

Solar Image Cloud Removal based on Improved Pix2Pix Network

Xukun Zhang¹, Wei Song^{1,2,3,*}, Ganghua Lin^{2,4} and Yuxi Shi⁵

¹School of Information Engineering, Minzu University of China, Beijing, 100081, China

²Key Laboratory of Solar Activity, Chinese Academy of Sciences (KLSA, CAS), Beijing, 100101, China

³National Language Resource Monitoring and Research Center of Minority Languages, Minzu University of China, Beijing, 100081, China

⁴National Astronomical Observatories, Chinese Academy of Sciences (NAOC, CAS), Beijing, 100101, China

⁵Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, New Jersey, 07102, USA

*Corresponding Author: Wei Song. Email: songwei@muc.edu.cn

Received: 12 January 2022; Accepted: 15 March 2022

Abstract: In ground-based observations of the Sun, solar images are often affected by appearance of thin clouds, which contaminate the images and affect the scientific results from data analysis. In this paper, the improved Pixel to Pixel Network (Pix2Pix) network is used to convert polluted images to clear images to remove the cloud shadow in the solar images. By adding attention module to the model, the hidden layer of Pix2Pix model can infer the attention map of the input feature vector according to the input feature vector. And then, the attention map is multiplied by the input feature map to give different weights to the hidden features in the feature map, adaptively refine the input feature map to make the model pay attention to important feature information and achieve better recovery effect. In order to further enhance the model's ability to recover detailed features, perceptual loss is added to the loss function. The model was tested on the full disk H-alpha images datasets provided by Huairou Solar Observing Station, National Astronomical Observatories. The experimental results show that the model can effectively remove the influence of thin clouds on the picture and restore the details of solar activity. The peak signal-to-noise ratio (PSNR) reaches 27.3012 and the learned perceptual image patch similarity (LPIPS) reaches 0.330, which is superior to the existed dehaze algorithms.

Keywords: Pix2Pix; decontamination; full-disk H-alpha images

1 Introduction

Study of solar activities depends on analyzing of observed time sequence solar images. Due to the influence of weather, solar images are often polluted, resulting in the loss of solar activity characteristics. At the Big Bear Solar Observatory (BBSO) in California, according to the site survey of the Global Oscillation Network Group (GONG) [1], the observation days affected by clouds account for 55% of the total observation days. At other observatories around the world, the percentage may



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

be higher. If the clouds block the sun, the shadow of the cloud will reduce the quality of observed image. Removing cloud pollution is of great significance for the study of solar activity. Huairou Solar Observatory (HSOS) is one of the ground based solar observatories which provides full disk observations in H-alpha, revealing the fine structure of solar chromosphere. In this paper, the cloud-contaminated full disk images and normal full-disk images provided by HSOS are used to remove cloud pollution.

Zhang et al., Feng et al. [2,3] proposed a method to remove pollution from the full-disk H-alpha images using a low-pass filter. This method obtains cloud shadow image from cloud image by special low-pass filter and corresponding clean image. The image shows the shape of the shadow created by the clouds on the original image and the extent to which the shadow blocks the original image. Cloud shadow images also represent cloud transmittance. Then the cloud image is corrected by the image to obtain a cloudless image. This method has the following limitations: If serious cloud pollution is eliminated by increasing the cut-off frequency of the filter, most solar features will also be removed. The image has a large intensity gradient at the edge of the sun, resulting in some edge effects during filtering. In this paper, the improved Pix2Pix network is used to convert cloud polluted images to clean images, avoiding the loss of solar activity details caused by the above method using filters, and achieving de-pollution effect.

The main technical components of this paper are as follows:

The loss function of Pix2Pix model adds perceptual loss, which improves the reconstruction ability of the model for deep image features and spatial structure.

The Pix2Pix model is improved by using the attention module. It makes the model paying attention to the polluted areas, and directly output the clean full-disk H-alpha images to restore the detailed characteristics of solar activity.

From the perspective of subjective visual perception, LPIPS index is introduced to evaluate the experimental results. The index uses the pre-training model to evaluate the depth characteristics of the image. From the experimental results, the improved Pix2Pix model can more clearly restore the solar activity characteristics affected by clouds in the full disk H-alpha images than the existed dehaze algorithms.

2 Related Previous Work

2.1 Dehaze

In the study of solar cloud shadow, this paper draws lessons from a method of removing haze. Methods for removing haze can be roughly divided into two categories: traditional methods based on prior knowledge and modern methods based on deep learning. Most of the methods based on prior knowledge are based on the atmospheric scattering model. For example, He et al. [4] proposed a method to estimate image prior information through an image dark channel. Fattal [5] proposed a new method to estimate the transport function from a single input image, adding surface shading variables to the atmospheric transport model. By assuming that the surface shadow and the transfer function are statistically independent, the author analyzes the atmospheric scattering model to obtain the transfer function and dehazes the image. These methods can only approximate the global atmospheric light and transmittance, and have great defects.

