The Astrophysical Journal, 923:76 (12pp), 2021 December 10 https://doi.org/10.3847/1538-4357/ac2aa2

Improving the Spatial Resolution of Solar Images Using Generative Adversarial Network and Self-attention Mechanism*

Junlan Deng[1], Wei Song[1,2,3], Dan Liu[1], Qin Li[4,5], Ganghua Lin[2,6], and Haimin Wang[4,5,7]
[1] School of Information Engineering, Minzu University of China, Beijing 100081, People's Republic of China; songwei@muc.edu.cn
[2] Key Laboratory of Solar Activity, Chinese Academy of Sciences (KLSA, CAS), Beijing 100081, People's Republic of China
[3] National Language Resource Monitoring and Research Center of Minority Languages, Minzu University of China, Beijing 100081, People's Republic of China
[4] Center for Solar-Terrestrial Research, New Jersey Institute of Technology, University Heights, Newark, NJ 07102-1982, USA
[5] Institute for Space Weather Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102-1982, USA
[6] National Astronomical Observatories, Chinese Academy of Sciences (NAOC, CAS), Beijing 100081, People's Republic of China
[7] Big Bear Solar Observatory, New Jersey Institute of Technology, 40386 North Shore Lane, Big Bear City, CA 92314-9672, USA
Received 2021 July 31; revised 2021 September 17; accepted 2021 September 20; published 2021 December 13

Abstract

In recent years, the new physics of the Sun has been revealed using advanced data with high spatial and temporal resolutions. The Helioseismic and Magnetic Imager (HMI) on board the Solar Dynamic Observatory has accumulated abundant observation data for the study of solar activity with sufficient cadence, but their spatial resolution (about 1″) is not enough to analyze the subarcsecond structure of the Sun. On the other hand, high-resolution observation from large-aperture ground-based telescopes, such as the 1.6 m Goode Solar Telescope (GST) at the Big Bear Solar Observatory, can achieve a much higher resolution on the order of 0.″1 (about 70 km). However, these high-resolution data only became available in the past 10 yr, with a limited time period during the day and with a very limited field of view. The Generative Adversarial Network (GAN) has greatly improved the perceptual quality of images in image translation tasks, and the self-attention mechanism can retrieve rich information from images. This paper uses HMI and GST images to construct a precisely aligned data set based on the scale-invariant feature transform algorithm and t0 reconstruct the HMI continuum images with four times better resolution. Neural networks based on the conditional GAN and self-attention mechanism are trained to restore the details of solar active regions and to predict the reconstruction error. The experimental results show that the reconstructed images are in good agreement with GST images, demonstrating the success of resolution improvement using machine learning.

Unified Astronomy Thesaurus concepts: Solar photosphere (1518); Neural networks (1933); Solar active regions (1974)

Supporting material: animation

1. Introduction

Solar activity has a significant impact on Earth. For example, coronal mass ejections generated by solar eruptions may cause geomagnetic storms and accelerate solar energetic particles. Those space weather events may affect radio communications, the operation of satellites, as well as power grids. Therefore, the study of solar activity is one of the most important subjects in solar physics. Modern observations clearly show that solar features have rich structures below 1″, for example, penumbral filaments and umbral dots. Spatial resolution below 0.″1 is needed to study these small-scale structures.

To achieve higher resolution, several meter-class large ground-based solar telescopes have been built. The Goode Solar Telescope (GST) of Big Bear Solar Observatory (BBSO) is one of the most capable solar telescopes. For the observation of the solar photosphere, the wave band used is TiO at a wavelength of 7057 Å. Diffraction-limited images are routinely obtained using a high-order adaptive optics (AO) system and speckle reconstruction. The nominal pixel size of GST/TiO images is 0.″034 (Table 1).

As discussed above, using AO and with increased telescope apertures, the spatial resolution of solar observations by ground-based telescopes has been greatly improved. However, due to the limited availability of daytime and local climate, the Sun cannot be observed continuously for an extended time. In addition, the high-resolution observation has a very limited field of view. In contrast, space solar observation has the advantage of observing round the clock, without interruptions caused by bad weather.

Since its launch in 2010, the Helioseismic and Magnetic Imager (HMI) on board the Solar Dynamic Observatory (SDO) has obtained observations of the entire Sun with a cadence between 45 s and 12 minutes, which is used to study the oscillation and magnetic fields in the solar photosphere (Fleck et al. 2011). It has an aperture of 14 cm, using a spectral line of Fe I 617.3 nm. The detector system is a 4096 by 4096 pixel CCD, achieving a pixel resolution of 0.″5 (Schou et al. 2012; Couvidat et al. 2016).

Many solar research works, such as morphological and dynamics studies of filaments, faculae, and sunspots, are based on full-disk observations provided by HMI. However, the resolution of an HMI image is limited by the diffraction effect of the telescope aperture, so it is difficult to analyze smaller-scale activity. In addition, the air currents and temperature and humidity variations in the optical path of telescope will degrade the images (Wachter et al. 2012). Various optical aberrations and stray light cause additional smearing in images (Yeo et al. 2014).

In recent years, deep learning has provided a new way to reconstruct solar images. The structures of granulation and sunspots observed by GST and HMI are similar. Therefore, although HMI images are blurred, due to the inherent defects of the recording instrument and the influence of turbulence, we

* Released on 2021 March 1st.

Table 1
Property Comparison of GST and HMI

Instrument Name (1)	Target line (Å) (2)	Aperture (m) (3)	Pixel size (arcsec) (4)	CCD (pix) (5)	FOV (arcsec) (6)	Cadence (seconds) (7)
GST	7057	1.6	0.034	1843 × 1843	62.7	15
HMI	6173	0.14	0.505	4096 × 4096	2068.5	45

Note. This table shows the property comparison between HMI and GST. Their spectral line is similar. The spatial resolution of GST image is about 16 times higher than HMI, but the field of view is much smaller than HMI.

can use precisely aligned GST images and HMI images as a training set, use deep neural network to learn the corresponding relationship between them, and then reconstruct HMI images with higher resolution. Such an improved resolution would have a high impact in research. For example, using the time-sequence high-resolution data covering entire solar active regions will allow researchers to derive flow fields, which play an important role in building energy for solar flare onset (Scherrer et al. 2012).

The rest of the paper is organized as follows: Section 2 introduces the traditional methods of solar image reconstruction and popular neural network models of single-image super-resolution (SISR). Section 3 introduces the data set registration process. Section 4 mainly describes the structure of the generator and discriminator, and introduces the theory of attention module in detail. Section 5 shows the test and validation results of super-resolution reconstruction and uses peak signal-to-noise ratio (PSNR), structural similarity (SSIM), rms, power spectrum curve, scatter plot, and gray histogram for quantitative analysis.

2. Background of Super-resolution Tools

2.1. Traditional SR Method

To eliminate the spatial resolution degradation caused by the point-spread function (PSF) and restore the details of the solar images, many different image reconstruction methods have been developed. The simplest super-resolution (SR) method is a shift-and-add method. The image is sampled to the target resolution, and the average is obtained by shifting on a common grid (Mboula et al. 2015). At present, the main traditional method of solar image reconstruction is speckle imaging and speckle phase diversity, which are quite effective in removing atmospheric turbulence. Speckle imaging was successfully applied to solar images starting in the 1970s (Von der Lühe 1993) and gradually developed into the Labeyrie method, Knox Thompson method, and speckle mask method. The three methods estimate the PSF by atmospheric imaging theory and linear theory and then deconvolute it from the average energy spectrum, average cross-spectrum, and triple correlation of a group of short-exposure images, respectively (Wöger et al. 2008), to obtain the Fourier amplitude and phase of the target image.

The blind-deconvolution-based speckle phase diversity algorithm finds the closest instantaneous PSF through an optimization algorithm and deconvolutes it in the optimization process. Compared with the speckle imaging method, the phase diversity method does not need the PSF function information, but the solution strongly depends on the prior assumption (Ramos et al. 2018). When the constraints are not complete, the reconstruction has difficulty converging, and even incorrect results will be reconstructed. Speckle imaging has good performance, but there are also some limitations: the PSF is rarely spatially invariant and is difficult to measure, and it takes dozens or even hundreds of frames to recover a frame of image (Puschmann & Sailer 2006).

