

Mask-Pix2Pix Network for Over-Exposure Region Recovery of Solar Image

Dong Zhao^{a,b}, Long Xu^{*,a}, Linjie Chen^a, Yihua Yan^a, Ling-Yu Duan^c

^aKey Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

^bUniversity of Chinese Academy of Sciences, Beijing, 100049, China

^cNational Engineering Lab for Video Technology, Peking University, Beijing, 100871, China

Abstract

Over-exposure may happen for imaging of solar observation as extremely violet solar bursts occur, which means that signal intensity goes beyond the dynamic range of imaging system of a telescope, resulting in loss of signal. For example, during solar flare, Atmospheric Imaging Assembly (AIA) of Solar Dynamics Observatory (SDO) often records over-exposed images/videos, resulting loss of fine structures of solar flare. This paper makes effort to retrieve/recover missing information of over-exposure by exploiting deep learning for its powerful nonlinear representation which makes it widely used in image reconstruction/restoration. First, a new model, namely mask-Pix2Pix network, is proposed for over-exposure recovery. It is built on a well-known Pix2Pix network of conditional generative adversarial network (cGAN). In addition, a hybrid loss function, including an adversarial loss, a masked L1 loss and a edge mass loss/smoothness, are integrated together for addressing challenges of over-exposure relative to conventional image restoration. Moreover, a new database of over-exposure is established for training the proposed model. Extensive experimental results demonstrate that the proposed mask-Pix2Pix network can well recover missing information of over-exposure, and outperforms the state-of-the-arts originally designed for image reconstruction tasks.

Key words: Deep learning, generative adversarial network, mask-pix2pix, over-exposure, solar flare, image inpainting

1. Introduction

Solar imaging can provide us more information than one-dimensional solar flux about solar activities, promoting our abilities of probing the mystery of the Sun. Especially, the instruments onboard satellites have more advantages than ground-based instruments, due to good seeing and non-occlusion. During the past few decades, number of satellites carrying solar probing instruments have been launched, including Yohkoh (1992) [1], SOHO (1995) [2][3][4], TRACE (1998) [5], STEREO (2006) [6][7], Hinode (2006) [8], and SDO (2010) [9][10][11][12]. These space instruments can provide us solar observation with unprecedented time cadence, high dynamic range, high spatial/frequency resolution. For example, Atmospheric Imaging Assembly (AIA) instrument onboard SDO can image solar atmosphere in 12 wavelengths every 12 seconds, with spatial resolution of 4096×4096 .

Nowadays, satellite carrying solar instruments can image the complete process of a solar activity with unprecedented high time cadence and high spatial resolution. These recorded images and videos provide the scientists the opportunity for uncovering the nature of solar activities, such as sunspot, filament, coronal loop, flare, or coronal mass emission (CME). However, for some extremely violet solar bursts, the intensity of

signal may go beyond the threshold of a telescope, leading to over-exposure, therefore information is missed in the captured images/videos. Although there were automatic exposure-control algorithms which were implemented in case of flares, controlling of exposure time to reduce over-exposure, for example, AIA avoids over-exposure by alternating between low and long exposures, they cannot guarantee complete avoidance of over-exposure. In addition, the time cadence would increase, so temporal resolution is compromised. For addressing over-exposure, an iterative algorithm by employing the PRiL approximation and EM algorithm is proposed in [13]. Different from [13], we solve this issue via deep learning.

Recently years, taking advantages of the powerful capability of deep learning in its non-linear representation, lots of traditional image processing/reconstruction problems achieved new breakthroughs, such as image noising, super-resolution, image inpainting. Therefore, in this work, we decide to restore over-exposure region (OER) by taking advantage of deep learning. Inspired by image inpainting, our task can be recast into a similar optimization problem of image inpainting. Both of them are proposed to recover missing region in an image using surrounding pixels according to continuity of signal. However, for our task, there are at least three challenges relative to traditional image inpainting. Firstly, the fidelity of signal is required, not just looks alike since the data concerned in our task is used for scientific purpose, for example, computing physical parameters. Secondly, larger OER is concerned in our task comparing to image painting. The last, more irregular shapes of OER results in complicated boundaries connecting OERs and non-

