Research in Astronomy and Astrophysics, 22:065009 (10pp), 2022 June https://doi.org/10.1088/1674-4527/ac69b7

Image Desaturation for SDO/AIA Using Mixed Convolution Network

Xuexin Yu[1,2], Long Xu[3,4], Zhixiang Ren[3], Dong Zhao[5], and Wenqing Sun[1,2]
[1] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; lxu@nao.cas.cn
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Peng Cheng National Laboratory, Shenzhen 518000, China
[4] National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China
[5] State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China
Received 2022 March 16; revised 2022 April 6; accepted 2022 April 22; published 2022 May 20

Abstract

The Atmospheric Imaging Assembly (AIA) onboard the Solar Dynamics Observatory (SDO) provides full-disk solar images with high temporal cadence and spatial resolution over seven extreme ultraviolet (EUV) wave bands. However, as violent solar flares happen, images captured in EUV wave bands may have saturation in active regions, resulting in signal loss. In this paper, we propose a deep learning model to restore the lost signal in saturated regions by referring to both unsaturated/normal regions within a solar image and statistical probability model of massive normal solar images. The proposed model, namely mixed convolution network (MCNet), is established over conditional generative adversarial network (GAN) and the combination of partial convolution (PC) and validness migratable convolution (VMC). These two convolutions were originally proposed for image inpainting. In addition, they are implemented only on unsaturated/valid pixels, followed by certain compensation to compensate the deviation of PC/VMC relative to normal convolution. Experimental results demonstrate that the proposed MCNet achieves favorable desaturated results for solar images and outperforms the state-of-the-art methods both quantitatively and qualitatively.

Key words: Sun: activity – Sun: atmosphere – Sun: chromosphere

1. Introduction

The Atmospheric Imaging Assembly (AIA) (Lemen et al. 2012) onboard the Solar Dynamics Observatory (SDO) (Pesnell et al. 2012) is composed of four dual-channel normal-incidence telescopes that capture full-disk images of the Sun's atmosphere over seven extreme ultraviolet (EUV) (94 Å, 131 Å, 171 Å, 193 Å, 211 Å, 304 Å, 335 Å) wave bands with spatial resolution of 4096 × 4096 and temporal cadence of 12s. These captured data provide an unprecedented EUV view for studying the structure and dynamics of the solar atmosphere.

However, when a solar flare occurs, the images captured by SDO/AIA in the EUV wave bands may present two kinds of artifacts, saturation and diffraction, as shown in Figure 1(a), which are closely associated with imaging process of AIA. Concretely, an image of AIA is the result of convolution between the incoming photon flux and point-spread function (PSF) which describes the response of AIA to an ideal point source. This process is formulated as

$$I = (A_c + A_d) \otimes f = A_c \otimes f + A_d \otimes f, \quad (1)$$

where I is an image recorded by AIA, $\otimes$ means convolution operator, f denotes the actual incoming photon flux, A_c and A_d are diffusion component and diffraction component of PSF, respectively. As shown in Figures 1(b) and (c), A_c is a core peak, and A_d is a peripheral regular diffraction pattern of varying intensity and replicates the core peak. Diffraction fringe is the convolution between A_d and f. It becomes apparent above the background with increasing intensity of a given peak in f. Saturation happens in $A_c \otimes f$ term, which is actually categorized into primary saturation and second saturation/blooming. The former occurs because the charge-coupled device (CCD) pixels cannot accommodate additional charges of incoming photon flux f, while the latter is the result that the additional charges spill into neighbor pixels. From Equation (1), intense incoming flux may lead to signal loss in the component of $A_c \otimes f$ in case of saturation, but it is also coherently and linearly scattered to other regions due to diffraction ($A_d \otimes f$) (Guastavino et al. 2019). Therefore, lost signal in primary saturation actually presents in diffraction fringes more or less, so it can be retrieved partially from diffraction fringes (Schwartz et al. 2014; Torre et al. 2015). In principle, DESAT (Schwartz et al. 2015) formulated lost signal recovery in saturated regions into an inverse diffraction issue, which is described as

$$I_d = A_d \otimes f + B_d, \quad (2)$$

where I_d is the known recorded image in diffraction regions, B_d is the unknown saturated image background related to diffraction fringes. In Schwartz et al. (2015), B_d is estimated from the interpolation of two neighbor unsaturated images.

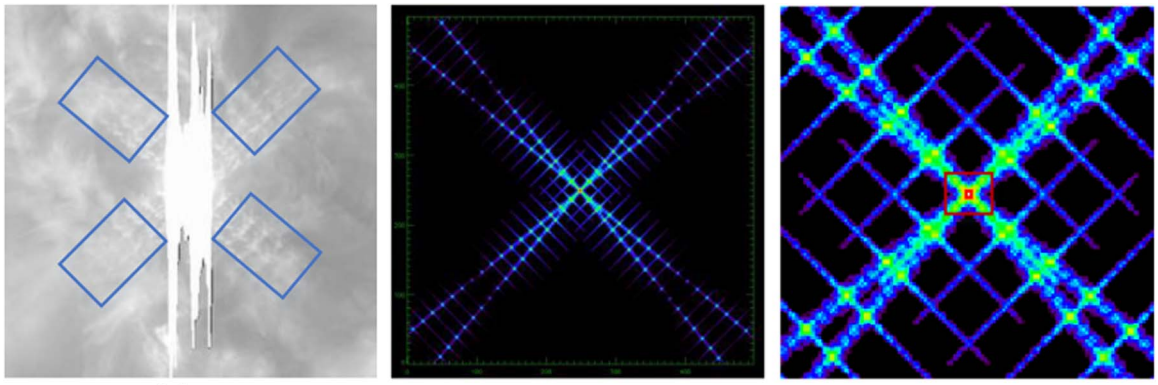


Figure 1. (a) An example of saturated image in active region 12 130 for SDO/AIA at 193 Å wave band with over-saturation region and the diffraction fringes highlighted by blue bounding box (the event occurred at 14:47:06 UT, 2014 August 1). (b) The complete point-spread function (PSF) of 193 Å wave band. (c) Zoomed-in central part of the PSF, where region highlighted by red bounding box is diffusion component denoted by A_c, and the rest parts are diffraction component denoted by A_d.

