Research in Astronomy and Astrophysics, 24:095015 (11pp), 2024 September https://doi.org/10.1088/1674-4527/ad4fc1

Super-resolution Imaging of Telescopic Systems based on Optical-neural Network Joint Optimization

You-Hong Sun[1,2], Tao Zhang[3], Hao-Dong Shi[1,2], Qiang Fu[1,2], Jia-Nan Liu[1,2], Kai-Kai Wang[1], and Chao Wang[1,2]
[1] Research Institute of Chongqing, Changchun University of Science and Technology, Changchun 130022, China; 1768811682@qq.com
[2] National and Local Joint Engineering Research Center for Space Photoelectric Technology, Changchun University of Science and Technology, Changchun 130022, China
[3] The Institute of Remote Sensing Satellites, China Academy of Space Technology (CAST), Beijing 100094, China
Received 2023 December 22; revised 2024 April 22; accepted 2024 May 7; published 2024 September 18

Abstract

Optical telescopes are an important tool for acquiring optical information about distant objects, and resolution is an important indicator that measures the ability to observe object details. However, due to the effects of system aberration, atmospheric seeing, and other factors, the observed image of ground-based telescopes is often degraded, resulting in reduced resolution. This paper proposes an optical-neural network joint optimization method to improve the resolution of the observed image by co-optimizing the point-spread function (PSF) of the telescopic system and the image super-resolution (SR) network. To improve the speed of image reconstruction, we designed a generative adversarial net (LCR-GAN) with light parameters, which is much faster than the latest unsupervised networks. To reconstruct the PSF trained by the network in the optical path, a phase mask is introduced. It improves the image reconstruction effect of LCR-GAN by reconstructing the PSF that best matches the network. The results of simulation and verification experiments show that compared with the pure deep learning method, the SR image reconstructed by this method is rich in detail and it is easier to distinguish stars or stripes.

Key words: Techniques: image processing – Telescopes – Stars: imaging

1. Introduction

Optical telescopes are widely used in astronomical observation, remote sensing, and optical surveillance, as an important tool for obtaining optical information on long-range targets (Wang et al. 2015, 2020; Mikhail et al. 2019; He et al. 2021). The resolution reflects its ability to distinguish two adjacent objects and is an important indicator that measures the ability to observe the object details. Higher resolution often means richer object detail. According to the Rayleigh criterion, the resolution of the system can theoretically reach the diffraction limit. However, due to the effects of system aberration, atmospheric seeing, and other factors, the observed image of ground-based telescopes is often degraded, resulting in reduced resolution (Li et al. 2018). Super-resolution (SR) technology can reconstruct high-resolution (HR) images from one or more low-resolution (LR) observation images to improve image resolution (Park et al. 2003). Among the mainstream image SR algorithms, the algorithm based on interpolation operation has a mediocre effect on the reconstruction of fine textures (Lehmann et al. 1999). The algorithm based on prior learning is sensitive to the selection of training samples, and cannot recover the high and low-frequency information in the original astronomical images at the same time (Yan et al. 2015). It is difficult for Iterative Back Projection (Irani & Park 1991) and Projection Onto Convex Sets (Youla & Webb 1982) to reconstruct the texture features of astronomical images.

In recent years, with the vigorous development of deep learning, a large number of neural networks continue to emerge. Through the combination of different convolution layers, linear layers, activation functions, etc., deep convolutional neural networks can learn the most complex functional features and thus solve unsolvable problems in many fields (Wang et al. 2023c). It is also widely used in telescopic systems, such as denoising and SR (Rahman et al. 2020; Sweere et al. 2022). In 2022, Shoubaneh et al. used an improved generative adversarial network model to improve the resolution of ground-based images from the Subaru Hyper Suprime-Cam to that of the Hubble Space Telescope (Shoubaneh et al. 2022). Most methods based on convolutional neural networks require paired data for supervised training, considering that the cost of producing paired data in astronomical observation is usually high, and there are some untrue details in the images synthesized by the algorithm, which affects the generalization ability of the model. In contrast, an unsupervised network breaks the limitation of a supervised learning network that relies on LR–HR paired data to improve image resolution (Chang & Wetzstein 2019). It can learn distribution information from HR observation images with similar celestial environments and use it as prior information to improve the resolution of observation images of distant galaxies.

From the perspective of the overall process of optical imaging, the imaging process of optical systems is essentially equivalent to optical calculation (Yuan et al. 2020; Wang et al. 2023a). The goal of the joint optimization method is to realize the optimal matching between the optical system and the image processing algorithm through the cooperative optimization of "optical system" and "digital image post-processing," so as to obtain a better imaging effect. In 2019, Peng et al. from Stanford University proposed an end-to-end design method from optical system to image processing, which achieves achromatic depth-of-field expansion and SR imaging (Peng et al. 2015, 2019). In parallel to ours, Sun Q et al. use an end-to-end framework to simultaneously optimize the single-lens imaging system and a reconstruction network to reconstruct SR images from raw measurement data (Sun et al. 2020).

This paper proposes an optical-neural network joint optimization method to improve the resolution of observation images of telescopic systems for known objects with basic outlines or shapes, such as stars and nebulae. First, the deep learning convolutional layer equivalent to the telescopic system point-spread function (PSF) is integrated into the front end of the neural network, and the PSF and network parameters are co-optimized through the training process. After the training, a phase mask is designed to reconstruct the PSF of network training in the optical path. Considering the error between the actual PSF of phase mask reconstruction and the PSF of network training, it is necessary to add the phase mask to the telescopic system and collect its observation image, and retrain the network to fine-tune the network parameters. Finally, we use Peak Signal-to-Noise Ratio (PSNR) and other indicators to evaluate the quality of reconstructed images. In this paper, HR observation images can be reconstructed quickly by designing a high-performance unsupervised network and co-optimizing the network parameters and the PSF of the telescopic system.

This paper is organized as follows: In Section 2, we give the theoretical background of joint optimization methods. In Section 3, we describe the unsupervised network structure and loss function. In Sections 4 and 5, we evaluate the network performance and the SR effect of the joint optimization method. In Section 6, we conclude.

2. Theoretical Background

2.1. Incoherent Imaging Model and Cooperative Optimization

First, we introduce the relationship between the incoherent imaging model of telescope systems and the convolution layer of deep learning. The illumination source of the telescope system is usually incoherent light, which belongs to a kind of passive imaging (Yang et al. 2022). When an object is imaged incoherently by the telescope system, the linear spatially invariant imaging model can be expressed as (Goodman 2005)

$$I_{im}(u, v) = |H(u, v)|^2 * I_g(u, v), \quad (1)$$

in which u and v are the spatial coordinates of the image plane, $I_g(u, v)$ is the ideal geometric irradiated image, and $|H(u, v)|^2$ is often referred to as the PSF, where $*$ denotes a two-dimensional convolution. The imaging of the telescope system can be modeled as a space-invariant convolution of the object with the PSF of the system.