*This work was partially supported by the National Natural Science Foundation of China (NSFC) under Grants 61572461, 11790305, 61811530282, CAS “100-Talents” (Dr. Xu Long)

Email addresses: dzhao@nao.cas.cn (Dong Zhao), lxu@nao.cas.cn (Long Xu), ljchen@nao.cas.cn (Linjie Chen), yyh@nao.cas.cn (Yihua Yan), lingyu@pku.edu.cn (Ling-Yu Duan)

OERs/normal regions. These boundaries should be suppressed to prevent unnatural change of pixel value.

The existing image inpainting approaches can be divided into three classes: diffusion-based, patch-based and learning based. From our experience, neither of diffusion-based and patch-based methods can contribute effective results for OER recovery. The diffusion-based methods [14][15] are highly restricted to locally available information, so they fail to recover meaningful structures of large missing regions, like OERs in our task. Patch-based methods [16][17][18][19] assume that missing region (OER) can find its most similar/relevant patches in the given image. This assumption however does not hold for OERs. The existing learning based inpainting methods are not competent in our task for two reasons. First, some of the learning based methods, such as [20], [21], [22], [23], and [24], trained their networks with fixed shapes and locations of missing regions. While the OERs commonly exhibit irregular shape and random location. Second, despite several studies for irregular missing region in image inpainting, such as [25], [26], [27] and [28], they focus on generating visually coherent completion or producing semantically plausible results. However, our task aim to restore missing OER with high fidelity, not just a visually plausible result.

To address the aforementioned challenges, we propose a learning-based model, namely mask-Pix2Pix network, to estimate the OER of solar image. Our network is built on the Pix2Pix [29], which is a general network for image-to-image translation using a conditional Generative Adversarial Network (cGAN) [30]. We adopt the similar architectures with the Pix2Pix: U-net [31] in generator and PatchGAN [30] in discriminator. The main improvements and contributions of this paper are: **(1)** Unlike the conventional Pix2Pix, our network utilizes the Convolution-SwitchNorm-LReLU/ReLU [32] modules (LReLU for encoder and ReLU for decoder) rather than the Convolution-BatchNorm-ReLU [33] ones. The former (i.e. switchable normalization) can switch between BatchNorm [33], LayerNorm [34] and InstanceNorm [35] by learning their importance weights in an end-to-end manner. The improved architecture of our model boosts the robustness of the network. **(2)** Our objective function contain an adversarial cGAN loss, a masked L1 loss and a edge mask loss/smoothness. The adversarial cGAN loss can capture the full entropy of the conditional distributions they model, and thereby produce highly realistic textures. The masked L1 loss calculates the L1 loss only in masked regions (OERs), enforcing correctness at low frequencies which guarantees restoration of high fidelity for OERs. The edge mask loss is used for smoothing edges of OERs and suppressing edge artifacts in final restored image. **(3)** Additionally, a new database of over-exposure is established for training and testing the proposed mask-Pix2Pix network, where 13700 images are collected from Large-scale Solar Dynamics Observatory image database (LSDO) [36].

2. Background

2.1. Image Inpainting

Image inpainting is the process of reconstructing lost or deteriorated parts of an image. Nowadays, it still remains theoretically and computationally challenging. It is highly ill-posed. Existing image inpainting algorithms can be categorized into three groups, *diffusion-based* [14][15] and *patch-based* [16][17][18][19] and learning based. The former two, typically use differential operators or patch similarity to propagate information from normal regions to missing regions. They work well for stationary textures. However, diffusion-based ones are highly restricted to locally available information, so they usually fail to recover meaningful structures of large missing regions. Patch-based ones, such as Planar Structure Guidance (PSG) [18] and Statistics of Similar Patches (SSP) [19], assume that missing regions can find their most similar/relevant patches from normal regions of the given input image. This assumption does not hold for our task where OER probably does not have similar patches.