With the development of artificial neural network models, they have been increasingly applied in both thin cloud removal and mist removal. At the same time, the neural network model combined with the method of calculating prior information has put forward more cloud removal models. For example,

Sun et al. [6] proposed a cloud removal method combining image repair and image denoising, called Cloud Aware Generation Network (CAGN). Multispectral remote sensing images are often polluted by haze, resulting in poor image quality. Qin et al. [7] proposed a new method for haze removal of multispectral remote sensing images based on a residual structure depth convolution neural network (CNN). Firstly, multiple CNN individuals with residual structures are connected in parallel, and each individual is used to learn the regression from the blurred image to the clear image. Then, the output of the CNN individual is fused with the weight map to obtain the final dehaze result. Enomoto et al. [8] proposed the cloud removal network based on the color information of visible images and the high penetrability of longwave images, which can not only predict the fuzzy area from RGB images, but also capture the predicted fuzzy areas from the longer wavelengths that partially or completely penetrate the clouds. According to the atmospheric scattering model, the All-in-One dehazing network (AOD-Net) [9] proposed using a lightweight CNN to estimate the parameters of the atmospheric scattering model and directly generate haze-free images. Qin et al. [10] proposed an end-to-end feature fusion attention network (FFA-NET) to directly recover haze-free images. Zhang et al. [11] proposed a transmission image deep convolutional regression network (FT-DCRN) dehazing algorithm that uses fine transmission image and atmospheric light value to compute dehazed image.

2.2 Generative Adversarial Network

Generative adversarial network (GAN) [12] is a generation model proposed in 2014. The model structure includes a generator and a discriminator. The generator generates data with the same distribution as the training set. The discriminator determines whether the input data distribution is a real data distribution. The two play a dynamic game to generate data similar to the distribution of training data. The proposal of Wasserstein generative adversarial network (WGAN) [13], conditional generative adversarial network (CGAN) [14] and deep convolutional generative adversarial network (DCGAN) [15] has gradually improved the training instability of the GAN model, and the GAN model has been used in an increasing number of image processing tasks, including image recognition [16], multi-class detection [17], style transfer [18] and super resolution [19].

2.3 Attention Model

The attention mechanism is crucial in the human perceptual system. An important feature of the human visual system is that people do not try to process the whole scene at once. To better capture the structure of an image, the human visual system selectively focuses on a part of the image. To make the model with the human attention mechanism, many people have carried out this aspect of the attempt [20]. By modeling the correlation between feature channels, the important features are given more weight in the subsequent calculation to improve accuracy. Self-attention [21] determines the attention weight by focusing on all positions in the same sequence and calculating the correlation coefficient between one position in the sequence and other positions. Vaswani et al. [22] proved that machine translation models can achieve the most advanced results by using self-attention models. Convolutional block attention module (CBAM) [23] infers the attention map along two independent dimensions (channel and spatial) and then multiplies the attention map by the input feature map for adaptive feature refinement.

3 Method

It is difficult to estimate the global atmospheric illumination and transmittance of the full-disk H-alpha images by appropriate methods. In this paper, the improved Pix2Pix model is used to convert

the elimination of cloud shadow of the full-disk H-alpha images into the conversion between clean image and cloud pollution image, so as to achieve the effect of removing cloud pollution.

3.1 Improved Decontamination Pix2Pix Model

The Pix2Pix model is a classical image conversion model based on the GAN model. The model structure is shown in Fig. 1. It mainly consists of generators and discriminators. The generator uses the U-Net encoder-decoder structures. Encoder downsamples images through a series of convolution operations and then restores the downsampled result image to the target image using a decoder structure. At the same time, the encoder upsampling layer will combine its generated feature images with those generated by the decoder convolution layer via jump connections to improve the quality of the generated images. This model combines the CBAM module with the generator of Pix2Pix model to improve the detail features of the generated haze free image. By inputting the feature map input by the hidden layer into the channel attention module and the spatial attention module respectively, the channel attention map and the spatial attention map are obtained. Then, the feature map output by the hidden layer is multiplied with these two parts of attention map to obtain the refined feature map, which is fed to the next hidden layer.

The discriminator of the Pix2Pix model uses the PatchGAN structure [24], which does not discriminate the overall picture but discriminates the picture by discriminating each $n \times n$ patch of the feature map of generated image and integrating the discrimination results of each image block. Because smaller image blocks are used, the model adopts smaller parameters, and the speed of the model is faster [25].