According to the theory, the incoherent imaging model for monochromatic light illumination can be modeled as a convolutional layer of deep learning (Chang et al. 2018). The imaging of the telescopic system for monochromatic illumination discussed in this paper can be modeled as a convolution layer: The flipped PSF is used as the convolution kernel of the convolution layer, the number of input and output channels are both 1, and the input and output feature map corresponds to the object and image, respectively. Therefore, we use a single-core convolutional layer to simulate the PSF of the optical system, which is integrated into the front end of the neural network, as diagrammed in Figure 1.

To improve the observation ability of the telescopic system on distant celestial bodies, as depicted in Figure 1, HR stellar images with similar celestial environments are used as the original HR images. The HR image is input into the joint optimization network for training, that is, the HR image is first convolved with the equivalent convolution layer to synthesize the LR image, which is equivalent to the incoherent imaging process of the telescope system. Then the synthesized LR image is input into the deep learning network to reconstruct the SR image. The loss function between the SR image output by the network and the HR image is calculated, and the parameters of the network and the equivalent convolution layer (PSF of optical systems) are updated simultaneously by gradient feedback. The training process is repeated until an optimal match between the optical system encoding and the neural network decoding is achieved. After training, the parameters of the equivalent convolution layer are derived as the PSF of the optical system.

Next, it is necessary to physically implement the PSF of network training in the optical path to realize the joint optimization of the optical-neural network in the physical sense. In this paper, a phase mask is used to modulate the two-dimensional phase distribution in the optical field, to obtain the ideal PSF in the image plane.

2.2. Phase Mask Solution

In this paper, a double-glued telescopic objective (DGTO) was used instead of the actual telescopic system to simplify the derivation process. As is displayed in the lower-left corner of Figure 1, a phase mask is placed immediately behind the DGTO with the same net aperture as the DGTO. The detector is placed in the image plane of the DGTO.

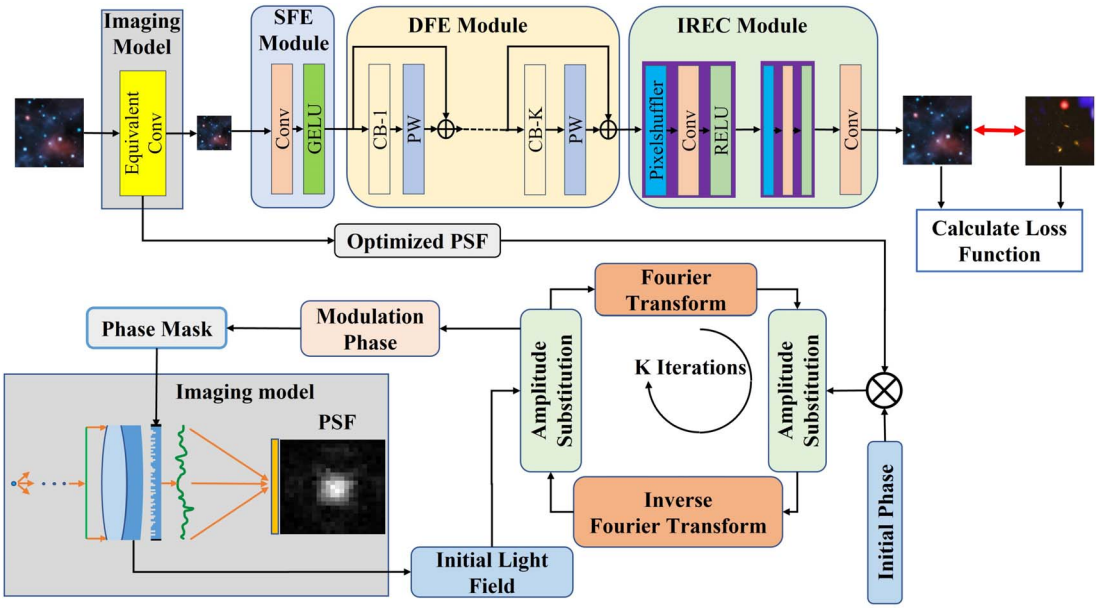

Figure 1. The framework for joint optimization method and example of a telescope system solving for a phase mask.

Suppose that the light emitted by a point light source on the surface of the object propagates to the telescope system, and the initial amplitude of the light field formed on the surface behind the last lens is $U_1(x_1, y_1)$ and the initial phase is $\psi(x_1, y_1)$, Suppose that the modulation phase of the phase mask is $\phi(x_1, y_1) + (-\frac{ik(x_1^2+y_1^2)}{2z_1})$, where $(-\frac{ik(x_1^2+y_1^2)}{2z_1})$ is the preset phase part. z_1 denotes the distance between the rear surface of the telescope system and the image surface. Considering that the phase mask is close to the rear surface of the telescope system and generally has a small thickness, (x_1, y_1) denotes the position coordinate of the rear surface of the telescopic system and also represents the position coordinate of the phase mask. According to the scalar diffraction theory (Goodman 2005; Wang et al. 2023b), the complex amplitude of the light field in the image plane is

$$U(x, y) = \frac{e^{ikz_1}}{i\lambda z_1} \iint_{\Sigma} U_1(x_1, y_1) e^{i\psi(x_1,y_1)} e^{\phi(x_1,y_1)} \times e^{-\frac{ik(x_1^2+y_1^2)}{2z_1}} e^{\frac{ik((x-x_1)^2+(y-y_1)^2)}{2z_1}} dx_1 dy_1, \quad (2)$$

where $k = \frac{2\pi}{\lambda}$ is the wavenumber. (x, y) denotes the position coordinates of the image surface. Removing unnecessary parameters and simplifying yields

$$U(x, y) = \iint_{\Sigma} U_1(x_1, y_1) e^{i\psi(x_1,y_1)} e^{\phi(x_1,y_1)} e^{-\frac{ik(xx_1+yy_1)}{z_1}} dx_1 dy_1. \quad (3)$$

Equation (3) is essentially a Fourier transform; then the PSF of the system can be expressed as

$$\mathrm{PSF} = |U(x, y)|^2 \propto |\boldsymbol{F}(U_1(x_1, y_1) e^{i\psi(x_1,y_1)} e^{\phi(x_1,y_1)})|^2. \quad (4)$$

$\boldsymbol{F}$ denotes the operator of the Fourier transform. According to Equation (4), we set the PSF of the telescope system as the PSF of the network training. We set up the telescope system in the non-sequential mode of optical design software, and then trace one million grid rays from the point source. A detector is set on the back surface of the system to record the coherent superposition of complex amplitudes of the plane waves corresponding to all rays, to obtain the initial light field of Equation (2), $U_1(x_1, y_1) e^{i\psi(x_1,y_1)}$.