Recently, deep learning based approaches have emerged as a promising paradigm for image inpainting. A significant advantage of them over conventional ones lies in their abilities to learn and understand semantics of images with complicated scenes. Context-Encoder (CE) [21] introduces a deep adversarial model to predict plausible structures with combined 2 pixel-wise reconstruction loss and adversarial loss. Although the adversarial loss improves the inpainting quality, the results exhibit blurriness and contain notable artifacts. To overcome the limitations of the CE [21] method, Iizuka et al. [23] propose a novel architecture with three networks: the completion network (using dilated convolutions [20]), the global and local context discriminators. However, the training model with two discriminators is time-consuming, and the output heavily relies on Poisson image blending [37] as a post-processing. Multi-scale neural patch synthesis (MNPS) [22] approach contains a content network to learn the semantics and global structure of the image, and a texture network to generate fine-detailed result by employing a pre-trained VGG-19 [38] network. Liu et al. [25] introduce a partial convolution (PC) model for irregularly shaped hole inpainting with an automatic mask updating step to reduce artifacts. In PC [25] method, the convolution is masked and re-normalized to be conditioned on only valid pixels. Both of the methods of MNPS [22] and PC [25] add the TV loss term to encourage smoothness on the 1-pixel dilation of the hole boundaries. Therefore, they avoid requiring post processing steps (such as Poisson image blending operation) to enforce texture coherency near the hole boundaries.

2.2. Image to Image Translation with Pix2Pix

Pix2Pix [29] utilizes a conditional generative adversarial network (cGAN) [30] to achieve the target of image-to-image translation. Instead of using a conventional encoder-decoder, the generator in Pix2Pix employs an U-Net [31] architecture, in which the encoder layers and decoder layers are directly connected by “skip connection”. Since the skip connection can shuttle the low-level information (which are commonly shared

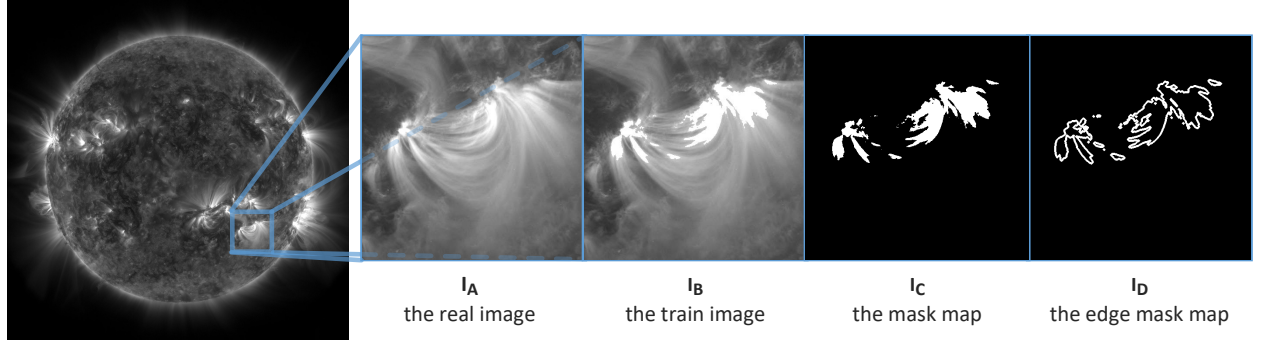


Figure 1: The Over-Exposure Region (OER) database built on the LSDO.

between the input and output images) across the bottleneck of the encoder-decoder net, it effectively improves the performance of the image-to-image translation. The synthesized images of the generator should not be distinguished from the real one by an adversarially trained discriminator. In order to model high-frequencies, the discriminator employs a convolutional PatchGAN classifier, i.e. only the structure at certain scale of patches is penalized. By this method, the PatchGAN is taken as a form of texture loss. To summary, there are at least three factors that enable the Pix2Pix network outperforms past works: the U-Net architecture for the generator, the convolutional PatchGAN classifier for discriminator, and the cGANs for network training.