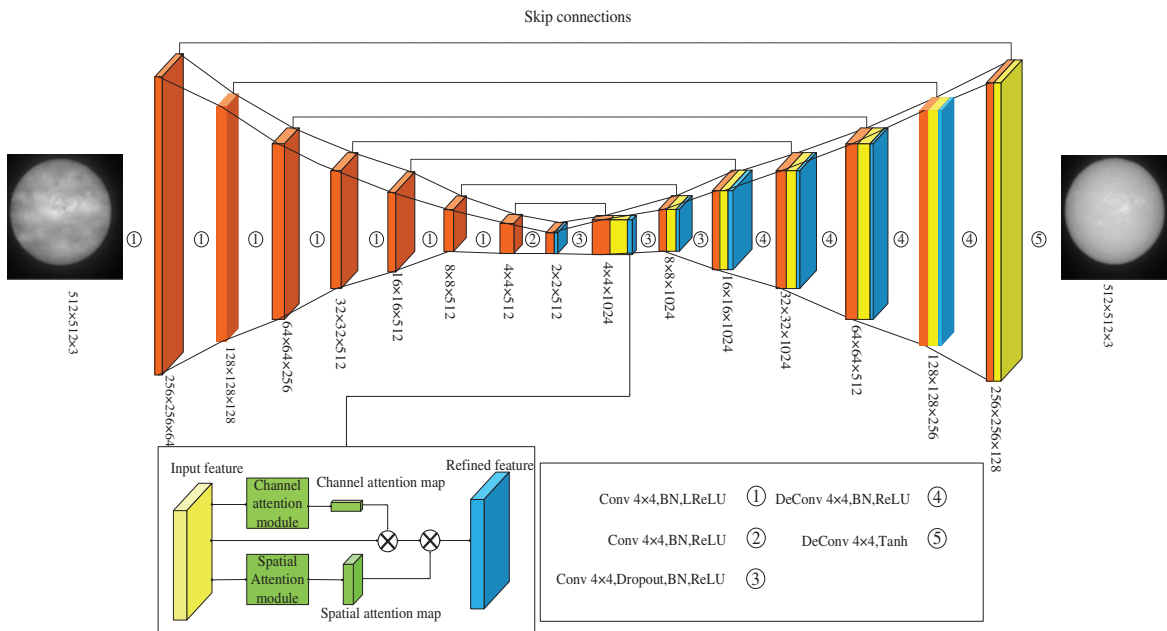


Figure 1: (Continued)

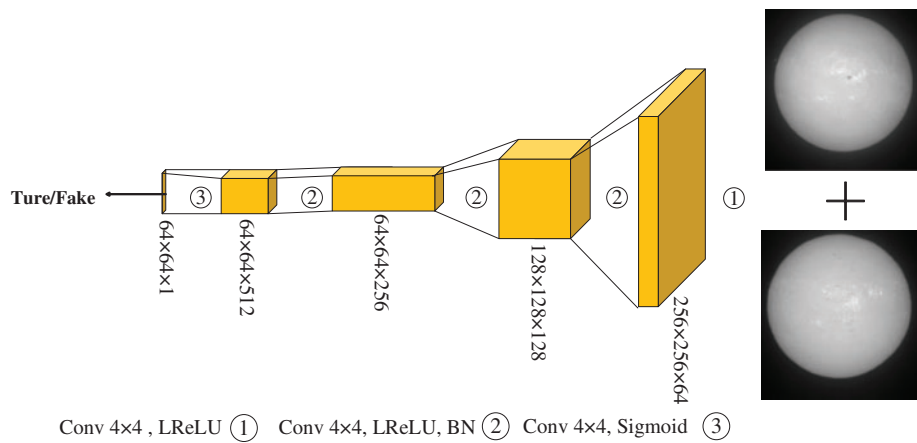


Figure 1: Improve Pix2Pix model structure

3.2 CBAM Model

CBAM [23] is a simple and effective attention module of a feedforward convolutional neural network. Given an intermediate feature map, the module computes the attention map sequentially along two independent dimensions (channel and space) and then multiplies the attention map by the input feature map for adaptive feature refinement. The model structure is shown Fig. 2.

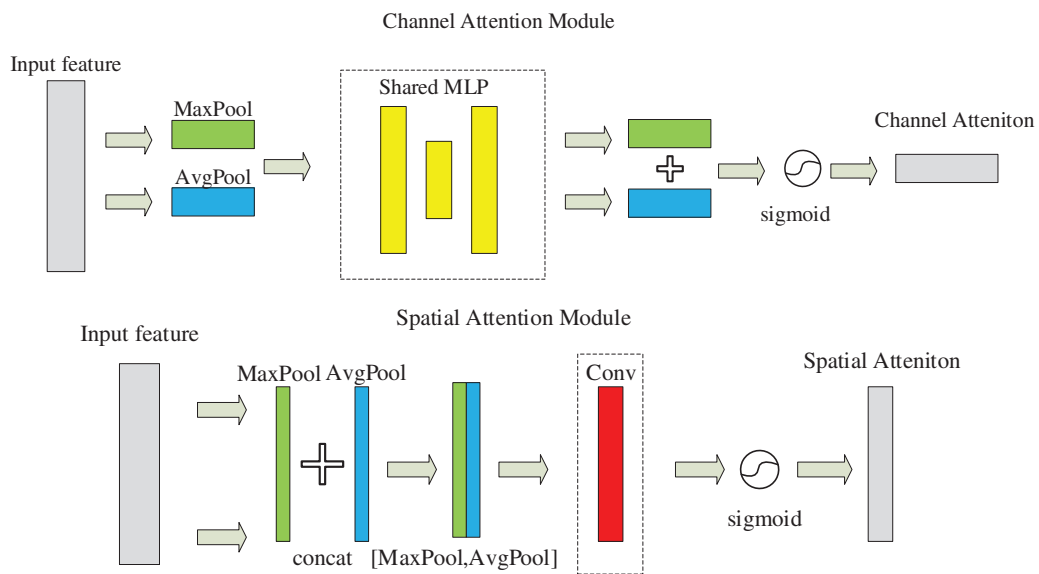


Figure 2: CBAM model structure

The CBAM module consists of channel attention and spatial attention. For channel attention, the average pooling and max pooling operations are firstly used to aggregate the spatial information of the feature matrix and generate two different attention scoring matrices: F_{avg} and F_{max} . Then, the two feature scoring matrices generate channel attention maps through a two-layer perceptron network. The perceptron network consists of multi-layer perceptron (MLP) and a hidden layer. After applying

the perceptron network to each attention score matrix, the two scores are added to obtain the channel attention weight. The calculation of channel attention is as follows:

$$\begin{aligned} A_c(\mathbf{F}) &= \sigma(\text{MLP}(\text{AvgPool}(\mathbf{F}))) + \text{MLP}(\text{MaxPool}(\mathbf{F})) \\ &= \sigma(\mathbf{M}_1(\mathbf{M}_0(\mathbf{F}_{avg}^c)) + \mathbf{M}_1(\mathbf{M}_0(\mathbf{F}_{max}^c))) \end{aligned} \quad (1)$$