These two unsaturated images are obtained by reducing exposure time, which is automatically triggered by the feedback system of SDO/AIA during the solar flares. However, DESAT becomes ineffective for large solar flares because the neighbor images of short exposure time may be also saturated, e.g., the super storm happening in 2017 September. To solve this issue, Guastavino et al. (2019) proposed Sparsity-Enhancing DESAT (SE-DESAT) where estimation of saturated image background is not from consecutive unsaturated images but from the current saturated image itself. However, in these two methods, segmentation of diffraction fringes and primary saturation regions and estimation of background would affect desaturated results. In addition, the blooming regions cannot be restored in both methods.

With significant success of deep learning in image inpainting, two learning-based approaches, Mask-Pix2Pix (Zhao et al. 2019) and PCGAN (Yu et al. 2021), were proposed to desaturate solar images in our previous efforts. They differ from DESAT (Schwartz et al. 2015) and SE-DESAT (Guastavino et al. 2019) in three aspects. First, DESAT (Schwartz et al. 2015) and SE-DESAT (Guastavino et al. 2019) explicitly model the problem and resolve it under the assumption that lost signal in saturated regions may present in diffraction fringes of unsaturated regions, while Mask-Pix2Pix and PCGAN model the problem implicitly by using neural network. Specially, a neural network first learns the distribution of unsaturated image from massive data, and then infers the lost signal in saturated regions from well-learnt distribution. Second, Mask-Pix2Pix and PCGAN have stronger representation ability than DESAT and SE-DESAT because neural networks can approximate any complex function theoretically (Hornik et al. 1989; Cybenko 1989; Leshno et al. 1993). Third, compared with DESAT and SE-DESAT, Mask-Pix2Pix and PCGAN automatically extract related information (including diffraction fringes) to restore the whole saturated regions including blooming and primary saturation in an end-to-end optimization manner. They do not need explicit segmentation of diffraction fringes and primary saturation regions and estimation of background. In an image, saturated region contains useless/invalid pixels. Once a standard convolution slides to the boundary of saturated region, invalid pixels would participate in convolution, resulting in deviation of convolution, e.g, Mask-Pix2Pix (Zhao et al. 2019). To overcome this problem, partial convolution (PC) (Liu et al. 2018) was employed to replace standard convolution in our previous effort (Yu et al. 2021), which excludes saturated pixels from block-wise convolution and compensates deviation of PC to approach normal convolution as far as possible.

In this paper, to further improve deaturating results, we propose a mixed convolution network (MCNet), where validness migratable convolution (VMC) (Wang et al. 2021) and partial convolution (PC) (Liu et al. 2018) are employed in encoder and decoder of generator, respectively. These two types of convolutions can extract features only from normal regions, and compensate deviation caused by saturated pixels in encoder and decoder in different ways respectively, which is beneficial to recovery of saturated regions.

The rest of paper is organized as follows. Section 2 introduces data set used in this work. Section 3 introduces network architecture, convolutions and loss functions of the proposed model in details. Experimental results are provided in Section 4. Conclusion and discussion are listed in Section 5.

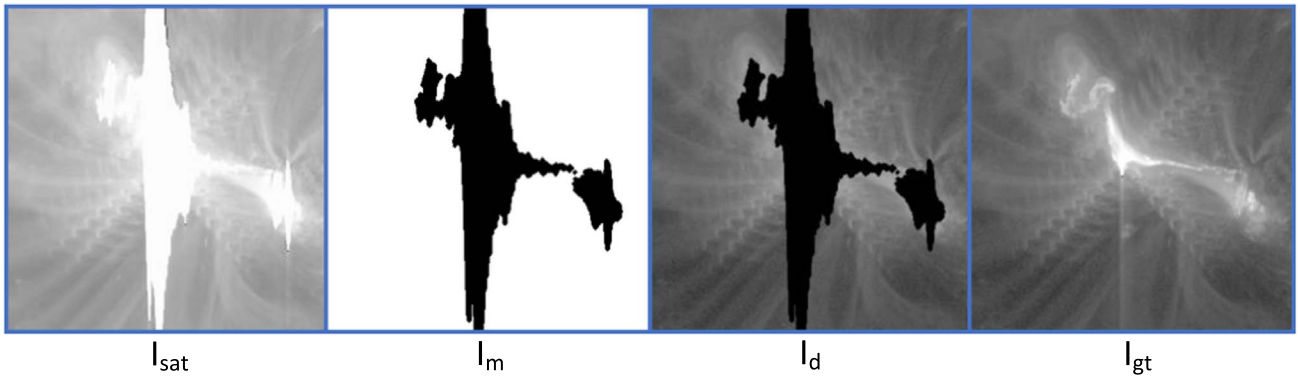


Figure 2. An sample in desaturation data set established by Yu et al. (2021), which is composed of four images: $I_{\rm sat}$, $I_{\rm gt}$, I_m and I_d. $I_{\rm sat}$ is a saturated image and $I_{\rm gt}$ is the unsaturated image following $I_{\rm sat}$ immediately. I_m is a binary mask which indicates saturated and unsaturated pixels of $I_{\rm sat}$ by 1 and 0, respectively. I_d is the degraded image which is obtained by $I_{\rm gt} \odot I_m$.