Next, the Gerchberg–Saxton (GS) phase retrieval algorithm is used to solve the modulation phase $\phi(x_1, y_1)$. The GS phase retrieval algorithm is a method to recover the phase of objects by iteration. It has strong anti-disturbance ability, and is an important tool for phase recovery (Gerchberg & Saxton 1972).

Finally, according to the equation that relates the phase difference with the thickness of the diffraction optical element (Wang et al. 2022; Xu et al. 2022), we design a mask about the physical height on the substrate with the refractive index n, which is the phase mask. As shown in the following equation

$$\phi(x_1, y_1) = (n - 1)\frac{2\pi h(x_1, y_1)}{\lambda}, \quad (5)$$

where λ is the wavelength of the point source. (x_1, y_1) denotes the spatial position of the phase mask. Considering the error between the actually recovered PSF and the trained PSF, the manufactured phase mask is added to the telescopic system, the observed image of the distant object is acquired as LR data, and the network is retrained to fine-tune the parameters of the reconstructed network.

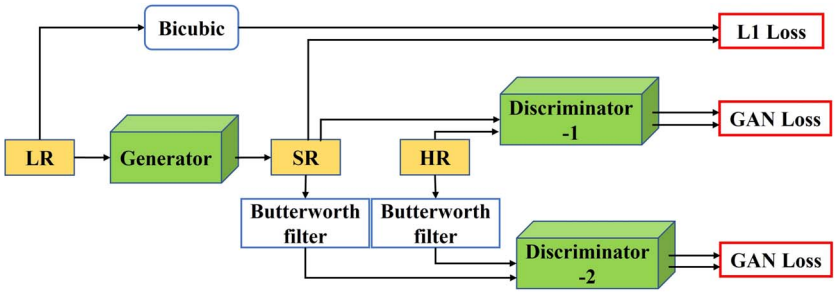

Figure 2. General architecture of the proposed LCR-GAN.

3. Unsupervised Image Super-resolution Network

The proposed joint optimization method based on deep learning uses the designed unsupervised network Lightweight Cascaded Residual Network Using Generative Adversarial Network (LCR-GAN) as the basis for efficient reconstruction of SR images. LCR-GAN has the advantages of light parameters and fast reconstruction speed, which can be easily applied to the real-time tasks of mobile low-cost devices, so it is conducive to the practical application of the joint optimization method. The model framework of LCR-GAN is drawn in Figure 2, which consists of three parts: generator, discriminator and loss function. These three parts are covered in Sections 3.2 and 3.3.

3.1. Network Architecture

A generative adversarial network obtains realistic images by game learning between generators that generate close to real images and discriminators that distinguish false images. It is one of the most promising methods of unsupervised learning (Ledig et al. 2017). In this paper, a generative adversarial network model is adopted to build an unsupervised network, as depicted in Figure 2. Bicubic denotes the interpolated up-sampling operation. The red boxes correspond to the mean absolute error and the adversarial loss. To improve the generator's ability to extract detailed features, we add a detail discriminator that enables the network to pay more attention to high-frequency components in the input image, thus making the reconstructed image texture clearer. For this purpose, a Butterworth high-pass filter is introduced before the second discriminator to filter high-frequency components from the input image. A second discriminator is then used to distinguish the spurious details generated by the generator.

In LCR-GAN, the generator extracts features of different levels from the input image, fuses them through the upper sampling layer, and reconstructs an image of the same size as the ground truth image. The discriminator takes the reconstructed image and the real image as input, and its function is to distinguish the fake image generated by the generator. The two discriminators have the same structure.

3.2. Generator and Discriminators

SR Using a Generative Adversarial Network (SRGAN) has been widely investigated in the direction of image SR because it can reconstruct fine texture details when the up-sampling factor is large (Ledig et al. 2017). To this end, we build a new SR network based on the SRGAN framework, as illustrated in Figure 3, to achieve high-performance reconstruction of SR images. SRGAN is modified as follows: First, the 16-layer residual block in the SRGAN network is modified to the 7-layer cascading block (CB). The CB can make full use of the feature information extracted at all levels and optimize and extend the propagation path of information, thus reducing the parameter redundancy of the network and improving the reconstruction efficiency of the model. Second. The activation function PReLU is replaced by the activation function GELU. GELU has smoother nonlinear characteristics, which can improve the performance of the model and accelerate the convergence rate of the model.

As shown in Figure 3, the network is divided into three parts. SFE Module, DFE Module, and Irec Module respectively represent Shallow Feature Extraction Module, Deep Feature Extraction Module, and Image reconstruction Module. K denotes the number of CBs in the network, and PW denotes pointwise convolution. After the LR image is input into the network, the deep feature information is extracted using seven CBs. Then the image reconstruction module fuses all the extracted features to reconstruct the SR image.

The structure of the CB is depicted in Figure 3. Among them, the residual block (ResBlock) is responsible for extracting feature information from the input feature map, and the concatenation operator collects feature information extracted at different levels. The PW layer fuses the feature maps extracted from previous layers and compresses the number of feature channels, thus reducing the number of training parameters of the residual block. The residual block first learns the feature information on the feature graph F_1 with input size $C \times H \times W$. Then it is cascaded with feature graph F_1 to obtain feature image F_2 with size $2C \times H \times W$. After two processes of point convolution layer compression feature channel number, the residual block extraction feature information and the

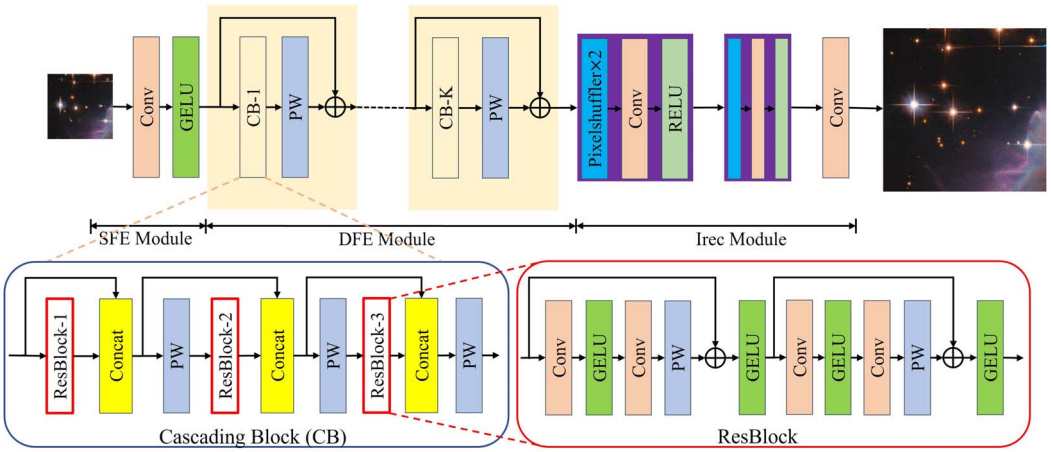

Figure 3. Network structure of generator (G) in LCR-GAN.