A_c indicates channel attention map. Where \mathbf{F} represents the feature graph output by the hidden layer, σ represents the sigmoid function. \mathbf{M}_0 and \mathbf{M}_1 are multi-layer perceptron that share parameters. The \mathbf{M}_0 and \mathbf{M}_1 layer are followed by the ReLU activation function.

For spatial attention, average pooling and maximum pooling are also used to aggregate channel information of feature mapping, and two 2D maps are generated: $\mathbf{F}_{avg} \in \mathbf{R}^{1 \times H \times W}$ and $\mathbf{F}_{max} \in \mathbf{R}^{1 \times H \times W}$. Each vector represents the average and maximum pool characteristics in the channel. Then, they are connected and convolved through a convolution layer to generate a 2D spatial attention diagram. The calculation of spatial attention is as follows:

$$\begin{aligned} A_s(\mathbf{F}) &= \sigma(\text{Conv}[\text{AvgPool}(\mathbf{F}) \oplus \text{MaxPool}(\mathbf{F})]) \\ &= \sigma(\text{Conv}([\mathbf{F}_{avg} \oplus \mathbf{F}_{max}])) \end{aligned} \quad (2)$$

A_s indicates Spatial attention map. Where \mathbf{F} represents the feature graph output by the hidden layer, σ represents the sigmoid function, \oplus represents the concat operation and Conv represents the convolution operation with a convolution kernel size of 7×7 .

3.3 Loss

3.3.1 GAN Loss

Pix2Pix model is a variant of CGAN. The original loss function consists of L1 loss and GAN loss. The loss function of CGAN is shown as follows:

$$L_{CGAN}(G, D) = E_{X,Y}[\log D(X, Y)] + E_{X,Z}[\log(1 - D(X, G(X, Z)))] \quad (3)$$

where G represents the generator, D represents the discriminator, $D(X, Y)$ represents the discriminant result of the discriminator when the real picture is input, and $D(X, G(X, Z))$ represents the result when the input is the picture generated by the generator. While generator G tries to minimize the target function, discriminator D tries to maximize the target:

$$G^* = \arg \min_G \max_D L_{CGAN}(G, D) \quad (4)$$

3.3.2 L1 Loss

Mixing GAN losses with more traditional losses (such as L2 distance) is beneficial to model training [26]. The generator's job is not only to mislead the discriminator but also to be closer to the real image distribution. Therefore, L1 loss is selected to be added to Pix2Pix model. Compared with L2 loss, L1 loss reduces the blur effect of images [24].

$$L_{L1}(G) = E_{X,Y,Z}[\|Y - G(X, Z)\|_1] \quad (5)$$

In the formula, Y is the real original image, and $G(X, Z)$ is the target picture generated by the generator corresponding to the conditional picture X .

3.3.3 Perceptual Loss

Perceptual loss (PLloss) [27] was first used in the literature of image style migration. It mainly uses a specific network to extract the characteristic image output by the hidden layer for loss calculation.

By adding this loss function, the model can learn the image detail information and the overall spatial structure, and restore the solar activity characteristics of the polluted image. $\varphi_i(\mathbf{Y})$ is a pre-training network φ output when processing image \mathbf{Y} at layer i . $\varphi_i(\mathbf{Y})$ is a characteristic graph with the shape $c \times h \times w$. The model uses the mean square error to calculate the perceptual loss between the generated image and the target image:

$$L_P^{\varphi,i}(\hat{\mathbf{Y}}, \mathbf{Y}) = \mathbb{E} \left[\left(\varphi_i(\hat{\mathbf{Y}}) - \varphi_i(\mathbf{Y}) \right)^2 \right] \quad (6)$$

In the formula, \mathbf{Y} is the real original image, $\hat{\mathbf{Y}}$ is the target image generated by the generator.

The function L of the model consists of three parts:

$$L = \alpha_1 L_{CGAN} + \alpha_2 L_{L1} + \alpha_3 L_P \quad (7)$$

In the formula, α_1 , α_2 and α_3 are hyperparameters.

4 Experiments

The method is performed on the TensorFlow framework and an NVIDIA GeForce RTX2080ti GPU. During training, the Adam optimizer [28] is adopted with a batch size of 1, and set a learning rate of 0.0002 is set. The discriminator tended to converge after 200 epochs on the training set.

4.1 Datasets

The dataset consists of two parts: the full-disk H-alpha images obscured by clouds and the full-disk H-alpha images without clouds. 82 pairs of cloud-contaminated and clear H-alpha images were selected as the original training set, and 15 pairs of H-alpha images were selected to test the model. Data enhancement is also critical to the variance and robustness of the network. Flip, cut, scale, and rotate were used to generate 2566 pairs of samples from 82 high-quality H-alpha global solar images.

4.2 Ablation Experiment

The CBAM module is combined with different structures of the original Pix2Pix model for experiments. The experimental results are shown in Figs. 3 and 4.