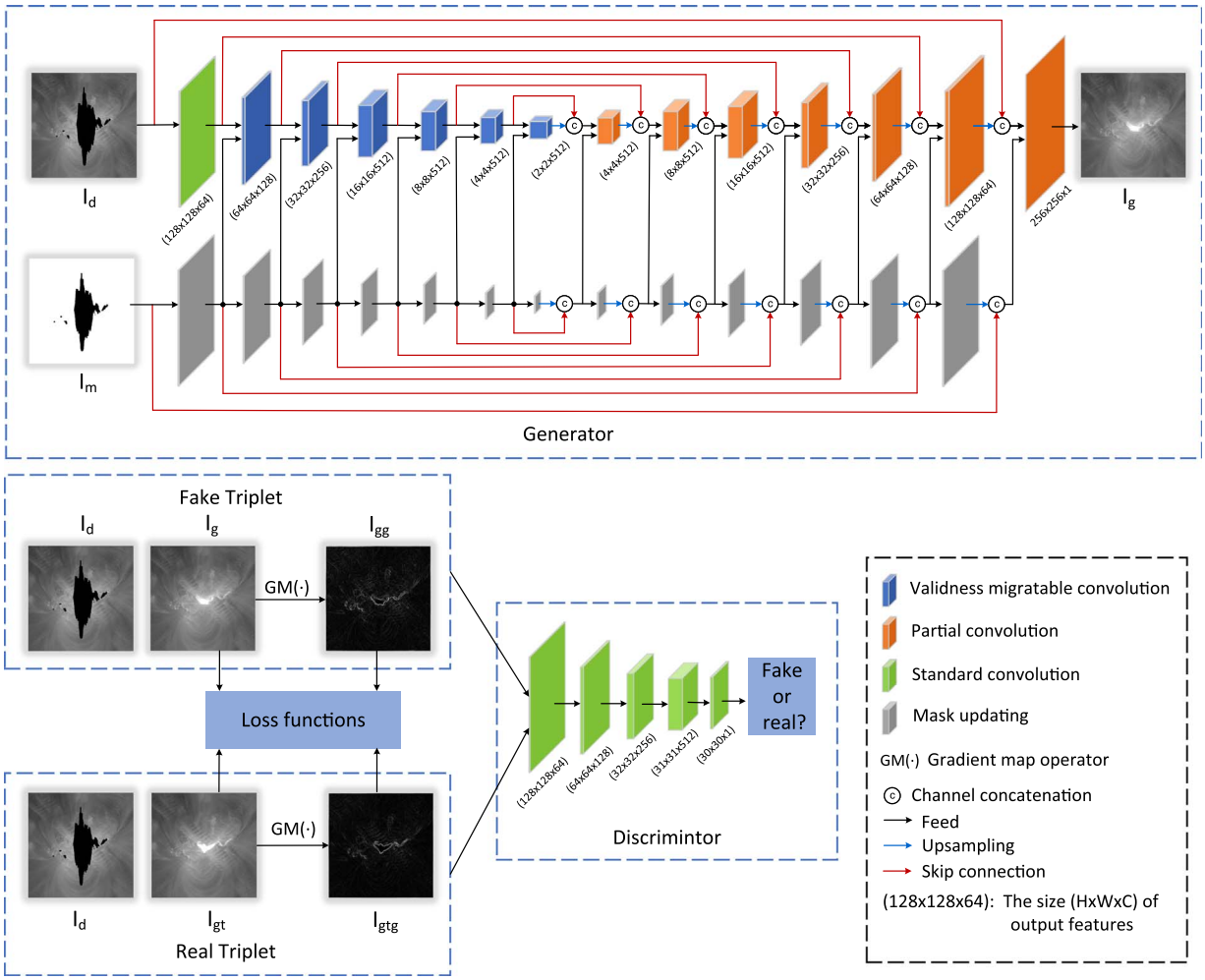


Figure 3. The architecture of the proposed mixed convolution network (MCNet). The generator learns a mapping from I_m and I_d to $I_{\rm gt}$. The discriminator and loss functions supervise the learning process of the generator by classifying fake $\{I_d, I_g, I_{gg}\}$ and real $\{I_d, I_{\rm gt}, I_{\rm gtg}\}$ pairs and minimizing distance between I_g and $I_{\rm gt}$ from pixel-level and feature-level, respectively.

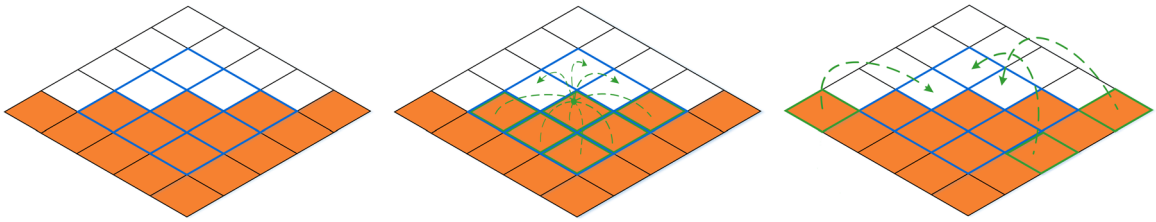


Figure 4. The differences among standard convolution, partial convolution and validness migratable convolution when receptive field of convolution contains saturated and unsautrated pixels. The white and orange boxes denote saturated and unsaturated pixels, respectively. The blue grid means receptive field. The green boxes represent unsaturated pixels which are used to fill saturation location in receptive field.