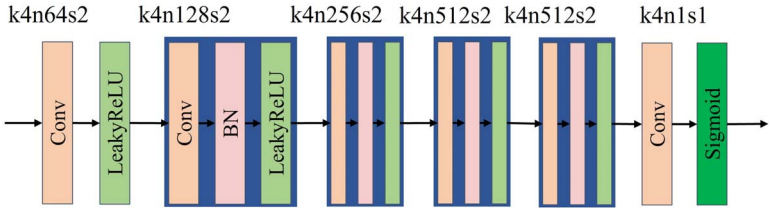

Figure 4. Network structure of the PatchGAN discriminator.

concatenation operator collection the output feature map of the residual block, a deep feature with a size of $C \times H \times W$ is obtained by point convolution layer compression. To learn deeper features, two tandem residual units are used to form a residual block, as shown in the lower right of Figure 3.

Considering that PatchGAN trains the discriminator by cutting the input image into small pieces to discriminate them separately, which enhances the local texture details (Phillip et al. 2017), we use PatchGAN as the discriminator, with both discriminators having the same architecture. The network model is visualized in Figure 4. BN represents the batch normalization layer. k, n, and s denote the size of the filter kernel, the number of output feature channels, and the stride of the convolution layer, respectively. A convolution layer, a BN layer, and an activation function LeakyReLU are integrated into each dark blue box. As demonstrated in Figure 4, the input passes through each convolution layer in turn and is mapped into an $N*N$ matrix, where each point represents a patch in the input image. Finally, a sigmoid activation function is used to discriminate the probability that different patches are true.

3.3. Loss Function

To accurately reconstruct SR images, we use the following loss function to train the generator

$$\text{Loss}_G = L_p + \lambda_1 L_G^1 + \lambda_2 L_G^2. \quad (6)$$

Loss_G is a weighted combination of the pixel loss L_p and the generative adversarial loss L_G^1 of the first discriminator, and the adversarial loss L_G^2 of the second discriminator. Here λ_1 and λ_2 are the weights of the two losses respectively. The expression for the pixel loss L_p is as follows

$$L_p = \frac{1}{N}\sum_N |I_{\text{SR}} - B_{\text{LR}}|. \quad (7)$$

Here, I_{SR} represents the output image of the network, and B_{LR} represents the results of the bicubic upsampling operation of the input LR image.

The expression for the generative adversarial loss L_G^1 of the first discriminator is as follows

$$L_G^1 = \frac{1}{N}\sum_N |1 - D_1(I_{\text{SR}})|. \quad (8)$$

I_{SR} represents the output image of the network, and D_1 represents the output result of the discriminator. The expression for the generative adversarial loss L_G^2 of the second discriminator is as follows

$$L_G^2 = \frac{1}{N}\sum_N |1 - D_2(W_2(I_{\text{SR}}))|. \quad (9)$$

D_2 represents the output of the second discriminator, and W_2 represents the weight of the convolution kernel of the Butterworth high-pass filter.

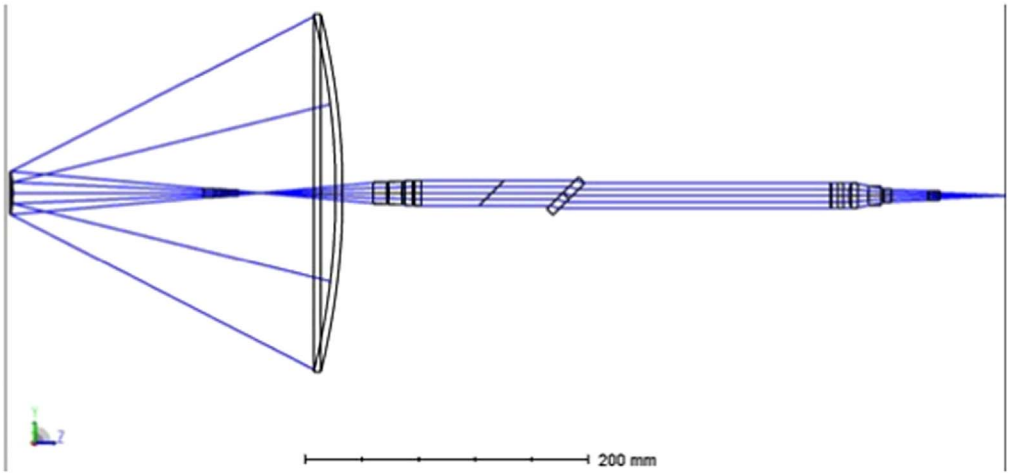

Figure 5. Optical path diagram of the astronomical telescopic system.

To better distinguish between the generator's reconstructed image and the HR images (label data), the first discriminator is trained using the following loss function

$$\mathrm{Loss}_{D_1} = \frac{1}{N}\sum_{N}\left(\frac{|D_1(I_{\mathrm{SR}})| + |1 - D_1(I_{\mathrm{HR}})|}{2}\right), \tag{10}$$

where I_{HR} represents the HR images (label data), and other parameters are as above.

To better distinguish the spurious detail features reconstructed by the generator, the second discriminator is trained using the following loss function

$$\mathrm{Loss}_{D_2} = \frac{1}{N}\sum_{N}\left(\frac{|D_2(W_2(I_{\mathrm{SR}}))| + |1 - D_2(W_2(I_{\mathrm{HR}}))|}{2}\right), \tag{11}$$

where D_2 represents the output result of the second discriminator, and W_2 represents the weight of the convolution kernel of the Butterworth high-pass filter.

4. Simulation of Joint Optimization Method

To verify the performance of the joint optimization method, we carried out SR reconstruction of degraded images by deep learning and joint optimization, and evaluated the image quality. Specifically, we first make the corresponding star atlas data sets in two ways, and use the produced data sets to train the two latest unsupervised image SR networks Direct Unsupervised Super-Resolution Using Generative Adversarial Network (DUS-GAN; Prajapati et al. 2021), Metric Learning based Interactive Modulation for Real-World Super-Resolution (MM-realSR; Mou et al. 2022), LCR-GAN and LCR-GAN of the joint optimization method until the networks converge. Then the degraded star atlas in the test set are input into the trained network to reconstruct the deep learning SR image and the jointly optimized SR image. Image quality evaluation indexes PSNR, Structural Similarity Index Metric (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) were used to evaluate the reconstruction effect.