Although these traditional methods are effective in removing atmospheric turbulence, the spatial resolution can only approach but cannot exceed the diffraction limit of the telescope. Miura et al. proposed a theoretical method to reconstruct solar images with two times the SR (Miura et al. 1999). K.G. Puschmann found that although Miura et al.'s method can enhance the contrast of the solar image, the improved spatial resolution does not exceed the diffraction limit, because the solar image is an extended target, not a point source (Puschmann & Kneer 2005). Extending the resolution beyond the diffraction limit means obtaining images with the same resolution as those observed by larger-aperture telescopes. With the development of deep learning, deep convolution neural networks can transform different aperture telescope images without any prior information, which provides a method for the SR reconstruction of solar images.

2.2. SR Method Based on Deep Learning

In deep learning, the SR problem is treated as a regression problem, that is, updating the parameters of the deep neural network to minimize the distance between the predicted value and the real value. The first SR convolution network, the Super-Resolution Convolution Neural Network (SRCNN), was proposed by Dong et al., and its performance is obviously better than all traditional SR algorithms (Dong et al. 2015). To further develop the advantages of deep neural networks in single-image SR, various networks have been proposed. Very Deep Super-Resolution used global residuals so that the network only needs to learn the different high-frequency details of a low-resolution (LR) image and a high-resolution (HR) image (Kim et al. 2016a). In the same year, the Deeply Recursive Convolutional Network used the same module to recurse 25 times and enhanced the representation of the network without adding network parameters (Kim et al. 2016b).

Lim et al. constructed Enhanced Deep Super-Resolution using 32 modified residual blocks with 256 channels and won the NSIR challenge "New Trends in Image Restoration and Enhancement" in 2017 (Lim et al. 2017). To pursue the visual quality of the image, a Generative Adversarial Network (GAN) is proposed. Its goal is to improve the visual perception quality of the image. The first breakthrough work in this respect is Super-Resolution GANs (Ledig et al. 2017). Then, a Cascading Residual Network uses a variant of the residual block with three convolution layers to cascade (Ahn et al. 2018). The concept of channel attention mechanism, which is mainly used in image classification, was introduced into the field of SR through Residual Channel Attention Networks (Zhang et al. 2018).

In the past, SISR methods were trained on synthetic LR images (such as bicubic downsampling of images). Recently, an increasing number of researchers have paid attention to the problem of SR reconstruction in real scenarios. Wang proposed the Text Super-Resolution Network based on a real scene text SR data set (Wang et al. 2020), and the experimental results were 13% higher than using a synthetic SR data set. Wei et al. proposed the method of component divide and conquer to deal with the real-world SR problem (Wei et al. 2020).

2.3. Self-attention Mechanism

The self-attention mechanism calculates the response at a position in a sequence by attending to all positions within the same sequence. In the problem of SR reconstruction, GANs can generate styles and textures that are difficult to distinguish as true or false, but the boundary of the overall structure of the object is prone to deviation. This is probably because the long-range dependency of general GANs is only realized by a multilayered convolution, and it is difficult for the network to carefully coordinate multiple layers to capture the long-range dependency. Therefore, the network can only use local regions with a fixed shape to generate details, but it is difficult to use the features of all positions to generate HR details.

Inspired by traditional nonlocal mean filtering algorithms, Wang et al. linked the self-attention mechanism in machine translation with nonlocal operations in computer vision to simulate the temporal and spatial dependence of video sequences (Wang et al. 2018b). Ding et al. proposed the Nonlocal Recurrent Network, which introduced nonlocal operations into recurrent neural networks for image restoration for the first time. Self-attention GANs introduced the self-attention mechanism into GANs to perform long-distance modeling (Zhang et al. 2019). The second-order channel attention module and nonlocally enhanced residual block were proposed in the Second-order Attention Network to better use image structure cues for SR tasks (Dai et al. 2019). Nonlocal operations are extended to multiple scales by Mei et al., and the performance was improved (Mei et al. 2020). The application of nonlocal attention mechanisms to solve various computer vision problems has gradually become a trend.

Our work is inspired by CJ Díaz Baso et al. (Baso & Ramos 2018). They used a solar photosphere image simulated by a magnetohydrodynamic model as the HR image, applied PSF convolution and downsampling on the HR image to generate the corresponding LR image, and used the full convolutional network on HMI to perform SR reconstruction. Peng et al. downsampled NVST/TiO images to obtain LR images and restored the images based on CycleGAN (Jia et al. 2019). The work of this paper is different in that it uses a real solar data set and generates higher-resolution solar images based on a GAN and a self-attention mechanism. Real solar images are more diverse in degraded forms and are noisy. Super-resolution reconstruction using real solar image data sets is more challenging but practical.

3. Data Set

As a powerful nonlinear function approximator, deep neural networks usually require a large amount of paired data for training. In the task of SR reconstruction of the solar image, this work targets the construction of an SR data set composed of lower-resolution HMI images and higher-resolution GST images.

First, select the GST image and the HMI image as the HR and LR image pairs, and then use the scale-invariant feature transform (SIFT) to achieve precise alignment of the GST and HMI data. If the LR image and the HR image are not accurately aligned, the network may learn the wrong pixel correspondence information, resulting in obvious artifacts.

SIFT is a classic image-matching method based on feature points. There are three main steps in using SIFT: feature point identification, feature point matching, and registration parameter determination. In the stage of feature point identification, the location of the rotation, translation, and scaling (RTS) invariant feature points is obtained by comparing adjacent layers of the image multiscale Gaussian pyramids. After low-contrast feature points and unstable feature points are removed by high-order function fitting, 128 dimensional feature vectors of the feature point are obtained through the image gradient.

Feature point matching establishes the matching relationship between the SIFT feature point descriptors in two images to be matched. The general nearest neighbor method is to minimize the Euclidean distance between the SIFT descriptors. When the ratio of the nearest distance to the second-nearest Euclidean distance is less than a certain threshold (set as 0.7 in this paper), the feature points are regarded to be matching. In the stage of registration parameter determination, the homogeneous coordinate transformation equations are solved by matching points to obtain the RTS value. In the process of solving, the sampling consensus algorithm (RANSAC) is used to eliminate the influence of the mismatch.

The experiment uses the OpenCV library to realize the identification and matching of feature points. It should be noted that the GST and HMI image resolutions are very different, and the GST data need to be downsampled and Gaussian-blurred to a resolution similar to that of the HMI before registration (Ji et al. 2019). For example, the HMI on 2015 January 5 obtained 788 feature points through the SIFT algorithm, and the preprocessed GST obtained 68 feature points. Through KNN nearest neighbor matching, 20 pairs of feature points are screened out and used to calculate the affine transformation matrix. Finally, the HMI image is rotated, translated, and zoomed to obtain an image that is precisely aligned with the GST image (Figure 1).

Based on the above method, 1597 pairs of accurately registered images from 2011 to 2020 were obtained. The spatial resolution of the original GST image was downsampled to $0\farcs126$, and 240×240 pixel patches cropped from the center of preprocessed images were used as the data set. One hundred pairs of image data that are more than one day apart from other training images are randomly selected as the test set. The amount and diversity of data sets are enough to represent the different evolution processes of solar activity and the different degradation processes in the light path.