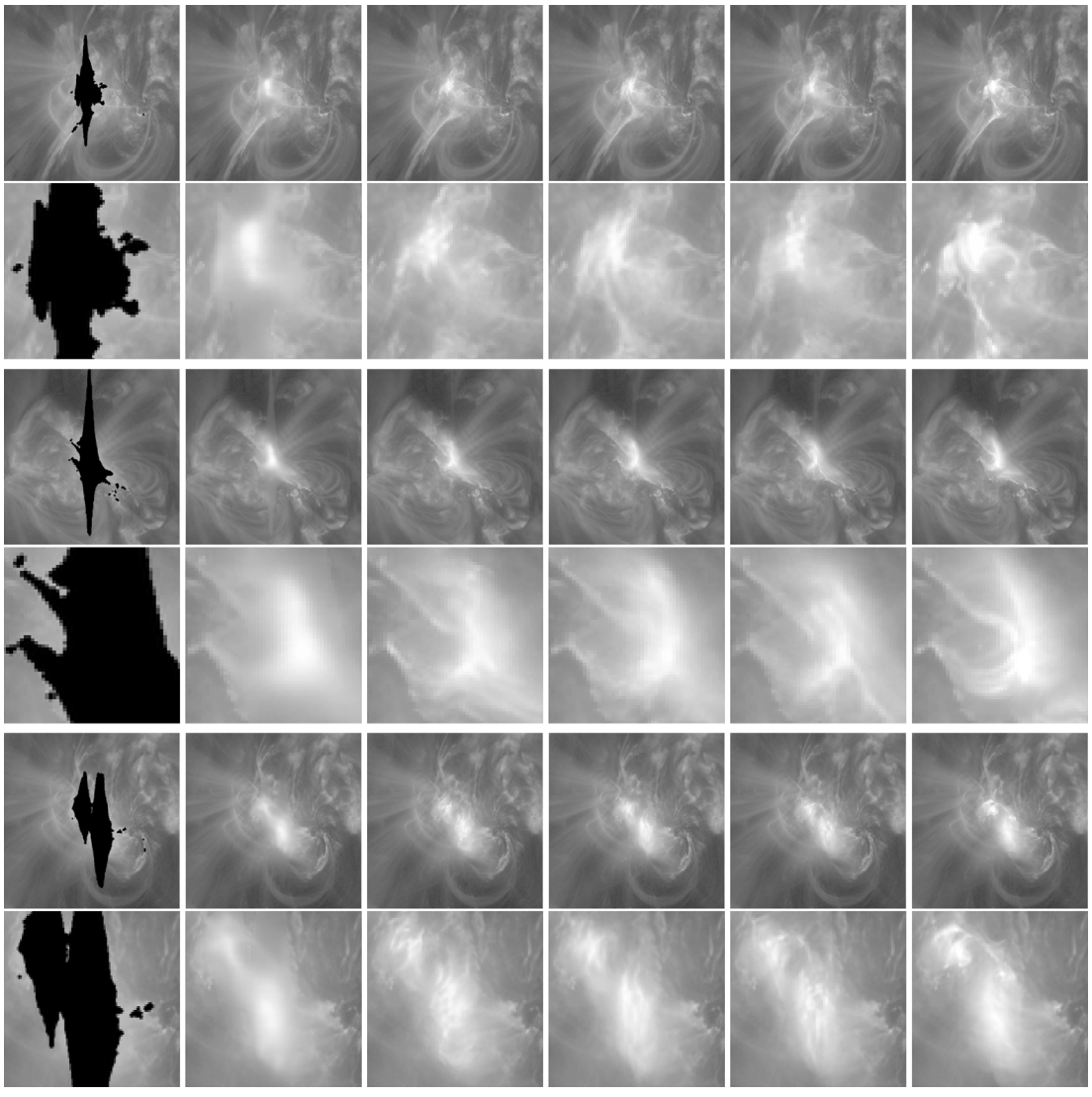


Figure 5. Visual quality comparison of desaturated results by Mask-Pix2Pix (Zhao et al. 2019), PCGAN (Yu et al. 2021), our MCNet and MCNet (w/o VMC). From top to bottom, the first, third and fifth rows are full-images, and the others are zoomed-in patches.

Table 1
Quantitative Comparison with State-of-the-art Methods on Testing Set

Methods	PSNR	SSIM
Mask-Pix2Pix (Zhao et al. 2019)	25.5839	0.7307
PCGAN (Yu et al. 2021)	29.2050	0.8146
MCNet (w/o VMC)	29.0644	0.8111
MCNet (ours)	29.6819	0.8306

2. Data set

To train and evaluate the proposed model, the desatuartion data set (Yu et al. 2021) is used in this work, which collected M-class and X-class solar flare data at 193 Å of SDO/AIA (Lemen et al. 2012) from 2010 to 2017. After a series of data pre-processing, such as normalization for exposure time, scaling using the log function, as shown in Figure 2, each sample is composed of image quadruple, saturated, mask, degraded and unsaturated image, whose size are 256×256. It is worth noting that saturated image $I_{\rm sat}$ is only utilized to obtain realistic shape of saturated regions by imposing a threshold on it, which is denoted by I_m. Degraded image I_d is the result of $I_{\rm gt} \odot I_m$ ($\odot$ represents element-multiplication operator). During the training of model, the triplet $\{I_d, I_m, I_{\rm gt}\}$ is fed into our model to optimize parameters of network. The whole data set contains about 18,700 samples. Following Yu et al. (2021), we split data set in chronological order: 2012-2017 for training (15,700 samples) and 2010-2011 for testing (3000 samples).

3. Method

In this section, we first introduce network architecture of the proposed model, and then convolutions and loss functions are discussed.

3.1. Network Architecture

The overall architecture of the proposed MCNet is shown in Figure 3, which is composed of a generator and a discriminator. The generator is a UNet-like architecture, which obtains favorable results in image inpainting. Concretely, it consists of encoder that extracts a representation of input, and decoder that utilizes the representation to output an image with the same as original input. The basic module of generator consists of VMC/PC, regional composite normalization (RCN) (Wang et al. 2021), and ReLU/LeakyReLU, and is stacked repeatedly in generator. In addition, skip connection is implemented in corresponding layer of encoder and decoder to shuttle different level information between them. The mask of image is nontrivial during encoding and decoding process because it indicates the unsaturated/normal regions and saturated regions/holes by 1 and 0, respectively. Therefore, it has an independent updating branch and then is fed into corresponding layer of image branch to effectively extract features from images. Following Isola et al. (2017) and Zhu et al. (2017), discriminator is a PatchGAN architecture which judge whether an input patch is real. For an input, the discriminator outputs a matrix where each element corresponds an overlap patch of input. In our work, its input is a triplet including input degraded image I_d, the generated image I_g or the ground truth $I_{\rm gt}$ and its corresponding gradient map I_{gg} or $I_{\rm gtg}$.

3.2. Convolutions

Our goal is to restore the lost signal in saturated/invalid regions or holes from unsaturated/valid/normal regions within an image. To achieve the goal, convolution need to meet two requirements. First, the output of convolution is only conditioned on unsaturated pixels because the lost signal is only in unsaturated regions. Second, deviation caused by unsaturated pixels should be compensated when receptive field of convolution crosses the boundary between normal regions and holes. Therefore, the validness migratable convolution (VMC) (Wang et al. 2021) and partial convolution (PC) (Liu et al. 2018) are employed in encoder and decoder of the generator respectively instead of standard convolution, because both of them can fulfill the above requirements. Given the input degraded image/feature map x, convolution weight w and bias b, the standard convolution is described as

$$y(p_0) = b + \sum_{p_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n), \tag{3}$$

where p_0 denotes each position on the output features map y, $\mathcal{R}$ defines the receptive field and dilation size of convolution. For example, 3×3 receptive field with dilation 1 is formulated as

$$\mathcal{R} = \{(-1, -1), (-1, 0),...,(0, 1), (1, 1)\}. \tag{4}$$

This process is shown in Figure 4 when receptive field of convolution contains valid and invalid pixels. We can see that standard convolution use both valid and invalid pixels and cannot treat these two kinds of pixels with differences in Figure 4, which cannot meet two proposed requirements for convolution.