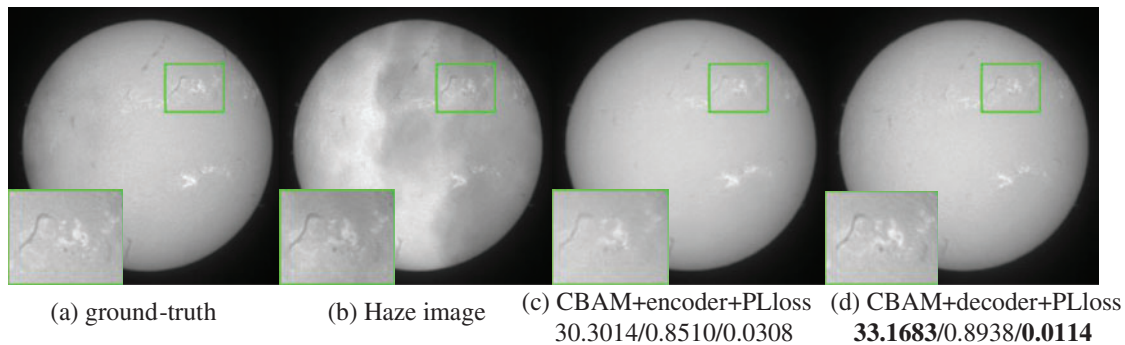


Figure 3: (Continued)

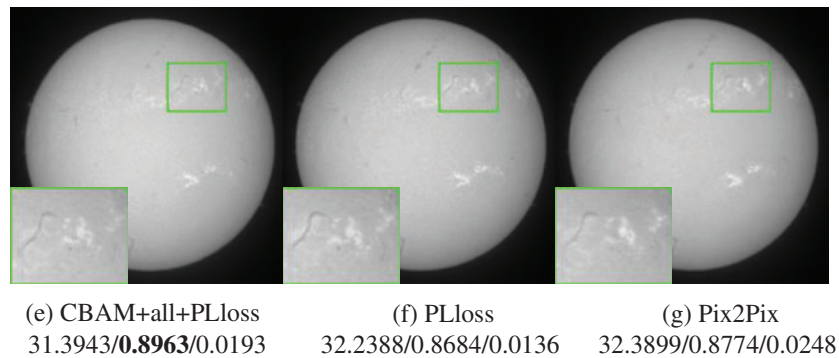


Figure 3: Comparison of the ablation experiment in detail repair of solar activity. The PSNR, structural similarity (SSIM) and LPIPS indexes of the picture are marked on the picture label

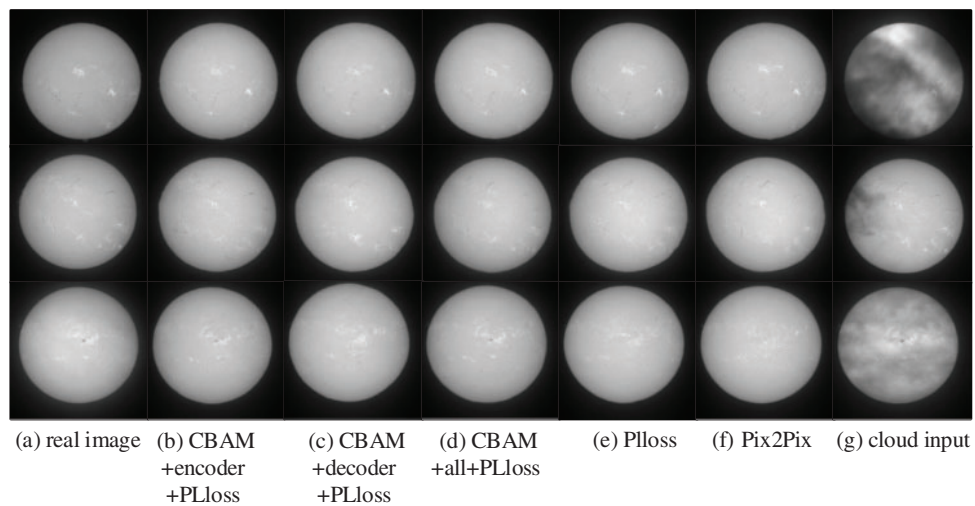


Figure 4: Results of the ablation experiment

The ablation experiment mainly considers the combination of CBAM module and the encoder and decoder of Pix2Pix model, as well as the impact of perceptual loss on the final generated image. PSNR, SSIM and LPIPS were selected to compare the experimental results quantitatively. The specific experimental results are shown in [Tab. 1](#).

Table 1: Ablation experiment results

Methods	PSNR	SSIM	LPIPS
PLloss + CBAM + encoder	26.3014	0.8356	0.0398
PLloss + CBAM + decoder	27.3012	0.8369	0.0330
PLloss + CBAM + all	26.9278	0.8419	0.0421
PLloss	26.9250	0.8351	0.0319
Pix2Pix	26.7709	0.8415	0.0351

It is difficult to evaluate the difference of ablation experimental results from the perspective of visual effect. From the experimental indicators, perceptual loss improves the ability of the model to restore deep features, and CBAM module improves the image quality generated by Pix2Pix model. But the combination of the two is not a simple additive relationship. It can be seen from LPIPS indicators that CBAM module will affect the restoration of deep features of the model to a certain extent. Therefore, it is also necessary to consider the combination relationship between CBAM module and different structures of the original model for finding the model with the best effect of removing cloud shadow. From the above experimental results, it can be concluded that the structure of CBAM combined with PLloss and decoder in Pix2Pix model generator has a good effect of removing cloud shadow and restoring image detail features. Although the LPIPS and SSIM is not optimal, there is little difference between the model and the optimal result of LPIPS and SSIM.

