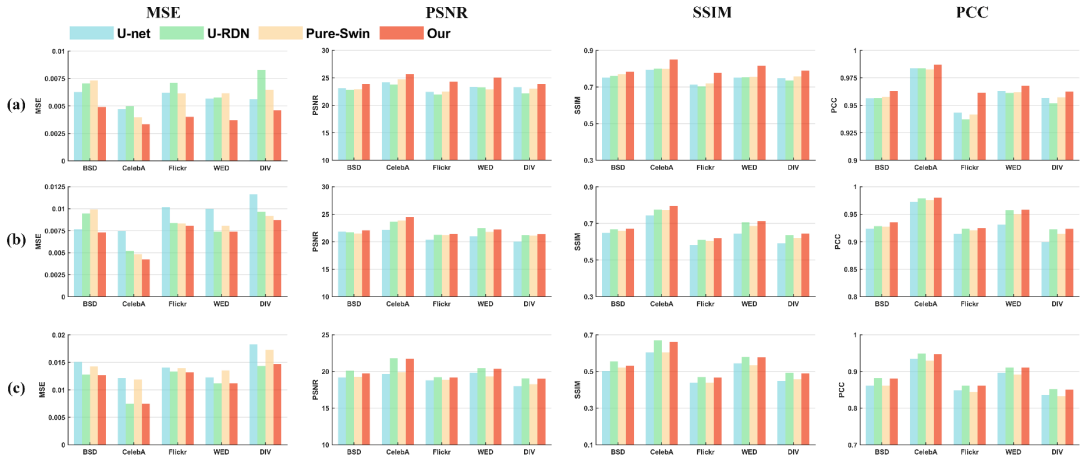

Figure S4. Performance of different methods on various datasets with different spatial coherence. (a) spatial coherence is 0.494. (b) spatial coherence is 0.368. (c) spatial coherence is 0.245.

Distance	**Model space**	**Coherence**	**Oceanic turbulence**	**Coherence**
$D=\mathrm{f}_1$	M_1	0.494	$\chi_t=10^{-9}\,\mathrm{K}^2/\mathrm{s}$	0.491
$D=1.1\mathrm{f}_1$	M_2	0.475	$\chi_t=10^{-7}\,\mathrm{K}^2/\mathrm{s}$	0.483
$D=1.2\mathrm{f}_1$	M_3	0.442	$\chi_t=2\times10^{-7}\,\mathrm{K}^2/\mathrm{s}$	0.447
$D=1.3\mathrm{f}_1$	M_4	0.419	$\chi_t=4\times10^{-7}\,\mathrm{K}^2/\mathrm{s}$	0.404
$D=1.4\mathrm{f}_1$	M_5	0.393	$\chi_t=10^{-6}\,\mathrm{K}^2/\mathrm{s}$	0.373
$D=1.5\mathrm{f}_1$	M_6	0.368	**Atmospheric turbulence**	
$D=1.6\mathrm{f}_1$	M_7	0.337	$C_n^2=10^{-15}\,\mathrm{m}^{3-\alpha}$	0.507
$D=1.7\mathrm{f}_1$	M_8	0.311	$C_n^2=1.5\times10^{-13}\,\mathrm{m}^{3-\alpha}$	0.459
$D=1.8\mathrm{f}_1$	M_9	0.285	$C_n^2=2.5\times10^{-13}\,\mathrm{m}^{3-\alpha}$	0.430
$D=1.9\mathrm{f}_1$	M_{10}	0.251	$C_n^2=3.5\times10^{-13}\,\mathrm{m}^{3-\alpha}$	0.403
$D=2\mathrm{f}_1$	M_{11}	0.245	$C_n^2=5\times10^{-13}\,\mathrm{m}^{3-\alpha}$	0.378

Table S1. Spatial coherence and model space under different distances between L1 and RD and different turbulences. D is the distance between L1 and RD in the experiment and f_1 is the focal length of L1. χ_t represents the rate of dissipation of mean-square temperature. C_n^2 denotes the refractive index structure constant.