4. Network Structure

4.1. Generator Structure

A Conditional Generative Adversarial Network (CGAN) that added the self-attention mechanism (Isola et al. 2017) is used to reconstruct the solar image (Figure 2). Input the HMI image to generate an HR solar image that is similar to a GST image. The basic structure of the generator follows U-Net, which is suitable for scenes with similar structures between input and output images. The unique structure allows the network to transmit

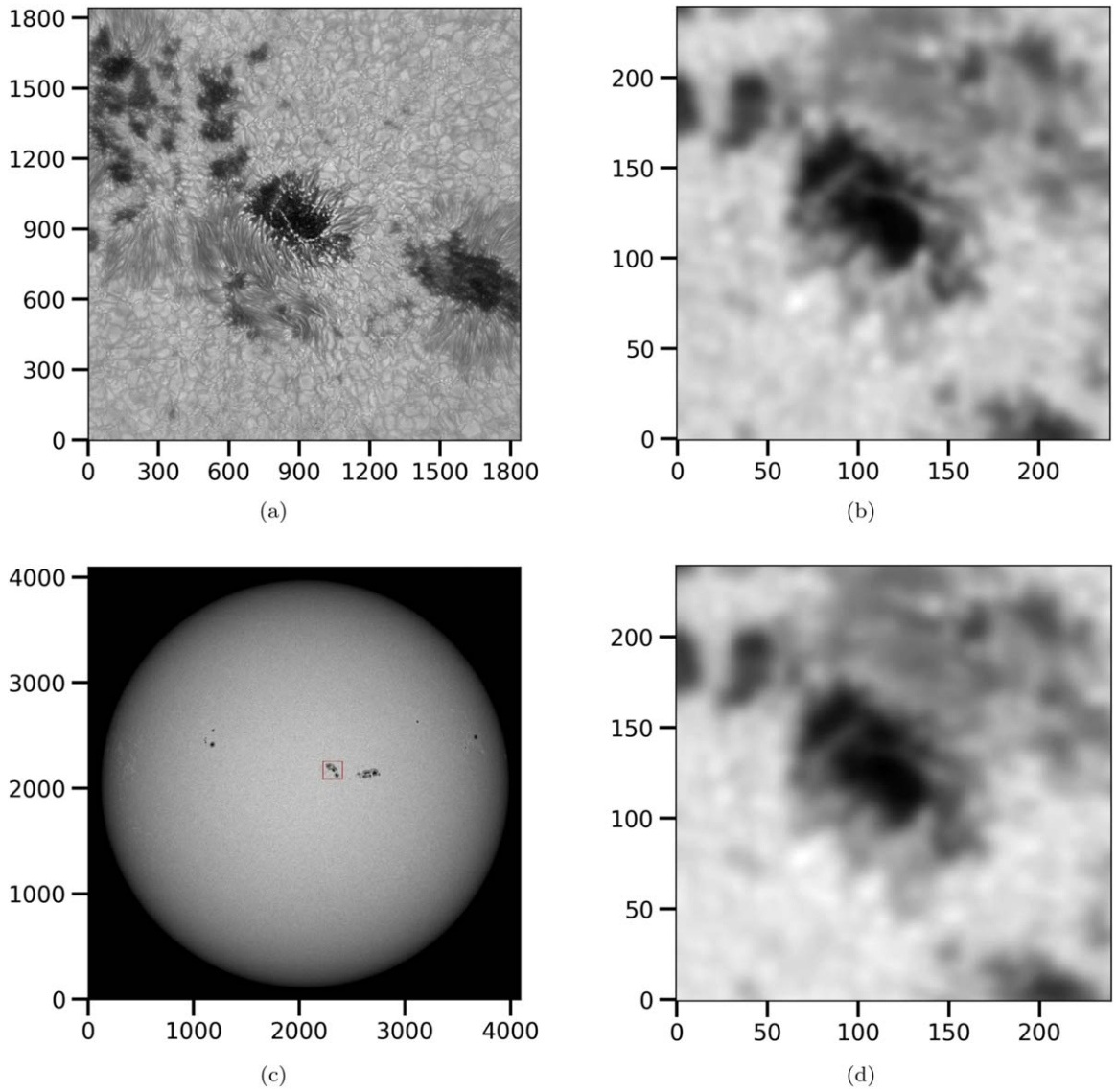

Figure 1. Image alignment results between GST and HMI data on 2015 January 5. (a) is the original GST image. (b) is the aligned GST image and blurred to HMI resolution. (c) is the original HMI image. The red box indicates the location of the GST image in the original HMI image. (d) is the HMI image obtained by the SIFT algorithm. The coordinates represent the range of pixels.

contextual information to higher layers for full integration. The shallow feature map usually contains low-frequency information of the image, such as the overall information of the image. The deep feature maps usually contain high-frequency information of the image, such as the edges and textures of the image. U-Net cascades high-level features and low-level features together through skip connection so that different levels of detailed information can be well preserved and integrated.

The downsampling part of the generator plays the role of feature extraction (Equation (1)):

$$F_{d4} = H_{d4}(H_{d3}(H_{d2}(H_{d1}(I_{\mathrm{LR}})))). \tag{1}$$

Among them, I_{LR} represents the input of the network and H_d represents the downsampling layer, consisting of a convolution kernel with a size of 4 and a stride of 2, a batch normalization, and a LeakyReLU activation function. After four downsampling layers, the feature map is represented as a latent variable F_{d4}, and the number of feature maps gradually doubles.

The upsampling part of the generator plays the role of feature fusion (Equation (2)):

$$F_{uk} = H_{uk}(S_k(F_{u(k-1)}) + F_{d(5-k)}). \tag{2}$$

Among them, F_{uk} represents the output result of the upsampling of the kth layer. H_u represents the upsampling module, which consists of a deconvolution kernel with a size of 4 and a stride of 2, batch normalization, and a ReLU activation function. In order to make the output match the input distribution of $[-1, 1]$, the upsampling module of the last layer changes the activation function to tanh. The upsampling result F_u in each layer combines two features: the context information extracted from the corresponding side of downsampling $F_{d(5-k)}$ and the self-attention map of the upper layer $S_k(F_{u(k-1)})$, where S represents the self-attention module.

The training and testing of the network take a 240×240 patch as input, the range of pixel value is standardized to $[-1, 1]$, the optimizer uses Adam, and the learning rate is set to $2 \times 10_{-4}$.

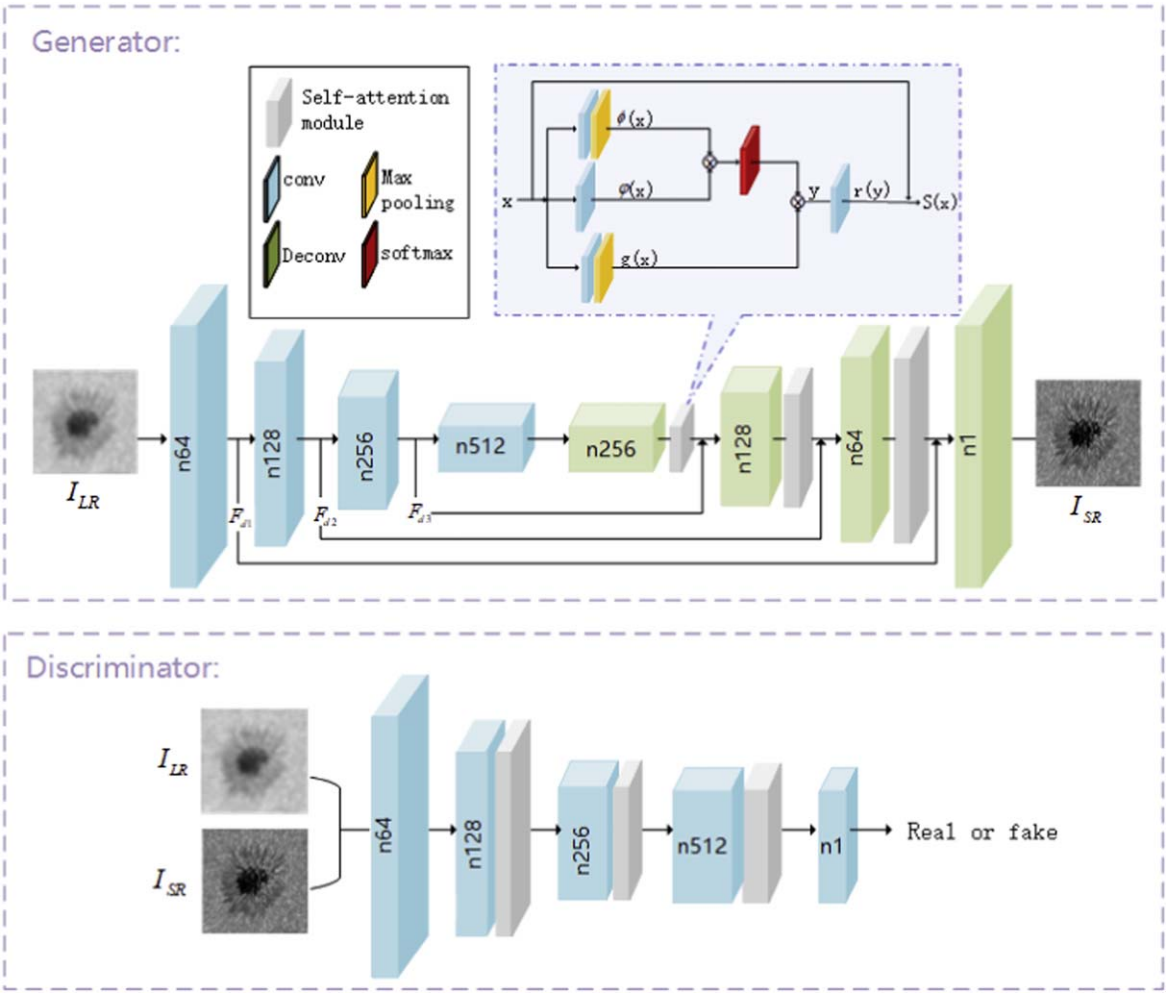

Figure 2. Generator network structure is divided into the downsampling part of feature extraction and the upsampling part of the feature fusion. N represents the number of channels, I_{LR} represents the HMI image upsampled four times in advance. The blue module represents the convolution layer, the green module represents the deconvolution layer, the gray module represents the self-attention module, the yellow module is Max pooling, and the red module is softmax.