To solve this issue, the PC is introduced, which is described as

$$y(p_0) = \begin{cases} b + \dfrac{\sum_{p_n \in \mathcal{R}} \mathbf{1}}{\sum_{p_n \in \mathcal{R}} m(p_0 + p_n)} \sum_{p_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n) \cdot m(p_0 + p_n), & if \sum_{p_n \in \mathcal{R}} m(p_0 + p_n) > 0 \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

$$m_0 = \begin{cases} 1, & if \sum_{p_n \in \mathcal{R}} m(p_0 + p_n) > 0 \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where m is corresponding mask (0 for saturated pixels and 1 for normal pixels) of x, $\mathbf{1}$ is a constant all-ones matrix with the same size as m, and m is updated by Equation (6) for the next layer. From Equations (5) and (6), output of PC only depends on valid pixels by introducing a mask, and deviation caused by invalid pixels is calibrated by scaling the standard output according to proportion of valid pixels in receptive field. This process can be equivalent to filling holes with valid pixels in current receptive field to exclude invalid pixels and compensate deviation caused by them before implementing standard convolution, which is shown in Figure 4.

VCM solves the above issue in a different way. It consists of feature migration, regional combination, convolution and mask updating. They are formulated by

$$y(p_0) = \sum_{p_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n + \triangle p_n), \tag{7}$$

where irregular receptive $\mathcal{R}$ is augmented with feature migration $\triangle p_n$, y is the deformed feature map of x. The $\triangle p_n$ is automatically learned from input feature map, and $n = 1,...,N$ and $N = |\mathcal{R}|$.

$$y_{rc} = y \odot (1 - m) + m \odot x, \tag{8}$$

$$y_{\text{out}}(p_0) = b + \sum_{p_n \in \mathcal{R}} w(p_n) \cdot y_{rc}(p_0 + p_n). \tag{9}$$

After validness migratable convolution, mask m is updated by Equation (6). From Equations (7)–(9), we can see that the holes of input feature map x is first filled by surround pixels in the learning way, then filled regions are copied into holes of x by Equation (8), and a standard convolution is imposed on new feature map y_{rc} lastly. This process is simply treated as filling holes with pixels outside receptive field to avoid the interference of invalid pixels before applying standard convolution, as shown in Figure 4.

Both these convolutions extract features only from normal regions, while their compensating deviation caused by saturated pixels are different. The PC uses valid pixels inside receptive field, while the VMC uses valid pixels outside receptive field as shown in Figure 4. To get more photorealistic recovery of saturated regions/holes, we apply VMC and PC to encoder and decoder of generator, respectively. To encoder, VCM has more choices to get valid pixels outside receptive field as shown in Figure 4(c) to fill holes while PC is limited to receptive field as shown in Figure 4(b). Roughly speaking, saturated regions/holes are gradually filled along with convolution progress, so encoder finally gets encoding features without holes for decoder. In decoder, encoding features with holes in each layer of encoder is provided by skip connection, while decoding features coming from the last layer of encoder are without holes. Thus, PC is employed in decoder, using decoding features within the receptive field to fill holes indicated by encoding features.

3.3. Loss Functions

To restore missing signal in saturated regions, multiple losses are integrated together to minimize the difference between generated image and ground truth from both pixel-level and feature-level. The pixel-level losses include pixel reconstruction loss (Liu et al. 2018), gradient loss (Ma et al. 2020) and total variation loss (Johnson et al. 2016), and feature-level losses includes perceptual loss (Johnson et al. 2016), style loss (Gatys et al. 2016) and adversarial loss (Mao et al. 2017).

Let I_d be the input degraded image, I_m initial binary mask (0 for saturated/invalid regions, 1 for normal/valid regions), I_g the generated image by generator, and I_{gt} the ground truth, we introduce pixel reconstruction loss to guarantee pixels similarity of output image. It is defined as

$$\begin{aligned} \mathcal{L}_{\text{rec}}(G) = {} & \lambda_{\text{hole\_rec}} \|(1 - I_m) \odot (I_g - I_{\text{gt}})\|_1 \\ & + \lambda_{\text{valid\_rec}} \|I_m \odot (I_g - I_{\text{gt}})\|_1, \end{aligned} \tag{10}$$

where $\|\cdot\|_1$ denotes L_1 loss, and $\lambda_{\text{hole\_rec}}$ and $\lambda_{\text{valid\_rec}}$ are the corresponding weights for saturated regions and normal regions, and empirically are set 100 and 10, respectively.

The gradient loss (Ma et al. 2020) is adopted to recover structural information, which imposes L_1 loss on gradient