Parameter	Definition	Value
κ	Spatial wavenumber	$\kappa = \sqrt{\kappa_x^{\ 2} + \kappa_y^{\ 2} + \kappa_z^{\ 2}}$
χ_t	the rate of dissipation of mean-square temperature	$10^{-10}\mathrm{K}^2/\mathrm{s} \sim 10^{-4}\mathrm{K}^2/\mathrm{s}$
ε	Dissipation rate of turbulent kinetic energy per unit mass	$10^{-4}\mathrm{m}^2/\mathrm{s}^3$
ω	The relative strength of temperature and salinity fluctuations	-3
η	Kolmogorov micro-scale	0.001m
A_T	Temperature dissipation constant	1.863×10^{-2}
A_S	Salinity dissipation constant	1.9×10^{-4}
A_{TS}	Temperature and salinity dissipation constant	9.41×10^{-3}
$C_n^{\ 2}$	The refractive index structure constant	$10^{-18}\mathrm{m}^{3-\alpha} \sim 10^{-13}\mathrm{m}^{3-\alpha}$
α	The refractive index power spectral density power law	11/3
l_0	Inner scale	0.01m
L_0	Outer scale	1m

Table S2. The parameters and the corresponding values in the oceanic turbulence and atmospheric turbulence power spectrum

Evaluation index[a)]	MSE					PSNR				
Spatial coherence	BSD	CelebA	Flickr	WED	DIV	BSD	CelebA	Flickr	WED	DIV
Input_f_1, SC=0.494	0.0499	0.0577	0.1879	0.0604	0.1482	13.4796	12.3874	7.261	12.1869	8.2926
Output_f_1	**0.0016**	**0.001**	**0.0017**	**0.0009**	**0.0030**	**27.8352**	**29.9178**	**27.6036**	**30.4842**	**25.2432**
Input_1.3f_1, SC=0.419	0.0488	0.0643	0.1994	0.0540	0.1598	13.4842	11.9197	7.0031	12.6781	7.9645
Output_1.3f_1	**0.0007**	**0.0008**	**0.0014**	**0.0013**	**0.0021**	**31.4565**	**30.942**	**28.4589**	**28.9769**	**26.8464**
Input_1.5f_1, SC=0.368	0.0381	0.0623	0.2432	0.0628	0.1683	14.1915	12.0522	6.1395	12.0223	7.7399
Output_1.5f_1	**0.0014**	**0.0023**	**0.0022**	**0.0030**	**0.0033**	**28.5689**	**26.4481**	**26.6498**	**25.2361**	**24.8589**
Input_1.7f_1, SC=0.311	0.0351	0.0658	0.2140	0.0615	0.1541	14.5519	11.8155	6.6958	12.1103	8.1221
Output_1.7f_1	**0.0016**	**0.0027**	**0.0040**	**0.0042**	**0.0051**	**28.0903**	**25.7329**	**24.0217**	**23.7305**	**22.8945**
Input_2f_1, SC=0.245	0.0511	0.0769	0.2089	0.0599	0.1725	12.9128	11.1432	6.8009	12.2263	7.6314
Output_2f_1	**0.0044**	**0.0028**	**0.0049**	**0.0051**	**0.0037**	**23.5742**	**25.4541**	**23.1178**	**22.9535**	**24.3529**
Ground truth	0	0	0	0	0					

[a)]**Bold values indicate the better index**

Table S3. Quantitative analysis of evaluation indexes (MSE, PSNR) at the different spatial coherence. f_1 is the focal length of L1. SC means spatial coherence of the light source.

Evaluation index[a)]	MSE					PSNR				
Ocean turbulence	**BSD**	**Celeb**	**Flickr**	**WED**	**DIV**	**BSD**	**Celeb**	**Flickr**	**WED**	**DIV**
Input, O1	0.1431	0.0246	0.0754	0.0424	0.0304	8.4441	16.089	11.2245	13.7200	15.1627
Output, O1	**0.0132**	**0.0051**	**0.0101**	**0.0067**	**0.0067**	**18.7857**	**22.8488**	**19.9562**	**21.7236**	**21.7279**
Input, O2	0.1442	0.0251	0.0797	0.0441	0.0789	8.4110	15.9995	10.9873	13.5560	11.0243
Output, O2	**0.0132**	**0.0056**	**0.0183**	**0.0068**	**0.0074**	**18.7800**	**22.5104**	**17.3632**	**21.6488**	**21.3175**
Input, O3	0.1496	0.0256	0.0828	0.0469	0.0790	8.2488	15.9204	10.8166	13.2849	11.0229
Output, O3	**0.0167**	**0.0084**	**0.0087**	**0.0078**	**0.0075**	**17.7549**	**20.7428**	**20.5745**	**21.0417**	**21.2228**
Input, O4	0.1477	0.0274	0.0843	0.0510	0.0815	8.3063	15.6173	10.7404	12.9202	10.8894
Output, O4	**0.0179**	**0.0095**	**0.0247**	**0.0129**	**0.0109**	**17.4734**	**20.2328**	**16.0698**	**18.8890**	**19.6057**
Input, O5	0.1566	0.0285	0.0931	0.0542	0.0817	8.0511	15.4382	10.3108	12.6533	10.8751
Output, O5	**0.0177**	**0.0084**	**0.0239**	**0.0136**	**0.0134**	**17.5229**	**20.7304**	**16.2072**	**118.6519**	**18.7298**
Ground truth	0	0	0	0	0					