4.2. Self-attention Mechanism

The self-attention mechanism refers to the weighted summation of the features of all nodes to obtain the response of a node in the network (Equation (3)):

$$S(x_i) = y_i = \sum_{j=1}^{N} \left(\frac{\exp\left(\phi(x_i)\right)^T \varphi(x_i)}{\sum_{j=1}^{N} \left(\exp\left(\phi(x_i)\right)^T \varphi(x_i)\right)} \right) g(x_j). \tag{3}$$

Among them, the functions $g(\cdot)$, $\phi(\cdot)$, $\varphi(\cdot)$ are unary functions, implemented by a 1^*1 convolution. Map the x_i and x_j features to different feature spaces through $\phi(\cdot)$ and $\varphi(\cdot)$, and then do the dot product operation to get the feature similarity. Standardization is realized through softmax operation, and the result indicates the degree of attention paid to location j when the network synthesizes location i. Multiply the attention map with $g(x_j)$ and get the final self-attention feature map y_i.

In order not to damage the original network structure when the attention map is superimposed on the network, the attention feature map y is multiplied by a learnable weight γ and initialized to 0. Use the skip connection to pass the upper deconvolution result x to the output (Equation (4)):

$$S(x) = \gamma(y) + x. \tag{4}$$

This makes the network first rely on the clues of local operations and then gradually learn to assign more weights to nonlocal attention, and finally make the feature map contain the global dependency of any two positions.

4.3. Discriminator Structure

The discriminator of the network uses PatchGAN with the self-attention mechanism. $G(x)$ is obtained through the HMI image and generator G. After being combined with the HMI image based on the channel dimension, $G(x)$ is put into the discriminator D to predict the probability value. In addition, the real image GST and HMI are also combined based on the channel dimension and input to the discriminator D to obtain the probability prediction value. In order to facilitate discriminator to combined HMI and GST based on channel, I_{LR} is the HMI image after four times bilinear interpolation.

PatchGAN finally outputs a feature matrix with a size of 28×28 (Table 2), which means that it evolves from predicting the true or false of the entire image to predicting the true or false of small areas. In addition, the discriminator adds a self-attention module, which can check whether similar features of the image at a distance are consistent and impose complicated geometric constraints on the global image structure.

Table 2
Structure of the Neural Network

Structure (1)	Layer (2)	Filter (3)	Stride (4)	Normalization (5)	Activation (6)	Output Size (7)
Generator	Input	⋯	⋯	⋯	⋯	240 × 240 × 1
	d1:Conv	4 × 4	2	⋯	Leaky ReLU	120 × 120 × 64
	d2:Conv	4 × 4	2	BN	Leaky ReLU	60 × 60 × 128
	d3:Conv	4 × 4	2	BN	Leaky ReLU	30 × 30 × 256
	d4:Conv	4 × 4	2	BN	ReLU	15 × 15 × 512
	u1:Deconv+attention	4 × 4	2	BN	ReLU	30 × 30 × 256
	u2:Deconv+attention	4 × 4	2	BN	ReLU	60 × 60 × 128
	u3:Deconv+attention	4 × 4	2	BN	ReLU	120 × 120 × 64
	u4:Deconv	4 × 4	2	⋯	Tanh	240 × 240 × 1
Discriminator	Input	⋯	⋯	⋯	⋯	240 × 240 × 2
	p1:Conv	4 × 4	2	⋯	Leaky ReLU	120 × 120 × 64
	p2:Conv +attention	4 × 4	2	BN	Leaky ReLU	60 × 60 × 128
	p3:Conv +attention	4 × 4	2	BN	Leaky ReLU	30 × 30 × 256
	p4:Conv +attention	4 × 4	1	BN	Leaky ReLU	29 × 29 × 512
	p5:Conv	4 × 4	1	⋯	Sigmoid	28 × 28 × 1

Note. The convolution kernel size and activation function used in each layer of the generator and discriminator are described in detail in the table.

The training effects of the network when using different loss functions are explored, such as L1 loss, L2 loss, CGAN loss, and Perceptual loss (Table 3). GAN loss is essential for GAN networks; CGAN loss and other losses are combined with a certain weight. L1 and L2 losses, both pixel losses, can accurately capture low-frequency information. Perceptual loss obtains the perceptual information of the image through a pretrained image classification network. The feature map of the seventh convolution of the VGG16 network is used as the perceptual information $V(x)$ of the image x. The definition of these losses are

$$L_{\mathrm{CGAN}} = E_{x,y}[\log D(x, y)] + E_x[\log(1 - D(x, G(x)))], \quad (5)$$

$$L_{l1} = E_{x,y}[|G(x) - y|], \quad (6)$$

$$L_{l2} = E_{x,y}[(G(x) - y)^2], \quad (7)$$

$$L_{\mathrm{perception}} = E_{x,y}[(V(G(x)) - V(y))^2]. \quad (8)$$

In the test results using L2 loss, some images are excessively smooth because Euclidean distance minimizes the loss by averaging all possible outputs, resulting in blur. In the test results using perceptual loss, the visual effect is similar to L1 loss, but L1 loss is better in terms of the index value.

5. Results Analysis

5.1. Evaluation Index

In order to evaluate the proposed method, we conducted extensive experiments on the data set. The model was trained on RTX 2080Ti for 4 hr, and then 100 images were tested and evaluated. Figure 3 shows some of these examples; the field of view is about $30'' \times 30''$. The first two rows mainly contain granulation in the quiet Sun and a small pore, and the last three rows are for active regions, focusing on sunspots. From the perspective of subjective evaluation, the reconstructed image can recover a lot of detailed information on the basis of keeping the contour edge of the HMI image. The boundary of the granulation structure in the quiet area is more obvious after the reconstruction, the contrast and brightness are consistent with the GST image, and the sharpness is greatly improved. In order to examine the reconstruction of small-scale structures, Figure 4 shows test result patches of pernumbral filaments, umbral dots, and granulation. Structures less than $1''$ can be observed, which achieves our initial goal.

Table 3
Test Results of Different Loss Functions

Function Type (1)	PSNR/dB (2)	SSIM (3)
L1 loss + CGAN loss	23.75	0.50
L2 loss + CGAN loss	22.32	0.45
perception loss + CGAN loss	22.77	0.45

Note. This table shows the average value of the PSRN and SSIM of the test results when using different loss functions.

To further verify the details of the recovery, we used entire 9-day images from 2017 September 1–9 for testing and made these images into an animation (Figure 5). The presented video shows much better the typical dynamics of a sunspot, such as materials moving away from the sunspot in the outer part of the penumbra and inward in the inner part.

The objective evaluation of the results uses three indicators: the PSNR between the predicted image and the target image, SSIM, and rms. PSNR (Equation (9)) is used to measure the ratio of the signal to noise in signal systems and is often used as a quantitative evaluation index for image compression and image SR reconstruction tasks. It is defined as

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{M^2}{\frac{1}{t}\sum_{i=1}^{t}(I_{\mathrm{LR}}(i) - I_{\mathrm{SR}}(i))^2}\right). \quad (9)$$

The PSNR value is determined by the maximum pixel value M of the image and the pixel mean square error MSE between I_{LR} and I_{HR}. t represents the total number of pixels in the image. For ordinary 8 bit deep images, the value of M is 255.