4.1. Data Preparation

The actual astronomical telescopic system was used for the simulation test, which is utilized to identify and track the space point target, and its optical path diagram is drawn in Figure 5. The entrance pupil diameter is 250 mm, the focal length is 2990 mm, and the wavelength is 0.623 μm. When the object height is set to 0.1° the image surface size is 5.23 mm × 5.23 mm. The pixel size is 192 × 192, then the spacing between adjacent pixels is about 27.24 μm. Select 300 clear and easily distinguishable HR star atlases on the NASA website.[4] These atlases are images of celestial objects captured by different astronomical telescopes, ensuring the authenticity of the data. The data are publicly available, and the data set can be provided as an attachment. Convert them to grayscale images as the original HR images.

Second, we introduce two methods for making degraded images or LR images of deep learning and joint optimization methods. Traditional deep-learning reconstruction methods require LR star atlases. Therefore, we construct a 1:1 imaging optical path of the actual astronomical telescopic system in optical design software, set the corresponding basic parameters according to the actual imaging scene, and use the image simulation function to get the image plane image of the telescopic system. For the joint optimization method proposed in this paper, as described in Section 2.1, a learnable single-core convolution layer (convolution kernel size is 23 × 23, step is 1) is used as the PSF of the actual telescopic system, and convolved with the HR image to generate an LR image to match the incoherent imaging process of the telescopic system.

[4] https://images.nasa.gov/

The images obtained by the two methods are downsampled 4 times as the LR image to match the possible low sampling rate detector. In this paper, in view of the large object distance of the telescope system and the fact that astronomical stars are difficult to reproduce in reality, weighing the fidelity of the data against the cost of production, we use both methods to produce the degraded data required for this article.

According to the degraded data production method of deep learning above, 300 pairs of LR–HR paired star atlases can be easily obtained. To gather enough data for training, we randomly rotated by 90° and flipped for data augmentation. For the obtained 1200 pairs of data, 960 pairs were extracted for the training set, 120 pairs for the verification set, and 120 pairs for the test set. Then we randomly cropped 48×48 and 192×192 unpaired patches from the training data set, for training the SR network DUS-GAN, MM-realSR, and LCR-GAN.

According to the degraded data production method of the joint optimization method above, the initial HR star atlas is randomly rotated by 90° and flipped for data augmentation. The 1200 images obtained were randomly cropped into 192×192 unpaired patches, such that 960 were used for the HR data of the training set, 120 were used for the HR data of the verification set, and 120 were used for the HR data of the test set. The HR images in the training set are convolved with a learnable single-core convolution layer, and then the obtained images were down-sampled 4 times to generate LR images, which were used to train the image reconstruction network LCR-GAN of the joint optimization method. As mentioned in Section 2.2, after the training, the PSF recovered by the GS algorithm is taken as the PSF of the actual telescopic system. The HR images in the test set are convolved with this PSF, and then the LR images of the test set are generated by 4 times down-sampling.

4.2. Simulation Result Analysis

Figure 6 displays the simulation results of the GS algorithm implementation described in Section 2.2. The average absolute error between the PSF generated by the algorithm and the PSF trained by the network converges quickly. The final Root Mean Square Error of the two is 0.0189, and the sum of squares due to error (SSE) is 0.0097, which means that the modulation phase obtained by the algorithm can more accurately realize the PSF trained by the network.

In the training phase, 960 sets of two kinds of unpaired images were used as the input pictures of their respective unsupervised networks, and 300,000 iterations were carried out. Of these, 960 are LR images and 960 are HR images. The learning rate is 0.0001, batch size is 1 and epoch is 300. Our implementation uses the PyTorch framework on a PC with the GeForce RTX 4090 (NVIDIA).

After the training, 60 sets of data are randomly selected from both the test sets of deep learning and joint optimization. The LR images are input into the corresponding network to reconstruct the SR star atlas. Three quantitative picture quality indices are used for performance evaluation, including PSNR, SSIM, and LPIPS.

Table 1 lists the average indicators of reconstructed images by different methods. We observe that the PSNR and SSIM values of LCR-GAN are better than MM-RealSR, and are comparable to bicubic upsampling and DUS-GAN, which indicate that the reconstructed images of LCR-GAN have less distortion, better structural similarity, and higher image quality. The image reconstruction time of LCR-GAN is much smaller than that of the other two networks, and the ultra-fast reconstruction speed is conducive to real-time reconstruction. Deep learning methods and bicubic upsampling methods are significantly lower than joint optimization methods in PSNR, SSIM, and LPIPS, which indicate that the SR image reconstructed by the joint optimization method is closer to the real image.

We compare the training parameters of several main unsupervised networks, as shown in Table 2. The parameters of LCR-GAN are smaller than those of other unsupervised networks. The simplified model not only improves the speed of image reconstruction but also can be deployed in low-cost computing devices, which is conducive to the practical application of deep learning.

Figure 7 shows the effect of the 4x super-resolved star atlas reconstructed by the different methods. According to the Rayleigh criterion, the diffraction limit of the system is about 9.12 μm, much smaller than the pixel spacing of 27.24 μm, which means that the system can theoretically distinguish between adjacent pixels. In Figure 7, the image in the image plane was down-sampled 4 times (matching the low sampling detector) to generate LR. It can be seen that due to the system aberration and other reasons, as well as the loss of detail information caused by 4x downsampling, the adjacent weak star points in LR are difficult to distinguish. The adjacent weak stars in the reconstructed image of LCR-GAN are easy to distinguish, which indicates that LCR-GAN has a good image SR effect. Compared with DUS-GAN and MM-realSR, the SR star atlas reconstructed by LCR-GAN has relatively clear star points and textures, and the weak star points are well preserved. Compared with the LCR-GAN, the star atlas reconstructed by the joint optimization method is rich in detail, with clearer nebular texture, and adjacent faint stars are distinguishable. We explain that the joint optimization method reconstructs the PSF of the telescopic system, so that the system retains the optimal information, and therefore the reconstruction results are closer to the original HR image.