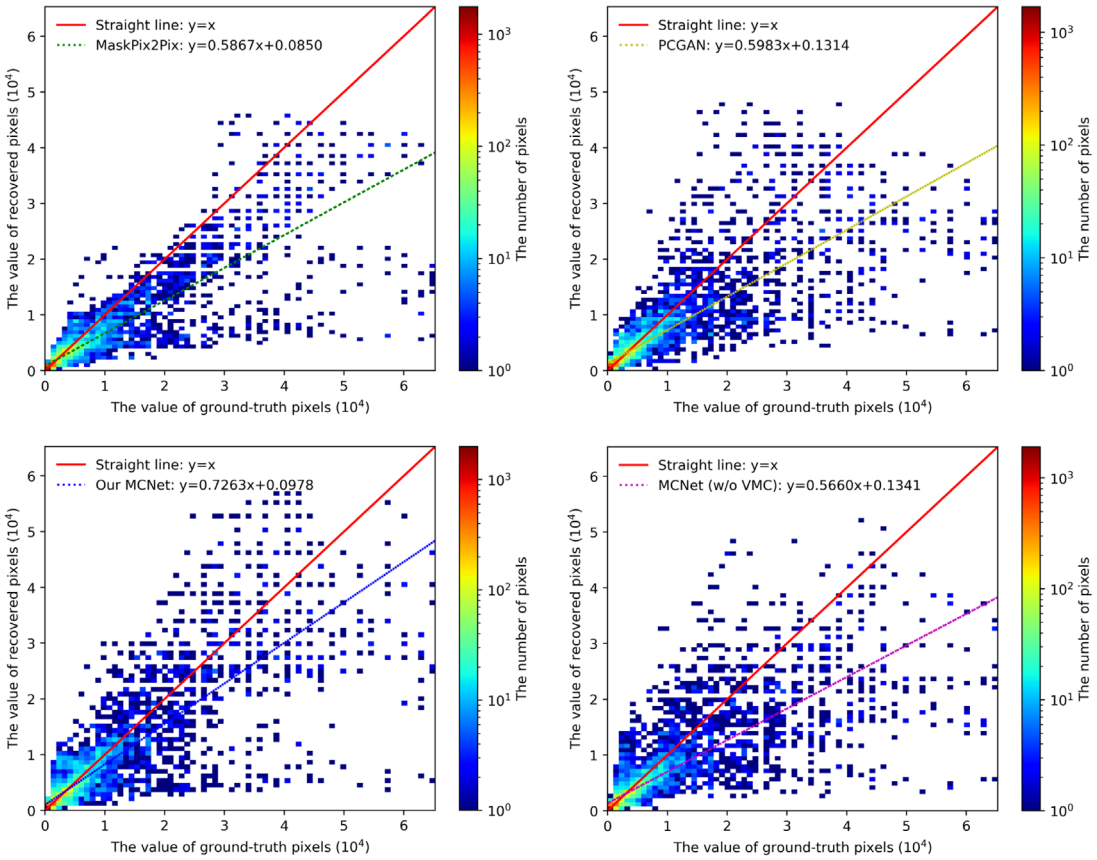


Figure 6. The linear fitting results between ground-truth pixels and recovered pixels of saturated regions for the last example in Figure 5, where x and y axes represent ground-truth pixels and recovered pixels respectively. The straight line in red color represents $x = y$, while dotted straight lines give the best linear fitting between x and y.

maps. Therefore, it is formulated as

$$\mathcal{L}_{\text{gra}}(G) = \lambda_{\text{hole\_gra}} \|(1 - I_m) \odot (GM(I_g) - GM(I_{\text{gt}}))\|_1 + \lambda_{valid\_\text{gra}} \|I_m \odot (GM(I_g) - GM(I_{\text{gt}}))\|_1, \tag{11}$$

where $\lambda_{\text{hole\_gra}}$ and $\lambda_{\text{valid\_gra}}$ are the corresponding weights of saturated regions and normal regions, empirically are set 300 and 10, respectively. Here, the $GM(\cdot)$ denotes an operator computing gradient map of an image (Ma et al. 2020), which is described as

$$\begin{aligned} X(i,j) &= I(i+1,j) - I(i-1,j), \\ Y(i,j) &= I(i,j+1) - I(i,j-1), \\ \nabla I(i,j) &= (X(i,j)), Y(i,j)), \\ GM(I) &= \|\nabla I\|_2, \end{aligned} \tag{12}$$

where $\|\cdot\|_2$ computes the length of vector $\nabla I(i,j)$ at each pixel location. To process margin pixels, the input image I is zero-padded by 1-pixel dilation before extraction of gradient map.

Total variation loss (Johnson et al. 2016) is included to eliminate artifacts in the boundary between saturated and unsaturated regions. It is calculated by

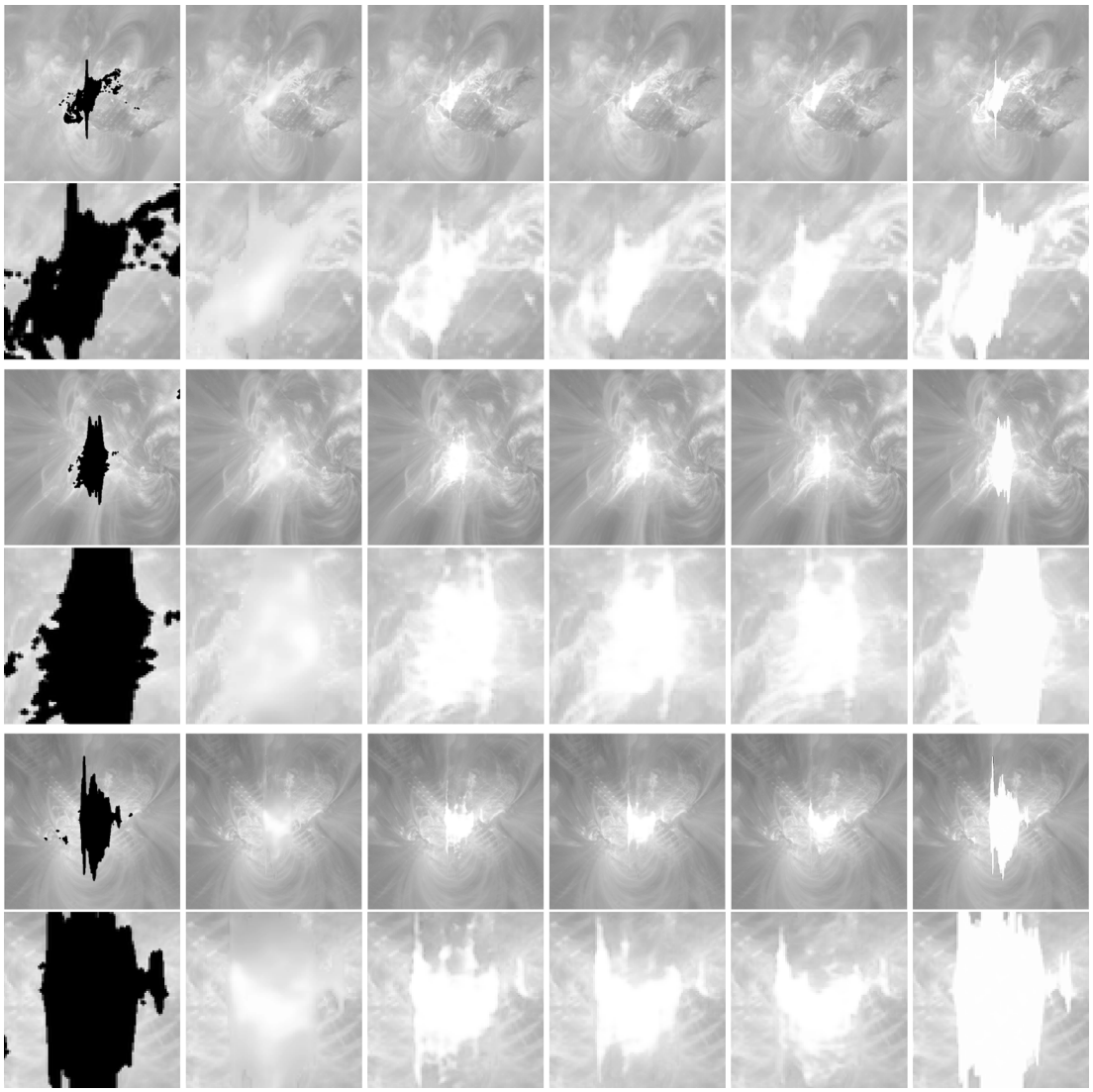
$$\mathcal{L}_{\text{tv}}(G) = \sum_{(i,j)\in P,(i,j+1)\in P} \|I_{\text{comp}}^{(i,j)} - I_{\text{comp}}^{(i,j+1)})\|_1 + \sum_{(i+1,j)\in P,(i,j)\in P} \|I_{\text{comp}}^{(i+1,j)} - I_{\text{comp}}^{(i,j)})\|_1, \tag{13}$$