PSNR focuses on the difference of pixels and does not consider the structural information of the image. SSIM (Equation (10)) proposes to evaluate the SSIM of the image by comparing the contrast μ, brightness σ, and structural details

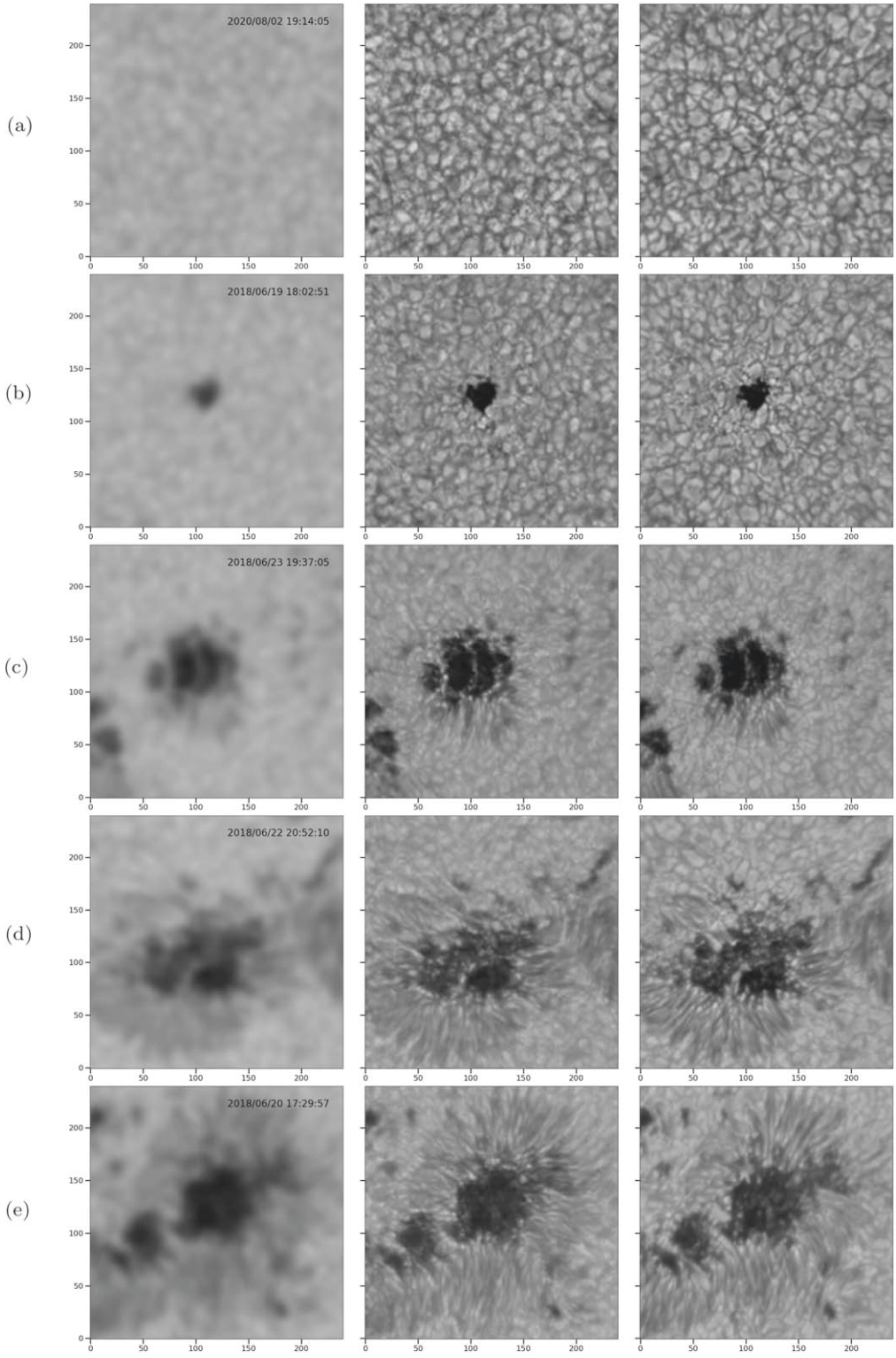

Figure 3. Testing results. The images from left to right correspond to the input HMI image, super-resolution reconstructed image, and target GST image. The field of view is about $30'' \times 30''$.

of the image. It is defined as

$$\mathrm{SSIM} = \frac{(2\mu_{\mathrm{LR}}\mu_{\mathrm{SR}} + c_1)(\sigma_{\mathrm{LRSR}} + c_2)}{(\mu_{\mathrm{LR}}^2 + \mu_{\mathrm{SR}}^2 + c_1)(\sigma_{\mathrm{LR}}^2 + \sigma_{\mathrm{SR}}^2 + c_2)}. \quad (10)$$

The brightness and contrast are obtained by calculating the average and standard deviation of image pixels. c1 and c2 are constant. μ_{LRSR} represents the covariance of I_{LR} and I_{HR}.

The rms (Equation (11)) does not need an HR image as a reference, and $\bar{I}$ is the average of the pixel values (Popowicz et al.

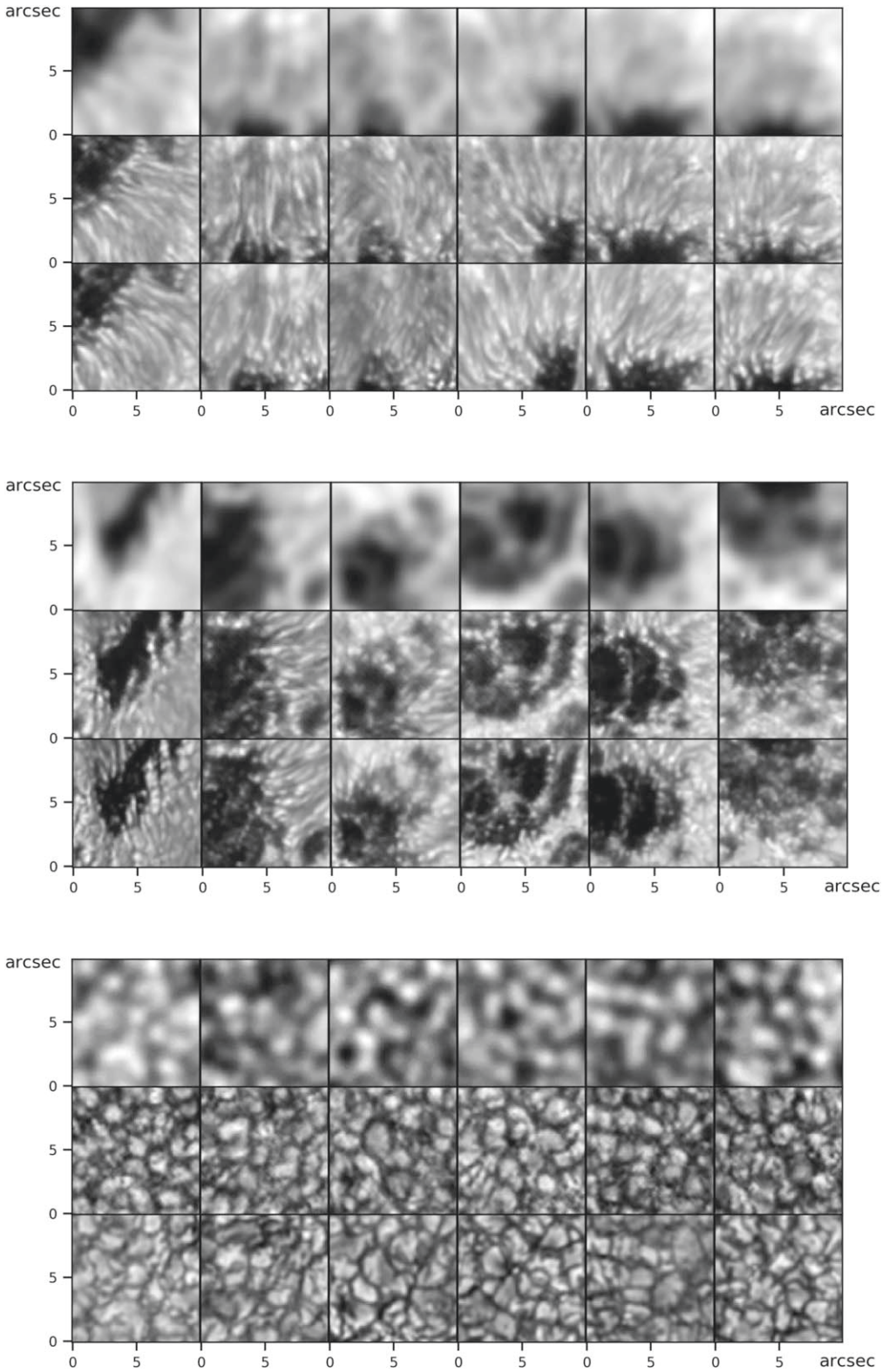

Figure 4. The test results of small-scale structures of the solar penumbra (top), umbra (middle), and granulation (bottom) are displayed respectively. The first row is the HMI image, the second row is the reconstructed image, and the third row is the GST image.