4.3 Comparison with Dehaze Methods

The method in this paper is qualitatively and quantitatively compared with the previous dehaze methods. Methods include dark channel prior (DCP) [4], color attenuation prior (CAP) [29], enhanced Pix2Pix dehazing network (EPDN) [30], FFA-net [10] and Pix2Pix [24]. The experimental results are shown in Figs. 5 and 6:

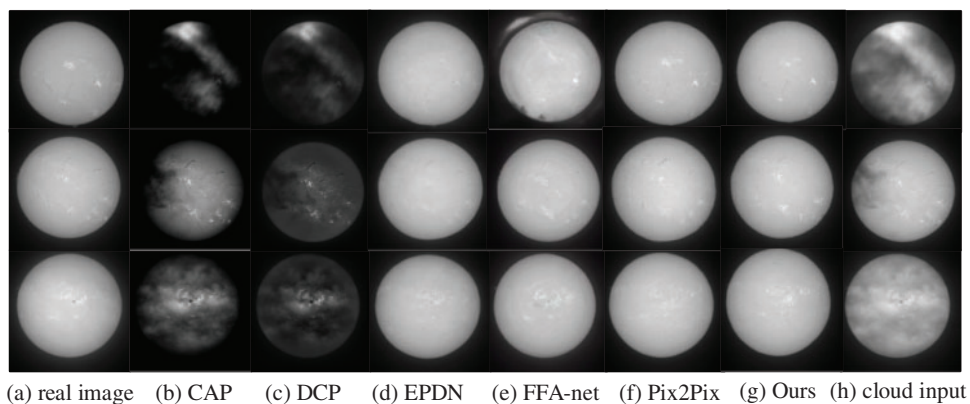


Figure 5: Comparison results of dehaze algorithms

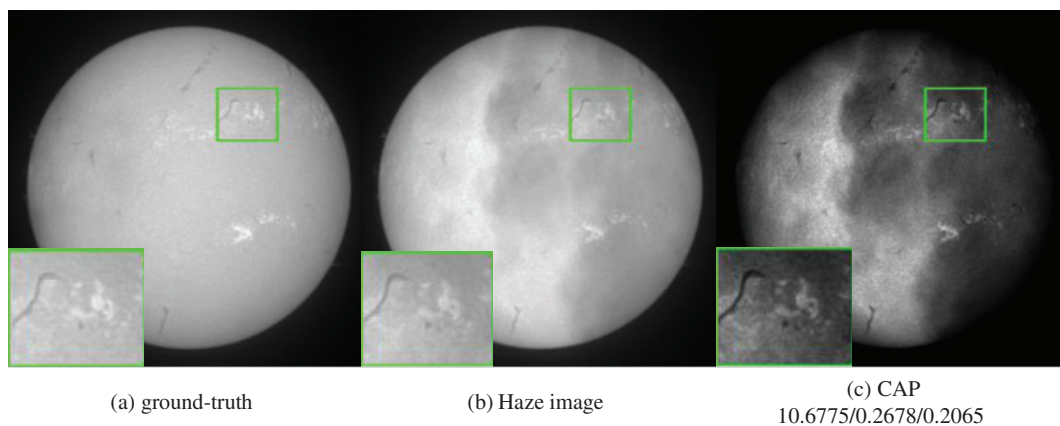


Figure 6: (Continued)