where P connects saturated and unsaturated regions.

The adversarial loss (Mao et al. 2017) is adopted to ensure recovered structure information more realistic from feature-level, which is formulated as

$$\mathcal{L}_{\text{adv}}(G, D) = \mathbb{E}_{(I_d,I_{\text{gt}})}[\log D(I_d, I_{\text{gt}}, GM(I_{\text{gt}}))] + \mathbb{E}_{(I_d,I_m)}[\log(1 - D(I_d, G(I_d, I_m), GM(G(I_d, I_m))))]. \tag{14}$$

To capture high-level semantics information and alleviate grid-shaped artifacts in recovered regions (Liu et al. 2018), we



(a) Input (b) Mask-Pix2Pix (c) PCGAN (d) MCNet (e) w/o VMC (f) Saturated

Figure 7. Visual quality comparison of desaturated results for real saturated images by Mask-Pix2Pix (Zhao et al. 2019), PCGAN (Yu et al. 2021), our MCNet and MCNet (w/o VMC). From top to bottom, the first, third and fifth rows are full-images, and the others are zoomed-in patches.

introduce perceptual loss (Johnson et al. 2016):

$$\mathcal{L}_{\text{perc}}(G) = \sum_{i=1}^{T} \|\Psi_i(I_g) - \Psi_i(I_{\text{gt}})\|_1 + \sum_{i=1}^{T} \|\Psi_i(I_{\text{comp}}) - \Psi_i(I_{\text{gt}})\|_1, \quad (15)$$

where $I_{\text{comp}} = (1 - I_m) \odot I_g + I_m \odot I_{\text{gt}}$, meaning combination of recovered regions in generated image and normal regions in ground truth. The perceptual loss computes L_1 loss in feature domain for both I_g and I_{comp} and I_{gt}, respectively. The Ψ_i is the feature map of the ith pooling layer of VGG-16 (Simonyan & Zisserman 2015). Here, the first three pooling layers are adopted in Equation (15).

Style loss (Gatys et al. 2016) is effective for capturing semantics information, which first computes Gram matrix for each feature map of VGG-16 and then calculates L_1 loss. Therefore, it is defined as

$$\mathcal{L}_{\text{sty}}(G) = \sum_{i=1}^{T} \|K_i((\Psi_i(I_g))^T(\Psi_i(I_g)) - (\Psi_i(I_{\text{gt}}))^T(\Psi_i(I_{\text{gt}})))\|_1 + \sum_{i=1}^{T} \|K_i((\Psi_i(I_{\text{comp}}))^T(\Psi_i(I_{\text{comp}})) - (\Psi_i(I_{\text{gt}}))^T(\Psi_i(I_{\text{gt}})))\|_1, \quad (16)$$

where K_i is a scaling factor, given by $1/H_iW_iC_i$ for the ith layer of VGG-16, and feature map $\Psi_i(I)$ is a tensor of size $H_i \times W_i \times C_i$.

4. Experimental Results

Experiments are conducted for evaluating the proposed MCNet. We first compare our model with two state-of-the-art desaturation methods, Mask-Pix2Pix (Zhao et al. 2019) and PCGAN (Yu et al. 2021). Then effect of VCM for overall performance is verified by an ablation experiment. The source code of MCNet can be accessed via GitHub (https://github.com/filterbank/MCNet).

4.1. Implementation Details

We evaluate the proposed model on the desaturation data set (Yu et al. 2021) which is described in detail in Section 2. In the following experiments, a series of data augmentation techniques are employed in training process, including randomly cropping input image triplet (the degraded image, corresponding mask and ground truth) from 350×350 to 256×256, and randomly rotating and flipping them in four angles (0°, 90°, 180° and 270°) and horizontal direction respectively. Our model is implemented on PyTorch platform. It is trained on a single NVIDIA GeForce RTX 3090 GPU with a batch size of 28, and the epoch number of 200. We initialize convolution weights using the initialization method proposed in He et al. (2015) and optimize them by the Adam algorithm (Kingma & Ba 2014) with $\beta_1 = 0.500$ and $\beta_2 = 0.999$. The initial learning rate is set to $2e-4$, and then decays half at the 100th and 150th epoch successively.