2017; Denker et al. 2018):

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(I-\bar{I})^2}\,/\bar{I}. \tag{11}$$

Although its sensitivity is related to the structure and content of the image, it is the most common method in solar image quality evaluation (Deng et al. 2015; Huang et al. 2019).

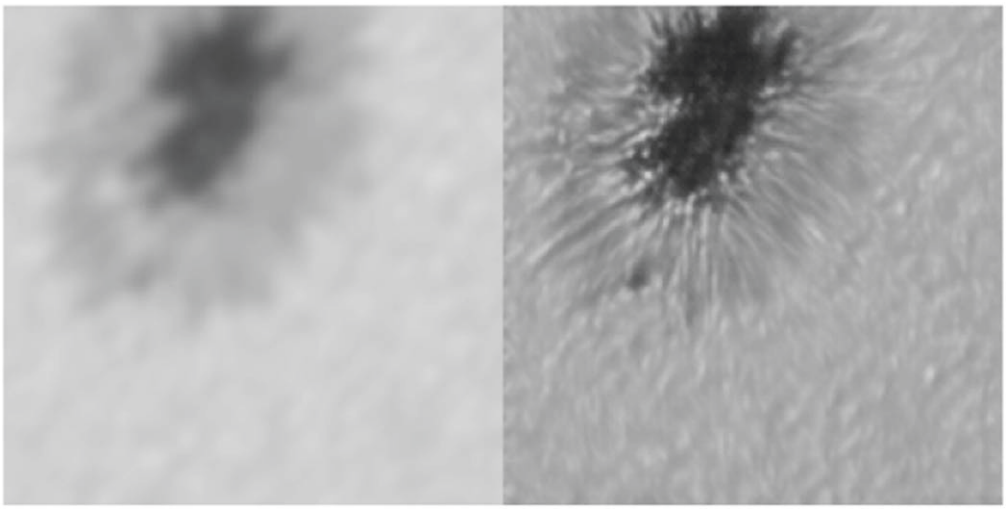

Figure 5. An animation of the super-resolution reconstructed results from 2017 September 1–9, comparing the input HMI images (left) to the SR images (right). The animation runs for 56 s, showing the much better typical dynamics of a sunspot from the reconstructed images. These improved dynamics include materials moving away from sunspots in the outer part of the penumbra and inward in the inner part.

(An animation of this figure is available.)

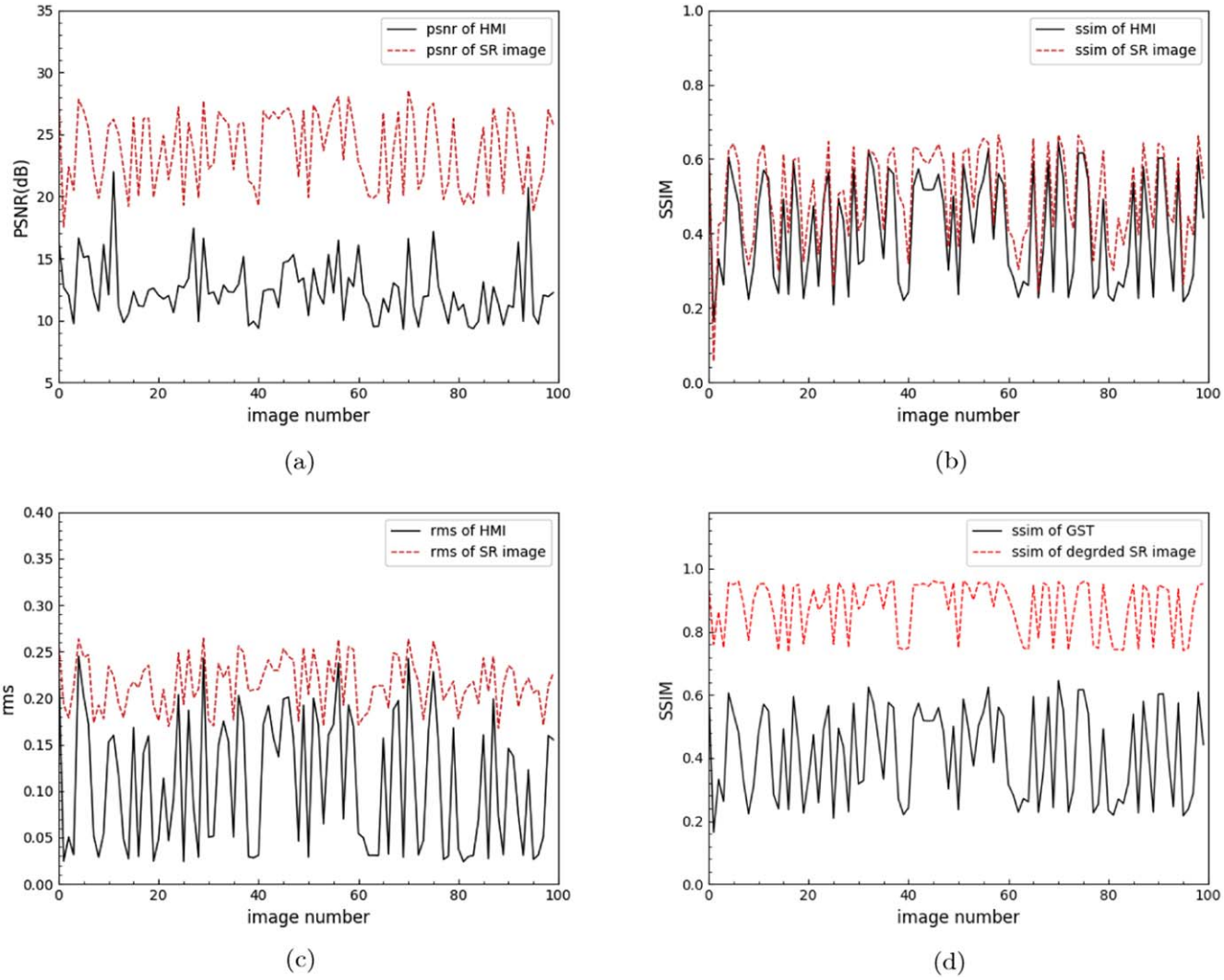

Figure 6. PSNR, SSIM, and rms index of test results. The black line in (a), (b), and (c) represents the index of the HMI image and GST image, and the red line represents the index of the SR image and GST image. (d) shows the SSIM index of the HMI image and GST image, as well as the index of the HMI image and the degraded SR image.