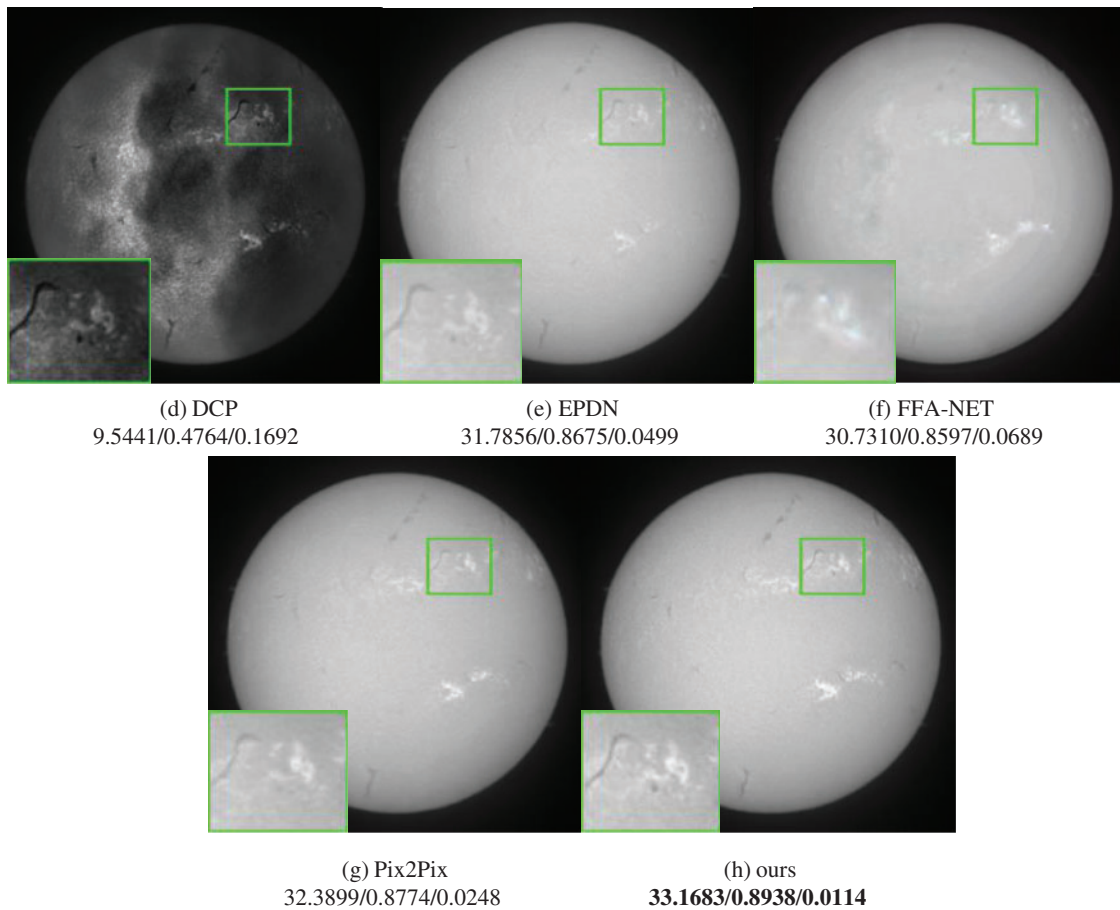


Figure 6: Comparison of dehaze algorithms in detail repair of solar activity. The PSNR, SSIM and LPIPS indexes of the picture are marked on the picture label

Fig. 6 shows the restoration of solar activity details by different haze removal models when they remove pollution from the full-disk H-alpha images. According to the experimental results, the traditional algorithm cannot remove pollution, and the shielding of pollution on solar activity details is enhanced. Deep learning-based haze removal algorithm can remove the shadows brought by thin clouds, but it cannot accurately restore the detailed features of solar activity. In addition, according to the LPIPS index in Fig. 6h, the enhanced Pix2Pix model also achieves better results in image deep feature restoration than other haze removal algorithms. Fig. 5 shows the performance of different algorithms on the test set. In general, the Pix2Pix model has better restoration results than other models. Tab. 2 shows the index performance of different haze algorithms in the test set. The improved model is slightly lower than the depth learning dehaze model in SSIM index, which may be because PLloss and CBAM module affect the overall image structure restoration. PSNR and LPIPS are superior to other dehaze algorithms. According to the above results, the improved Pix2Pix model can effectively enhance the recovery of solar activity details under the influence of thin clouds, and improve the quality of the generated pollution-free image.

Table 2: Comparison with dehaze model test results

	DCP	CAP	EPDN	FFA-net	Pix2Pix	Ours
PSNR	8.6247	9.9705	27.2328	26.1623	26.7709	27.3012
SSIM	0.4295	0.2431	0.8460	0.8414	0.8415	0.8369
LPIPS	0.1784	0.2065	0.0499	0.0784	0.0351	0.0330

5 Conclusion

Solar image analysis is an important part of the study of solar activity. Removing cloud pollution from images is important for all Ground based observations. In this paper, Pix2Pix network is used to convert polluted images to clear images, which has achieved the effect of removing the pollution from the full-disk H-alpha images. In addition, the CBAM module and perceptual loss are used to improve the model, so that the visual effect of the generated cloudless image is more realistic, and the purpose of image cloud removal is realized. The experimental results show that compared with the traditional mainstream depth learning image dehaze algorithm, the model can use the data set of the HSOS to learn more effective solar image features and synthesize clearer cloudless images. The results of this study will be deployed to solar physics community to improve the image quality of the full disk dataset of solar images. Future work will focus on generating larger solar image data sets, obtaining data sets with various solar activity conditions, making the network have better generalization ability.

Acknowledgement: Authors are thankful to the Huairou Solar Observing Station (HSOS) which provided the data. Authors gratefully acknowledge technical and financial support from CAS Key Laboratory of Solar Activity, Media Computing Lab of Minzu University of China and New Jersey Institute of Technology. We acknowledge for the data resources from “National Space Science Data Center, National Science Technology Infrastructure of China” (<https://www.nssdc.ac.cn>).

Funding Statement: Funding for this study was received from the open project of CAS Key Laboratory of Solar Activity (Grant No: KLSA202114), the crossdiscipline research project of Minzu University of China (2020MDJC08).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] F. Hill, G. Fischer, S. Forgach, J. Grier, J. W. Leibacher *et al.*, “The global oscillation network group site survey,” *Solar Physics*, vol. 152, no. 2, pp. 321–349, 1994.
- [2] N. Zhang, Y. Yang, R. Li and K. Ji, “A real-time image processing system for detecting and removing cloud shadows on h-alpha full-disk solar (in Chinese),” *Astronomical Research and Technology*, vol. 13, no. 2, pp. 242–249, 2016.
- [3] S. Feng, J. Lin, Y. Yang, H. Zhu, F. Wang *et al.*, “Automated detecting and removing cloud shadows in full-disk solar images,” *New Astronomy*, vol. 32, no. 4, pp. 24–30, 2014.
- [4] K. He, J. Sun and X. Tang, “Single image haze removal using dark channel prior,” in *Proc. of the 2009 IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, Florida, USA, vol. 32, pp. 1956–1963, 2009.
- [5] R. Fattal, “Single image dehazing,” *ACM Transactions on Graphics*, vol. 27, pp. 1–9, 2008.