3. Proposed Method

3.1. Problem Description

The over-exposure problem is simply explained in Fig. 1, where ground truth image I_A , over-exposure image I_B , binary mask map I_C and edge mask map I_D are concerned. The mission of OER recovery is to recover the region labeled by I_C in I_B , and outside region of I_C keeps unchanged. Given OER Ω_M and non-OER $\Omega_{\bar{M}}$, we can get $\Omega_M = I_B \odot I_C$ and $\Omega_{\bar{M}} = I_B \odot (1 - I_C)$, where \odot is the element-wise product operator. Inspired by image inpainting, the neural network, GAN, can be employed to retrieve missing region of I_B (i.e., Ω_M). In a GAN, a generator G is trained on the pairs of ground truth and degraded ones. Then the generator is applied to a degraded image to output a repaired one, i.e., $I_G = G(I_B)$.

To our task, I_G should meet the following criteria:

- The restored image should have realistic textures as far as possible, i.e. visually coherent and semantically plausible relative to I_A ;
- The restored region $I_G \odot I_C$ should have high fidelity relative to corresponding regions in ground truth I_A ;
- It should be avoid artificial edges (I_C) surrounding OERs, i.e. the boundary between OERs and normal regions should transit smoothly.

To achieve a) and b), conventional Pix2Pix network is employed and modified by adding a mask L1 loss term. To address c), a edge mask loss/smoothness term is introduced, where ground-truth edge mask I_D is given by the edge of mask map I_C as shown in Fig. 1.

3.2. Solar OER recovery Database

For training our model, a database for over-exposure is established. The raw data is cropped from LSDO [36]. The LSDO database records consist of three different parts: event records, corresponding images, and extracted image parameters. Event records encompass the list of solar events with generated and extracted attributes. Among these attributes, the *bounding box attribute* consists of four coordinate pixel values. LSDO has pre-processed all original polygon values of event records from their Helioprojective Coordinates (HPC) solar coordinate system into pixels. This preprocessing step is vital to make polygons compatible with the high-resolution images in the database. More information in detail of the LSDO can be referenced from the literature [36].

In this work, ground-truth I_A is cropped from Active Region (AR) of images in LSDO. The database is established as following:

- 1) According to polygon values of AR in LSDO, corresponding AR square regions are cropped from non-over-exposed AIA/SDO images, we get I_A ;
- 2) Scale I_A into the same size of (256×256), and label OERs in I_A by a threshold segmentation: pixel values (from 0 to 1) over a threshold (e.g. 0.8) are set to be 1. Thereby, we get I_B ;
- 3) Label OERs (Ω_M) in I_B as mask map I_C , and dilate I_C with circular kernel (the radius is 3). The dilated mask map I_C is used to calculate the masked L1 loss in subsection 3.3.
- 4) Extract the edge of I_C , with edge width of 5 pixel. This edge mask map I_D accounts for edge mask loss/smoothness in subsection 3.3.

Thereby, each sample in database contains four parts: real image I_A , over-exposed (fake) image I_B , mask map I_C and edge mask map I_D .

Table 1: Architecture of the generator.