5. Verification Experiment

5.1. Experimental Setup

By using deep learning, Sun's method (Sun et al. 2020) and the joint optimization method in this paper, the SR reconstruction experiment of the resolution plate is carried out to verify

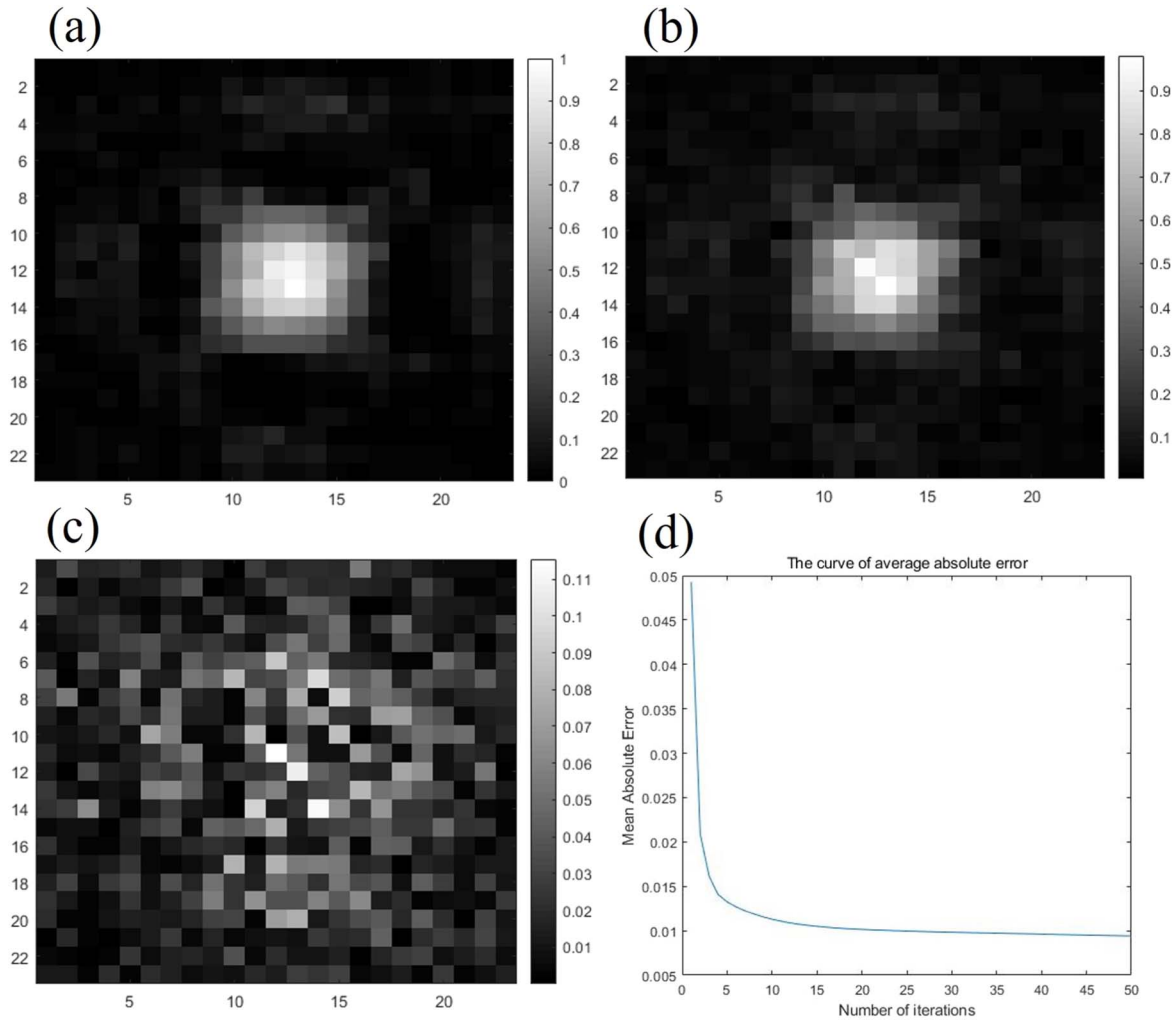

Figure 6. Results of the implementation of the GS phase retrieval algorithm. (a) PSF trained by the deep learning network; (b) PSF recovered by the GS algorithm; (c) absolute error between (a) and (b); (d) the curve of the mean absolute error between (A) and (B) with iteration.

Table 1
Comparison of PSNR, SSIM, LPIPS, and Total Reconstruction Times for some up-sampling Methods

Method	PSNR	SSIM	LPIPS	Total Time/s
Bicubic	28.17	0.85	0.32	0
DUS-GAN	27.86	0.83	0.28	4.11
MM-RealSR	21.76	0.77	0.13	4.63
LCR-GAN	27.35	0.82	0.21	0.32
Co-optimization	32.33	0.91	0.11	0

Table 2
Comparison of Unsupervised Network Parameters

Model	CinCGAN	USISResNet	DUS-GAN	MM-RealSR	LCR-GAN
Param	>43M	5.1M	15.9M	26.1M	3.7M

Note. M denotes a million.

the performance of the joint optimization method. The optical path for the validation experiment is illustrated in Figure 8. The light emitting diode emits scattered light with a central wavelength of 623 nm, and then filters out 623 nm monochromatic light with a narrow band filter. The intensity of the beam is adjusted by using the attenuator plate, and the beam forms many scattered point sources through the aperture of the resolution plate, simulating the light emitted by the star point. A single lens is placed at a focal length from the resolution plate to correct the light into parallel light, simulating an infinite distance. The focal length of the lens is 100 mm. The beam splitter prism splits the light into two beams, and one is

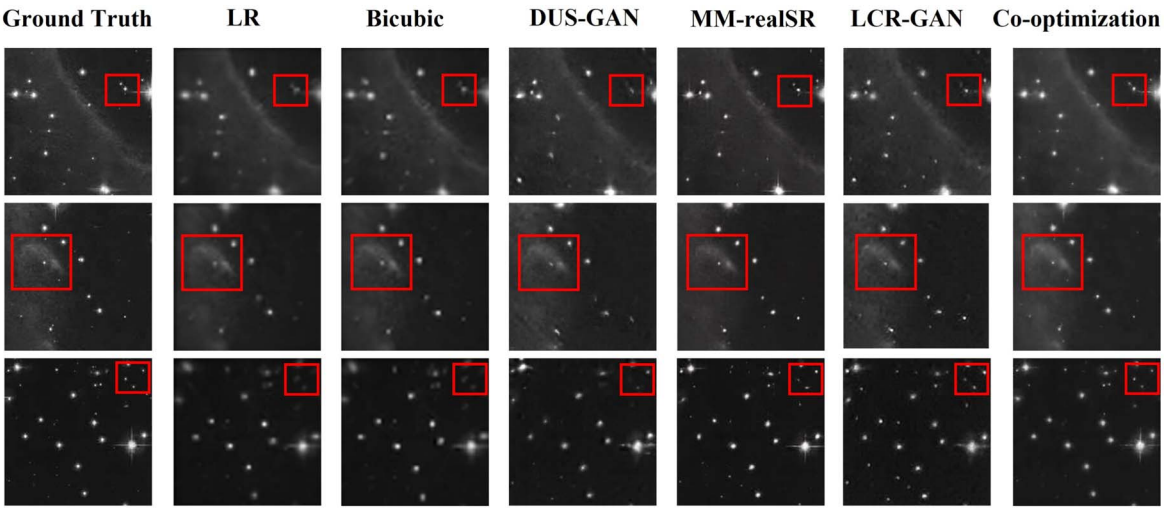

Figure 7. Comparison of SR images reconstructed by deep learning and the joint optimal method.