- [6] L. Sun, Y. Zhang, X. Chang and J. Xu, "Cloud-aware generative network: Removing cloud from optical remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 4, pp. 691–695, 2020.
- [7] M. Qin, F. Xie, W. Li, Z. Shi and H. Zhang, "Dehazing for multispectral remote sensing images based on a convolutional neural network with the residual architecture," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 5, pp. 1645–1655, 2018.
- [8] K. Enomoto, K. Sakurada, W. Wang, H. Fukui, M. Matsuoka *et al.*, "Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets," in *Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Hawaii, USA, pp. 1533–1541, 2017.
- [9] B. Li, X. Peng, Z. Wang, J. Xu and D. Feng, "An all-in-one network for dehazing and beyond," *ArXiv preprint arXiv:1707.06543*, pp. 1–12, 2017.
- [10] X. Qin, Z. Wang, Y. Bai, X. Xie and H. Jia, "FFA-net: Feature fusion attention network for single image dehazing," in *Proc. of the AAAI Conf. on Artificial Intelligence*, New York, USA, vol. 34, no. 7, pp. 11908–11915, 2020.
- [11] J. Zhang, X. Qi, S. H. Myint and Z. Wen, "Deep-learning-empowered 3d reconstruction for dehazed images in iot-enhanced smart cities," *Computers Materials & Continua*, vol. 68, no. 2, pp. 2807–2824, 2021.
- [12] I. Goodfellow, J. P. Abadie, M. Mirza, B. Xu and D. Farley, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2014.
- [13] M. Arjovsky, S. Chintala and L. Bottou, "Wasserstein GAN," *ArXiv preprint arXiv:1701.07875*, pp. 1–32, 2017.
- [14] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *ArXiv preprint arXiv:1411.1784*, pp. 1–7, 2014.
- [15] Y. Yang, Z. Gong and P. Zhong, "Unsupervised representation learning with deep convolutional neural network for remote sensing images," in *Proc. of the 9th Int. Conf. on Image and Graphics*, Shanghai, China, 10667, pp. 97–108, 2017.
- [16] W. Fang, F. H. Zhang, V. S. Sheng and Y. W. Ding, "A method for improving CNN-based image recognition using DCGAN," *Computers Materials & Continua*, vol. 57, no. 1, pp. 167–178, 2018.
- [17] X. Hao, X. Meng, Y. Zhang, J. Xue and J. Xia, "Conveyor-belt detection of conditional deep convolutional generative adversarial network," *Computers Materials & Continua*, vol. 69, no. 2, pp. 2671–2685, 2021.
- [18] T. Zhang, Z. Zhang, W. Jia, X. He and J. Yang, "Generating cartoon images from face photos with cycle-consistent adversarial networks," *Computers Materials & Continua*, vol. 69, no. 2, pp. 2733–2747, 2021.
- [19] K. Fu, J. Peng, H. Zhang, X. Wang and F. Jiang, "Image super-resolution based on generative adversarial networks: A brief review," *Computers Materials & Continua*, vol. 64, no. 3, pp. 1977–1997, 2020.
- [20] J. Hu, L. Shen, S. Albanie, G. Sun and E. Wu, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.
- [21] H. Zhang, I. Goodfellow, D. Metaxas and A. Odena, "Self-attention generative adversarial networks," in *Proc. of the 36th Int. Conf. on Machine Learning*, Long Beach, California, USA, pp. 7354–7363, 2019.
- [22] A. Vaswani, N. Shazeer, N. Parmar, J. U. Szokoreit, L. Jons *et al.*, "Attention is all you need," in *Proc. of the Neural Information Processing Systems*, Long Beach, California, USA, pp. 5998–6008, 2017.
- [23] S. Woo, J. Park, J. Park and J. Lee, "CBAM: Convolutional block attention module," in *Proc. of the European Conf. on Computer Vision (ECCV)*, Munich, Germany, pp. 3–19, 2018.
- [24] P. Isola, J. Zhu, T. Zhou and A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Hawaii, USA, vol. 12, pp. 5967–5976, 2017.
- [25] C. Li and M. Wand, "Combining markov random fields and convolutional neural networks for image synthesis," in *Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 2479–2486, 2016.
- [26] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 2536–2544, 2016.

- [27] J. Johnson, A. Alahi and F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. of the European Conf. on Computer Vision (ECCV)*, Amsterdam, Netherlands, vol. 9906, pp. 694–711, 2016.
- [28] D. Kingma and J. Ba, "Adam: A method for stochastic optimization, compute," in *Proc. of the 3rd Int. Conf. for Learning Representations*, San Diego, CA, USA, vol. 26, pp. 14–17, 2015.
- [29] Q. Zhu, J. Mai and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3522–3533, 2015.
- [30] Y. Qu, Y. Chen, J. Huang and Y. Xie, "Enhanced Pix2Pix dehazing network," in *Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, USA, pp. 8152–8160, 2019.