4.2. Comparisons with State-of-the-Art

Our MCNet is compared with two benchmarks and its variant, Mask-Pix2Pix (Zhao et al. 2019), PCGAN (Yu et al. 2021), and MCNet (w/o VMC). In MCNet (w/o VMC), VCMs in encoder are replaced by PCs to verify its contribution to the proposed model.

Desaturated results in the testing set are shown in Figure 5, our method and other two benchmarks effectively recovery the overall intensity distribution of the lost signal compared with ground truths. However, Mask-Pix2Pix struggles to generate structure information and there are apparent artifacts in boundary between recovered regions and valid regions. Results of PCGAN and our approach are sharp and contain rich structure information, but fine structures in our results are more consistent with the ground truths. Although MCNet (w/o VMC) also generates favorable results, MCNet is superior to it in details of structures, which indicates the benefits of VCMs which automatically fill holes by copying surrounding pixels before convolutional operation. Following Mask-Pix2Pix and PCGAN, we also employed peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) (Wang et al. 2004) as metrics to evaluate performance of the models objectively. Table 1 shows the results of PSNR and SSIM on the whole testing set, where our model outperforms other two methods and its variant. Concretely, the proposed model receptively improves 4.0980 dB, 0.4769 dB and 0.6175 dB in PSRN, and 0.0999, 0.0160 and 0.0195 in SSIM, compared with Mask-Pix2Pix, PCGAN and MCNet (w/o VMC). Following Zhang et al. (2020), we also analyze the linear fitting result between the ground-truth pixels and recovered pixels in saturated regions. The linear fitting lines for the last example in Figure 5 are shown in Figure 6. The linear fitting slopes of MaskPix2Pix, PCGAN and MCNet (w/o VMC) are comparable, which are all close to 0.6000, while our MCNet outperforms them largely and achieves linear fitting slope of 0.7263. Our model is also superior to SE-DESAT (Guastavino et al. 2019) because it is demonstrated that its performance is inferior to PCGAN in Yu et al. (2021). Therefore, our MCNet achieves better results in qualitative and quantitative and VCMs have contribution to overall performance.

In addition, we also conduct experiments on real saturated images of solar flare to evaluate the proposed model. Figure 7 shows desaturated results of Mask-Pix2Pix, PCGAN, MCNet and MCNet (w/o VMC). It can be seen that although these models cannot completely restore all missing content in saturated regions /holes, the size of saturated regions becomes small obviously. Mask-Pix2Pix fails to recover sharp structural content. Our model MCNet performs the best, while the three benchmarks achieve similar efficiency. However, the achievement of our model is slightly below the results presented in Figure 5. The main reason lies in intensity gap between training images and real saturated images, resulting in performance drop.

5. Conclusions and Discussion

This paper proposes an MCNet model to recover saturated regions of SDO/AIA images. Compared with benchmarks, the proposed model can achieve better results in both visual experience and quantitative comparison, which attributes to applying different types of specialized convolutions (VCM and PC) to the encoder and decoder of the generator. However, there are still unrecovered or mis-recovered regions in real saturated image. Three solutions can possibly further improve our model. First, training images and real saturated images are normalized regarding to exposure time to bridge the gap between them. Second, EUV image and corresponding magnetogram of photosphere are used jointly as the input of model, which can provide additional information for recovering the lost signal. Third, physical principle of SDO/AIA imaging is integrated with deep learning model to improve robustness of model and reducing risk of overfitting. In addition, the

proposed model can be easily transferred to similar observation devices through transferring learning.

This work was supported by the Peng Cheng Laboratory Cloud Brain (No. PCL2021A13), the National Natural Science Foundation of China (NSFC) under 11790305, 11973058 and 12103064.

References

Cybenko, G. 1989, Math. Control Signals Syst., 2, 303
Gatys, L. A., Ecker, A. S., & Bethge, M. 2016, in IEEE Conf. on Computer Vision and Pattern Recognition (IEEE Computer Society), 2414
Guastavino, S., Piana, M., Massone, A. M., Schwartz, R., & Benvenuto, F. 2019, ApJ, 882, 109
He, K., Zhang, X., Ren, S., & Sun, J. 2015, in IEEE Int. Conf. on Computer Vision (IEEE Computer Society), 1026
Hornik, K., Stinchcombe, M., & White, H. 1989, NN, 2, 359
Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. 2017, in IEEE Conf. on Computer Vision and Pattern Recognition (IEEE Computer Society), 5967
Johnson, J., Alahi, A., & Fei-Fei, L. 2016, in European Conf. on Computer Vision (Berlin: Springer), 694
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Lemen, J. R., Title, A. M., Akin, D. J., et al. 2012, SoPh, 275, 17
Leshno, M., Lin, V. Y., Pinkus, A., & Schocken, S. 1993, NN, 6, 861
Liu, G., Reda, F. A., Shih, K. J., et al. 2018, in European Conf. on Computer Vision (Berlin: Springer), 89
Ma, C., Rao, Y., Cheng, Y., et al. 2020, in IEEE Conf. on Computer Vision and Pattern Recognition (Computer Vision Foundation / IEEE), 7766
Mao, X., Li, Q., Xie, H., et al. 2017, in IEEE Int. Conf. on Computer Vision (IEEE Computer Society), 2813
Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. 2012, SoPh, 275, 3
Schwartz, R., Torre, G., Massone, A., & Piana, M. 2015, A&C, 13, 117
Schwartz, R. A., Torre, G., & Piana, M. 2014, ApJ, 793, L23
Simonyan, K., & Zisserman, A. 2015, arXiv:1409.1556S
Torre, G., Schwartz, R. A., Benvenuto, F., Massone, A. M., & Piana, M. 2015, InvPr, 31, 095006
Wang, N., Zhang, Y., & Zhang, L. 2021, IEEE Trans. Image Process., 30, 1784
Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. 2004, IEEE Trans. Image Process., 13, 600
Yu, X., Xu, L., & Yan, Y. 2021, SoPh, 296, 1
Zhang, P.-J., Wang, C.-B., & Pu, G.-S. 2020, RAA, 20, 204
Zhao, D., Xu, L., Chen, L., Yan, Y., & Duan, L.-Y. 2019, AdAst, 2019, 5343254
Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. 2017, in IEEE Int. Conf. on Computer Vision (IEEE Computer Society), 2242