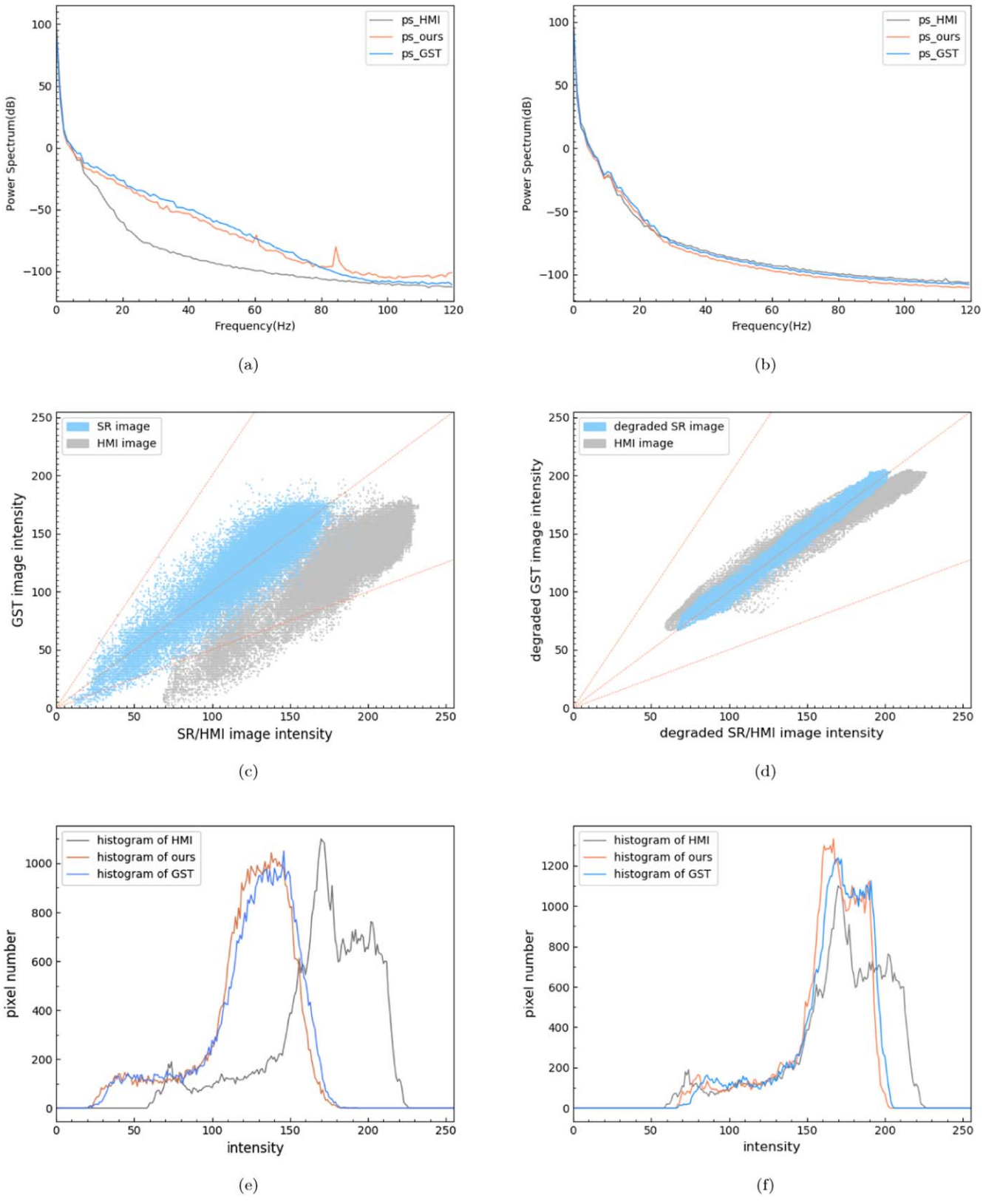

Figure 7. The three rows are a comparison of the power spectrum, scatter plot, and gray histogram between an SR image and GST image on 2020 June 20 at 17:29:57 UT, respectively. The dotted lines in the scatter plot represent the auxiliary lines with slopes of 0.5, 1, and 2. The left column compares the HMI image, reconstructed image, and GST image. The right column compares the HMI image, blurred reconstructed image, and blurred GST image.

5.2. Comparative Experiment

Figure 6 provides the PSNR, SSIM, and rms measurement values of all test images. The quantitative results show that compared with the original HMI image, the similarity between the output image and the GST image in terms of pixels and structure has been improved. In order to explore the effect of the self-attention mechanism, a comparative experiment was

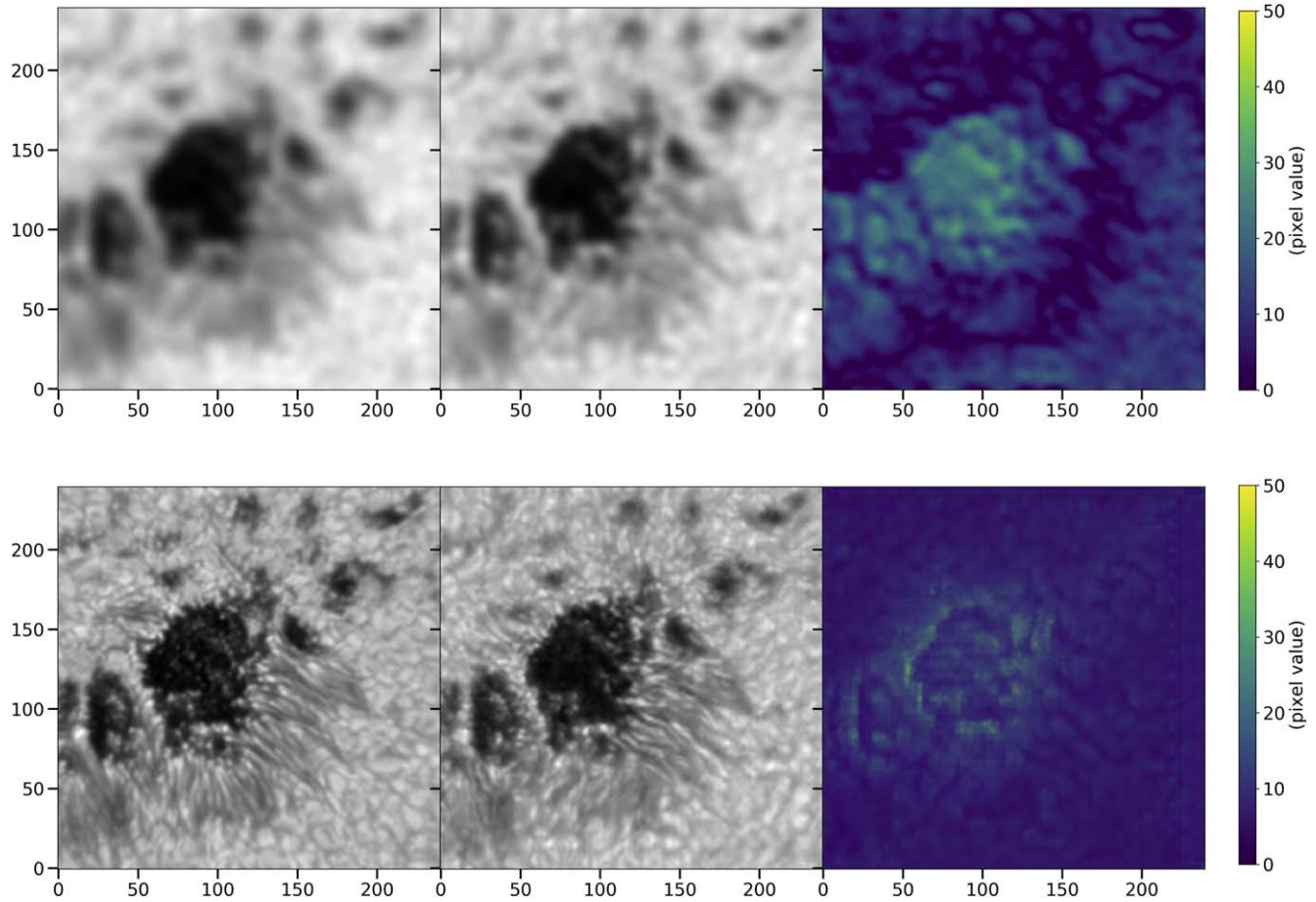

Figure 8. The first row, from left to right, shows the HMI image, the reconstructed image blurred to the HMI resolution, and the residual image, respectively. The second row, from left to right, shows the GST image, reconstructed image, and predicted residual image, respectively. The residual image shows the difference between the pixel values of the two images.

set up between the model in this paper and the basic model without attention. The test results showed that the average PSNR value of the improved CGAN network on the test set increased by 0.82, and the SSIM index increased by 0.03, with better performance.

The reason why the improved CGAN network has a high performance is that there are many similar texture structures in the HMI image, and the self-attention mechanism can directly search for high-frequency details from the LR image, so that the network has the ability to find similar patches from the image. At the same time, the fine details of each position in the image are coordinated with the details of other positions in the image. The network excavates as much as possible the inherent priors of the image, thereby obtaining rich structural clues in the image features and reconstructing more texture.