[a)]Bold values indicate the better index

Table S4. Quantitative analysis of evaluation indexes (MSE, PSNR) at the different oceanic turbulence intensities.

Evaluation index[a)]	MSE					PSNR				
Atmosphere turbulence	**BSD**	**Celeb**	**Flickr**	**WED**	**DIV**	**BSD**	**Celeb**	**Flickr**	**WED**	**DIV**
Input, A1	0.1239	0.0223	0.0689	0.0321	0.0575	9.0678	16.5139	11.6172	14.9256	12.4018
Output, A1	**0.0115**	**0.0063**	**0.0097**	**0.0052**	**0.0062**	**19.3732**	**21.9646**	**20.1209**	**22.7763**	**22.0238**
Input, A2	0.1251	0.0231	0.0710	0.0352	0.0633	9.0247	16.3624	11.4835	14.5263	11.9837
Output, A2	**0.0138**	**0.0094**	**0.0224**	**0.0099**	**0.0103**	**18.5798**	**20.2327**	**16.4828**	**20.0040**	**19.8565**
Input, A3	0.1299	0.0248	0.0728	0.0379	0.0598	8.8639	16.0506	11.3795	14.2074	12.2265
Output, A3	**0.0231**	**0.0127**	**0.0249**	**0.0127**	**0.0128**	**16.3665**	**18.9606**	**16.0310**	**18.9301**	**18.9327**
Input, A4	0.1339	0.0247	0.0782	0.0374	0.0638	8.7308	16.0675	11.0632	14.2674	11.9454
Output, A4	**0.0164**	**0.0114**	**0.0287**	**0.0131**	**0.0107**	**17.8445**	**19.4237**	**15.4141**	**18.8084**	**19.6832**
Input, A5	0.1304	0.0253	0.0747	0.0401	0.0665	8.8460	15.9555	11.2651	13.9684	11.7714
Output, A5	**0.0207**	**0.0076**	**0.0225**	**0.0142**	**0.0113**	**16.8227**	**21.1424**	**16.4754**	**18.4686**	**19.4552**
Ground truth	0	0	0	0	0					

[a)]Bold values indicate the better index

Table S5. Quantitative analysis of evaluation indexes (MSE, PSNR) at different atmospheric turbulence

intensities.