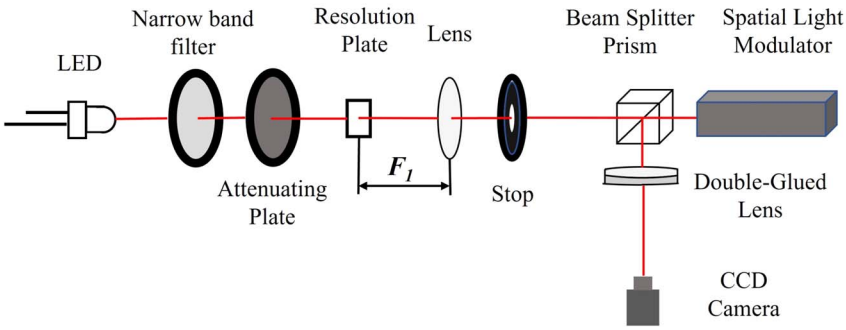

Figure 8. Optical path diagram of the resolution plate validation experiment.

transmitted to the surface of the spatial light modulator, and reflected back to the prism after phase modulation. Then the prism deflects the reflected light to the perpendicular direction of the incident light. The light is imaged by the DGTO in the back focal plane, which is recorded by the low-illumination camera. The focal length of the double-glued objective is 200 mm. The pixel size of the detector is 3.45 μm. The length of the spatial light modulator is about 13 mm, and width is about 8 mm.

In terms of waveform modulation, the telescopic system is equivalent to a lens with an arbitrary modulated wave front shape. In this paper, a non-ideal lens, a double-glued objective, is used instead of the telescopic system to verify the joint optimization method. As shown in Figure 8, since the light incident on the DGTO is a plane wave and the thickness of the double-glued lens is small, the spatial light modulator can be placed in front of the double-glued objective. Since the fringe width of the resolution plate is less than 30 μm, in order to better characterize the performance of the joint optimization method, the resolution plate is used instead of the star atlas in the experiment. In addition, as described in Section 4.1, we add two images of resolution plates to the original images and retrain the corresponding network, to help reconstruct the SR image of the resolution plate.

In order to verify the performance of the joint optimization method, we generate a modulation phase mask according to the method described in Section 2. According to the joint optimization method based on the ideal lens proposed by Sun Q., the corresponding phase mask is solved. The spatial light modulator modulates the phase of the light field according to the grayscale image of the phase mask. The image of the phase mask is fed into the spatial light modulator. In the off and on state of the spatial light modulator, the detector collects the images of the phase-free mask, the phase mask with the Sun Q. method, and the phase mask with the joint optimization method in this paper, as the inputs of their respective networks. The SR images corresponding to deep learning, the Sun's method, and the joint optimization method in this paper are all reconstructed.

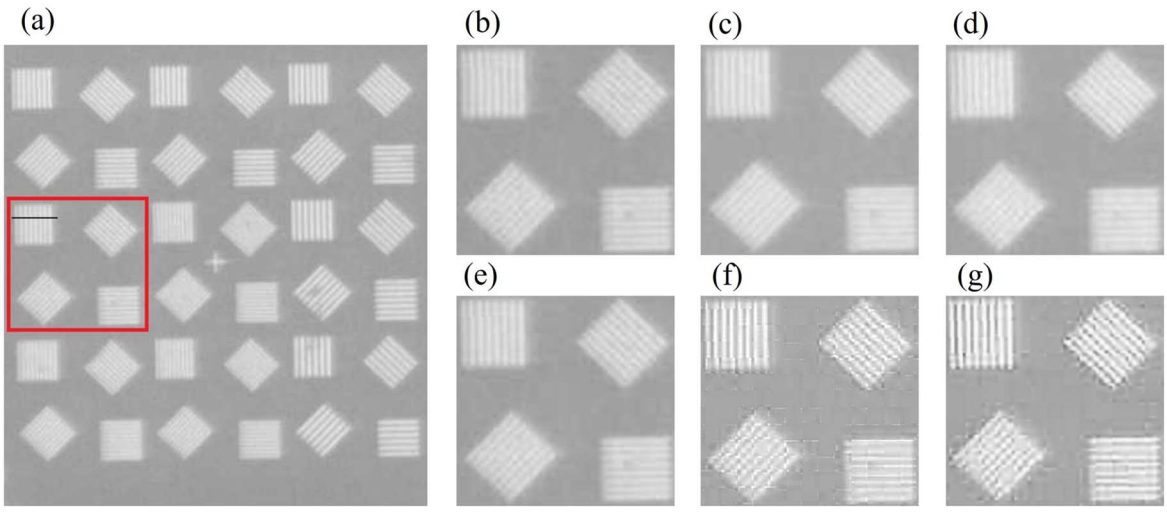

Figure 9. Experimental results of SR reconstruction of resolution plates.

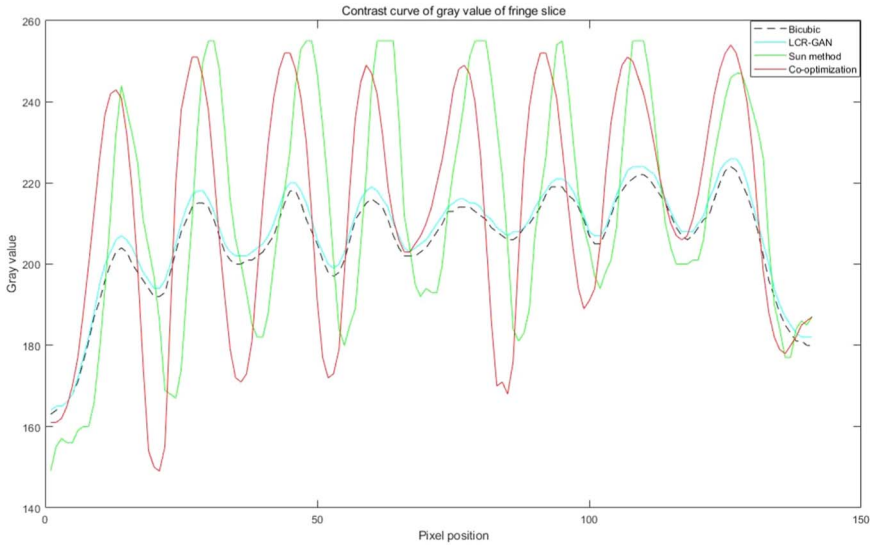

Figure 10. The gray value curves corresponding to the fringe slices of several reconstructed images.