In order to compare the information content of each frequency of HMI image, SR reconstruction image, and GST image, the paper draws the azimuthally average power spectrum of the three images, i.e., first, square the amplitude of the two-dimensional Fourier transform of the image, and then calculate the average value on the ring of constant frequency after polar coordinate transformation (Wedemeyer-Böhm & van der Voort 2009). The power spectrum of Figure 3(e) is shown in Figure 7(a), and the resulting unit is converted to dB. It can be observed that the midfrequency of the HMI image is very low, indicating little detail in the image. After SR reconstruction, the mid-to-high frequency is improved to a certain extent, which is consistent with the GST data, and the high frequency is slightly higher than the GST and HMI, meaning that the basic structure of the image is enhanced and some details are restored. However, at the frequency of 85, the curve has fluctuations and deviations, which indicate the noise of the generated image.

The pixel value of the generated image reflects the intensity fluctuation in the solar photosphere. In order to compare the pixel values of the corresponding points between the generated image and the target image, Figures 7(c) and (e) show the scatter plot and histogram between the reconstructed image and GST image in Figure 3(e), respectively. The gray points indicate the correlation between the HMI data and GST data, and the cross-correlation value is 0.88. The blue points indicate the correlation between the super-resolution image intensity and GST image intensity, and the cross-correlation value has been improved to 0.92. The results show that most of the blue points are distributed near the line with a slope of 1, and the pixel values of the generated image and GST image show a linear relationship without significant error. The gray histogram also shows the consistency of pixel distribution between the reconstructed image and GST image.

In addition to the reconstructed image and the target image needing to be as similar as possible, the degraded SR image should also be consistent with the input image. To check the inverse process, the GST images and reconstructed images are blurred by a Gaussian kernel to the resolution of HMI. Figure 7(b)

shows that the power spectrum of the degraded image is almost the same as that of the HMI image. Figure 7(d) indicates that the inversed image and HMI image have a strong linear relationship. The reason for the difference in the histogram in Figure 7(f) is that the GST image is taken as the reconstruction target, and the contrast of the image is changed after reconstruction. The SSIM value in Figure 6(d) demonstrates that the degraded reconstructed image maintains a highly similar image structure to the input image.

For scientific applications, estimating the errors of a reconstruction image is as important as the reconstruction itself (Gitiaux et al. 2019). In order to estimate the reconstruction error and uncertainty, an existing network was trained with HMI images and the residual images, which are calculated from the reconstruction images and GST images. The test results show that the difference between the reconstruction results and GST images will be greater at the junction of the umbra and other regions. The predicted results are similar to the real difference between the blurred reconstructed results and the HMI images (Figure 8). A possible explanation is that although the small-scale structure observed by the Fe I spectral line and TiO spectral line is consistent, there are some differences in the intensity distribution of the observed values.

6. Conclusion

This paper proposes an SR reconstruction model of the solar images based on CGAN and the self-attention mechanism to improve the resolution of images by a factor of around 4 with an average cross-correlation of 91.6% in active regions. The network combines nonlocal operations and convolution operations to effectively capture the local and nonlocal features of the image.

One immediate usage of our data is to derive accurate flow fields by tracking moving features. Wang et al. (2018a) demonstrated that using HR data, the tracked velocity is much higher than that from lower-resolution data. Our next step is to apply the same technique to train vector magnetograms to improve the resolution of magnetograms of the HMI. Those improved magnetograms will disclose more clearly the flux emergence and cancellation in detail, which may be the primary triggering mechanism of solar eruptions (Toriumi & Wang 2019).

For the methodology aspect, future work will focus on creating a larger and more diversified data set from solar images, making the network have better generalization capabilities, and studying some different quantitative performance indicators to evaluate our method. In the future, even higher-resolution observations for the 4 m DKISK can be used for a similar kind of training to improve the resolution of GST.

We thank Dr. Vasyl Yurchyshyn for providing accurate image alignment information of GST TiO. BBSO operation is supported by NJIT, US NSF AGS-1821294, and NSFC 11729301 grants. The GST operation is partly supported by the Korea Astronomy and Space Science Institute, Seoul National University, and the Strategic Priority Research Program of CAS with grant No. XDB09000000. H.W. is supported by US NSF under grants AGS-1927578 and AGS-1954737. The work is supported by the open project of CAS Key Laboratory of Solar Activity (grant No: KLSA202114) and the cross-discipline research project of Minzu University of China (2020MDJC08).

ORCID iDs

Wei Song https://orcid.org/0000-0002-2324-4302
Haimin Wang https://orcid.org/0000-0002-5233-565X

References

Ahn, N., Kang, B., & Sohn, K.-A. 2018, in ECCV (Cham: Springer), 252
Baso, C. D., & Ramos, A. A. 2018, A&A, 614, A5
Couvidat, S., Schou, J., Hoeksema, J. T., et al. 2016, SoPh, 291, 1887
Dai, T., Cai, J., Zhang, Y., Xia, S.-T., & Zhang, L. 2019, in IEEE CVPR (Long Beach, CA: IEEE), 11065
Deng, H., Zhang, D., Wang, T., et al. 2015, SoPh, 290, 1479
Denker, C., Dineva, E., Balthasar, H., et al. 2018, SoPh, 293, 1
Dong, C., Loy, C. C., He, K., & Tang, X. 2015, PAMI, 38, 295
Fleck, B., Couvidat, S., & Straus, T. 2011, SoPh, 271, 27
Gitiaux, X., Maloney, S. A., Jungbluth, A., et al. 2019, arXiv:1911.01486
Huang, Y., Jia, P., Cai, D., & Cai, B. 2019, SoPh, 294, 1
Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. 2017, in IEEE CVPR (Honolulu, HI: IEEE), 1125
Ji, K., Liu, H., Jing, Z., Shang, Z., & Zhenping, Q. 2019, ChSBu, 16
Jia, P., Huang, Y., Cai, B., & Cai, D. 2019, ApJL, 881, L30
Kim, J., Lee, J. K., & Lee, K. M. 2016a, in IEEE CVPR (Las Vegas, NV: IEEE), 1646
Kim, J., Lee, J. K., & Lee, K. M. 2016b, in IEEE CVPR (Las Vegas, NV: IEEE), 1637
Ledig, C., Theis, L., Huszár, F., et al. 2017, in IEEE CVPR (Honolulu, HI: IEEE), 4681
Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. 2017, in IEEE CVPR (Honolulu, HI: IEEE), 136
Mboula, F. N., Starck, J.-L., Ronayette, S., Okumura, K., & Amiaux, J. 2015, A&A, 575, A86
Mei, Y., Fan, Y., Zhou, Y., et al. 2020, in IEEE CVPR (Seattle, WA: IEEE), 5690
Miura, N., Baba, N., Sakurai, T., et al. 1999, SoPh, 187, 347
Popowicz, A., Radlak, K., Bernacki, K., & Orlov, V. 2017, SoPh, 292, 1
Puschmann, K., & Sailer, M. 2006, A&A, 454, 1011
Puschmann, K. G., & Kneer, F. 2005, A&A, 436, 373
Ramos, A. A., de la Cruz Rodríguez, J., & Yabar, A. P. 2018, A&A, 620, A73
Scherrer, P. H., Schou, J., Bush, R., et al. 2012, SoPh, 275, 207
Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, SoPh, 275, 229
Toriumi, S., & Wang, H. 2019, LRSP, 16, 1
Von der Lühe, O. 1993, A&A, 268, 374
Wachter, R., Schou, J., Rabello-Soares, M., et al. 2012, SoPh, 275, 261
Wang, J., Liu, C., Deng, N., & Wang, H. 2018a, Apj, 853, 143
Wang, W., Xie, E., Liu, X., et al. 2020, ECCV (Berlin: Springer), 650
Wang, X., Girshick, R., Gupta, A., & He, K. 2018b, in IEEE CVPR (Salt Lake City, UT: IEEE), 7794
Wedemeyer-Böhm, S., & van der Voort, L. R. 2009, A&A, 503, 225
Wei, P., Xie, Z., Lu, H., et al. 2020, in ECCV (Berlin: Springer), 101
Wöger, F., Von der Lühe, O., & Reardon, K. 2008, A&A, 488, 375
Yeo, K., Feller, A., Solanki, S., et al. 2014, A&A, 561, A22
Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. 2019, ICML, PMLR, 7354
Zhang, Y., Li, K., Li, K., et al. 2018, in IEEE ECCV (Berlin: Springer), 286