Evaluation index[a)]	MSE					PSNR				
Methods	**BSD**	**Celeb**	**Flickr**	**WED**	**DIV**	**BSD**	**Celeb**	**Flickr**	**WED**	**DIV**
Input, SC=0.494	0.0449	0.0577	0.1879	0.0604	0.1482	13.4796	12.3874	7.2610	12.1869	8.2926
U-net	0.0014	0.0026	0.0017	0.0021	0.0034	27.5212	25.8284	27.6284	26.8457	24.6679
U-RDN	0.0032	0.0031	0.0019	0.0029	0.0047	24.9734	25.0346	27.2389	25.3533	23.2954
Pure-Swin	0.0022	0.0016	0.0018	0.0031	0.0042	27.1912	28.0367	27.3320	25.1208	23.7172
Our	**0.0016**	**0.001**	**0.0017**	**0.0009**	**0.0016**	**27.8352**	**29.9178**	**27.6036**	**30.4842**	**25.2432**
Input, SC=0.368	0.0381	0.0623	0.2432	0.0628	0.1683	14.1915	12.0522	6.1395	12.0223	7.7399
U-net	0.0024	0.0057	0.0062	0.0025	0.0066	26.2406	22.4193	22.0768	25.9566	21.7759
U-RDN	**0.0012**	0.0013	0.0038	**0.0024**	0.0016	29.2139	**28.9604**	24.2183	26.2102	27.8569
Pure-Swin	0.0020	0.0026	0.0048	0.0023	0.0021	26.9187	25.8403	23.1514	**26.3914**	26.7619
Our	0.0014	**0.0023**	**0.0022**	0.0030	**0.0033**	**28.5689**	26.4481	**26.6498**	25.2361	**27.8589**
Input, SC=0.245	0.0511	0.0769	0.2089	0.0599	0.1725	12.9128	11.1432	6.8009	12.2263	7.6314
U-net	0.0048	0.0045	0.0069	0.0073	0.0117	23.1964	23.4604	21.6195	21.3607	19.3551
U-RDN	**0.0023**	0.0035	**0.0045**	0.0053	0.0042	**26.3226**	24.5103	**23.4867**	22.7962	23.7401
Pure-Swin	0.0032	0.0060	0.0086	0.0075	0.0052	24.9252	22.1834	20.6685	21.2618	22.8502
Our	0.0044	**0.0028**	0.0049	**0.0051**	**0.0037**	23.5742	**25.4541**	23.1178	**22.9535**	**24.3529**
Ground truth	0	0	0	0	0					

[a)]Bold values indicate the better index

Table S6. Quantitative comparison of evaluation indexes (MSE, PSNR) between different methods under low-spatial coherence. These values are calculated from samples of Figure S2.

Evaluation index[a)]	SSIM					PCC				
Methods	**BSD**	**Celeb**	**Flickr**	**WED**	**DIV**	**BSD**	**Celeb**	**Flickr**	**WED**	**DIV**
Input, SC=0.494	0.5893	0.5943	0.4296	0.6155	0.4625	0.9368	0.9575	0.9210	0.9146	0.8753
U-net	0.8623	0.8433	0.8313	0.8528	0.8379	0.9752	0.9808	0.9834	0.9875	0.9745
U-RDN	0.8722	0.8283	0.8289	0.8260	0.8462	0.9787	0.9888	0.9800	0.9858	0.9712
Pure-Swin	0.8633	0.8308	0.8073	0.8396	0.8127	0.9763	0.9866	0.9798	0.9808	0.9783
Our	**0.8984**	**0.8908**	**0.8523**	**0.9019**	**0.8940**	**0.9807**	**0.9893**	**0.9848**	**0.9930**	**0.9819**
Input, SC=0.368	0.6178	0.5394	0.2777	0.5677	0.3892	0.8957	0.9211	0.8396	0.8961	0.8144
U-net	0.8471	0.7828	0.7712	0.8452	0.8257	0.9329	0.9790	0.9698	0.9807	0.9339
U-RDN	0.8888	0.8450	0.7937	0.8534	**0.8662**	0.9718	**0.9888**	0.9763	0.9865	0.9714
Pure-Swin	0.8615	0.8298	0.7822	0.8299	0.8567	0.9504	0.9844	0.9748	0.9845	0.9593
Our	**0.8906**	**0.8513**	**0.8171**	**0.8541**	0.8622	**0.9691**	0.9881	**0.9783**	**0.9869**	**0.9780**
Input, SC=0.245	0.4881	0.4469	0.3073	0.5271	0.3643	0.8072	0.8817	0.7557	0.8326	0.7196
U-net	0.7709	0.7042	0.6385	0.6792	0.6942	0.9307	0.9655	0.9339	0.9528	0.9121
U-RDN	**0.8287**	**0.7701**	**0.6980**	**0.7914**	**0.7732**	**0.9443**	**0.9739**	0.9490	**0.9661**	**0.9362**
Pure-Swin	0.7910	0.7038	0.6244	0.6810	0.6905	0.9326	0.9525	0.9236	0.9405	0.9147
Our	0.8146	0.754	0.6962	0.7722	0.7572	0.9431	0.9713	**0.9505**	0.9631	0.9341
Ground truth	1	1	1	1	1	1	1	1	1	1

[a)]Bold values indicate the better index

Table S7. Quantitative comparison of evaluation indexes (SSIM, PCC) between different methods under low-spatial coherence. These values are calculated from samples in Figure S2.