5.2. Experimental Result Analysis

The experimental results are displayed in Figure 9. Panel (a) is the image of the standard resolution plate actually taken by the detector, and the red box corresponds to the reference position of the control group. Panels (b), (c), and (d) are the images without phase mask, phase mask with Sun method, and phase mask with joint optimization method, respectively. Panels (c), (f), and (g) are SR images reconstructed by LCR-GAN, Sun's method, and joint optimization method respectively. According to the Rayleigh criterion, the lateral diffraction limit of the system is about 11.69 μm, and the longitudinal diffraction limit is about 19 μm. The fringe width is about 8.3 pixels, or 28.64 μm, which is larger than the diffraction limit, and theoretically, the system can distinguish between light and dark fringes. Comparing Panels (b) and (e), due to the system aberration, noise, and other reasons, it is difficult to distinguish light and dark fringes in images captured by the detector, and LCR-GAN has a certain SR effect. Comparing Panels (d) and (g), it can be seen that the SR image reconstructed by the joint optimization method in this paper is rich in detail, and the fringes are clearly distinguishable, which greatly improves the resolution of the observation image.

Comparing Panels (f) and (g), it can be seen that the SR effect of the image reconstructed by Sun's method is lower than that of the joint optimization method. We explain that the modulation phase of the double-glued objective lens is different from that of the ideal lens, and the phase mask solved by the Sun method has an error, resulting in the loss of some high-frequency information in the image plane. These results verify the simulation results.

To quantitatively measure the SR effect, we take tangents to the stripes in the reconstructed image, and the position corresponds to the black line in Figure 9(a). The pixel position is taken as the *X*-axis, and the gray value at the tangent line is plotted. The legend in the upper right corner of Figure 10 shows the curves corresponding to different reconstruction methods. For comparison, we performed a 4-fold bicubic upsampling on the real image, to match the upsampling ratio of other SR reconstruction methods. Since the stripes are equally spaced, we measure the contrast of the fringe by calculating the average absolute error between the seven pairs of clear crests and troughs in Figure 10. The mean absolute errors of bicubic upsampling, LCR-GAN, Sun's method, and joint optimization method are 17, 17.7, 68.7 and 71.4, respectively. The joint optimization method in this paper has the highest contrast, and the stripes in the reconstructed image are easier to distinguish.

6. Conclusions

In this paper, an optical-neural network joint optimization method based on the telescopic system is proposed. First, the deep learning convolutional layer equivalent to the telescopic system PSF is integrated into the front end of the neural network, and the PSF and network parameters are jointly optimized through the training process. After the training, a phase mask is constructed to reconstruct the trained PSF in the telescopic system, the manufactured phase mask is added to the system and the observation image is collected, and the network is retrained to fine-tune the network parameters. This paper also constructs an unsupervised network LCR-GAN, which has the advantages of light parameters and fast reconstruction speed, and is suitable for real-time tasks of mobile computing devices. Simulation and experiments show that the reconstructed star atlas is rich in detail and easy to distinguish, and the resolution of the observed image can be improved without adding complex hardware equipment. It is worth noting that since the network in this paper is trained using a data set composed of known star atlases, it can only be applied to the observation images of known targets with basic outlines or shapes.

Acknowledgments

Funding is provided by the National Natural Science Foundation of China (NSFC, Grant Nos. 62375027 and 62127813), Natural Science Foundation of Chongqing Municipality (CSTB2023NSCQ-MSX0504), Natural Science Foundation of Jilin Provincial (YDZJ202201ZYTS411), and Jilin Provincial Education Department Fund of China (JJKH20240920KJ).

References

Chang, J., Sitzmann, Q., Dun, X., et al. 2018, NatSR, 8, 12324
Chang, J.-L., & Wetzstein, G. 2019, in IEEE/CVF Int. Conf. on Computer Vision (ICCV) (Piscataway, NJ: IEEE), 10192
Gerchberg, R. W., & Saxton, W. O. 1972, Optik, 35, 237
Goodman, J. W. 2005, Introduction to Fourier Optics (3rd ed.; Englewood, CO: Roberts & Co. Publishers)
He, Z.-Q., Yin, K., Wu, S.-T., et al. 2021, LSA, 134, 10
Irani, M., & Park, S. 1991, ITIP, 53, 231
Lehmann, T. M., Park, M. K., Kang, M. G., et al. 1999, IEEE Trans. Med. Imaging, 18, 1049
Ledig, C., Theis, L., Huszar, F., et al. 2017, in 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 105
Li, Z., Peng, Q.-Y., Bhanu, B., et al. 2018, Ap&SS, 363, 92
Mikhail, K., Ma, Y.-Q., Chen, Y.-B., et al. 2019, LSA, 118, 8
Mou, C., Wu, Y.-Z., Wang, X.-T., et al. 2022, in Computer Vision – ECCV 2022, ed. S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Cham: Springer), 723
Park, S. C., Park, M. K., Kang, M. G., et al. 2003, IEEE Signal Process Mag., 20, 21
Peng, Y.-F., Fu, Q., Amata, H., et al. 2015, OExpr, 23, 31393
Peng, Y.-F., Sun, Q.-L., Dun, X., et al. 2019, ACM Trans. Graph, 38, 219:1
Phillip, I., Zhu, J.-Y., Zhou, T.-H., et al. 2017, in 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 5967
Prajapati, K., Chudasama, V., Patel, H., et al. 2021, ITIP, 30, 8251
Rahman, S., Moon, Y. J., Park, E., et al. 2020, ApJL, 897, L32
Shoubaneh, H., Eric, H., Hooshang, N., et al. 2022, ApJ, 941, 141
Sweere, S. F., Valtchanov, I., Lieu, M., et al. 2022, MNRAS, 517, 4054
Sun, Q.-L., Zhang, J., Dun, X., et al. 2020, ACM Trans. Graph, 39, 9:1
Wang, C., Shi, H.-D., Liu, Z., et al. 2020, ApOpt, 59, 7883
Wang, C., Xing, S.-Y., Xu, M., et al. 2022, Senso, 22, 2717
Wang, C.-T., Tang, D.-L., Wang, Y., et al. 2015, NatSR, 5, 18485
Wang, K.-K., Fu, Q., Shi, H.-D., et al. 2023a, InPhT, 2023, 105032
Wang, K.-K., Fu, Q., Wang, B., et al. 2023b, OExpr, 31, 30750
Wang, S., Jiang, X.-N., Guo, H.-J., et al. 2023c, OptCo, 545, 129634
Xu, M., Wang, C., Wang, K.-K., et al. 2022, Senso, 22, 9676
Yan, Q., Xu, Y., Yang, X., et al. 2015, ITIP, 24, 3187
Yang, M.-Y., Fan, X.-W., Wang, Y.-M., et al. 2022, RemS, 14, 2264
Youla, D. C., & Webb, H. 1982, IEEE Trans. Med. Imaging, 1, 81
Yuan, R., Zhao, C.-Z., Guo, Y., et al. 2020, LOP, 57, 232201