Layers	Architecture of Generator	output size
input	I_A, I_B, I_C, I_D	$(256 \times 256 \times 1)$
D1	Conv.($4 \times 4 \times 64$), LReLU	$(128 \times 128 \times 64)$
D2	Conv.($4 \times 4 \times 128$), SwitchNorm, LReLU	$(64 \times 64 \times 128)$
D3	Conv.($4 \times 4 \times 256$), SwitchNorm, LReLU	$(32 \times 32 \times 256)$
D4	Conv.($4 \times 4 \times 512$), SwitchNorm, LReLU , Dropout	$(16 \times 16 \times 512)$
D5	Conv.($4 \times 4 \times 512$), SwitchNorm, LReLU , Dropout	$(8 \times 8 \times 512)$
D6	Conv.($4 \times 4 \times 512$), SwitchNorm, LReLU , Dropout	$(4 \times 4 \times 512)$
D7	Conv.($4 \times 4 \times 512$), SwitchNorm, LReLU , Dropout	$(2 \times 2 \times 512)$
D8	Conv.($4 \times 4 \times 512$), LReLU , Dropout	$(1 \times 1 \times 512)$
U1	Concatenate(D8, D7), DeConv.($4 \times 4 \times 512$), SwitchNorm, ReLU , Dropout	$(2 \times 2 \times 512)$
U2	Concatenate(U1, D6), DeConv.($4 \times 4 \times 512$), SwitchNorm, ReLU , Dropout	$(4 \times 4 \times 512)$
U3	Concatenate(U2, D5), DeConv.($4 \times 4 \times 512$), SwitchNorm, ReLU , Dropout	$(8 \times 8 \times 512)$
U4	Concatenate(U3, D4), DeConv.($4 \times 4 \times 512$), SwitchNorm, ReLU , Dropout	$(16 \times 16 \times 512)$
U5	Concatenate(U4, D3), DeConv.($4 \times 4 \times 256$), SwitchNorm, ReLU	$(32 \times 32 \times 256)$
U6	Concatenate(U5, D2), DeConv.($4 \times 4 \times 128$), SwitchNorm, ReLU	$(64 \times 64 \times 128)$
U7	Concatenate(U6, D1), DeConv.($4 \times 4 \times 64$), SwitchNorm, ReLU	$(128 \times 128 \times 64)$
final	Upsample($4 \times 4 \times 1$), ZeroPad, Conv.($4 \times 4 \times 1$), Tanh	$(256 \times 256 \times 1)$
output	I_G	$(256 \times 256 \times 1)$

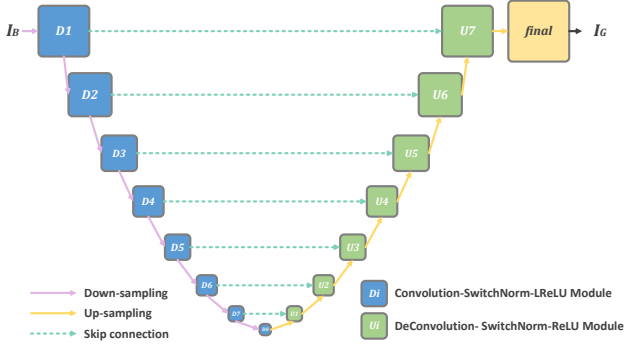


Figure 2: Architecture of the generator.

3.3. Network Architecture

The proposed mask-Pix2Pix network is composed of a discriminator and a generator. The generator employs an U-Net architecture as demonstrated in Fig. 2. The U-Net is named after its shape which looks like a “U”. It consist of an encoder of 8 layers, and a decoder of 8 layers. The parameters of each layer is explained in Table 1 in detail. In addition, skip connections are added between encoder and decoder at the same layer as shown in Fig.2 with dotted lines. Each skip connection simply concatenates feature map of encoder with that of decoder at the same layer (e.g., D_l and $U_{n,l}$ at the l -th layer in Fig.2). This cross-layer connection can reduce semantic gap between corresponding layers of encoder and decoder since they are too far away in U-Net structure. The discriminator is a PatchGAN network, the structure of which is explained in Table 2. In a GAN framework, the discriminator is to judge “fake” instances from “real” ones. In our work, the output of the discriminator is a 16×16 image, each pixel value of which ranges from 0 to 1, for measuring how real is the output. In addition, we adopt Convolution-SwitchNorm-LReLU/ReLU [32]

Table 2: Architecture of the discriminator.

Layers	Architecture of Discriminator	output size
input	$[I_A, I_B]$ or $[I_G^*, I_B]$	$(256 \times 256 \times 2)$
C1	Conv.($4 \times 4 \times 64$), LReLU	$(128 \times 128 \times 64)$
C2	Conv.($4 \times 4 \times 128$), SwitchNorm, LReLU	$(64 \times 64 \times 128)$
C3	Conv.($4 \times 4 \times 256$), SwitchNorm, LReLU	$(32 \times 32 \times 256)$
C4	Conv.($4 \times 4 \times 512$), SwitchNorm, LReLU	$(16 \times 16 \times 512)$
C5	ZeroPad2d, Conv.($4 \times 4 \times 1$)	$(16 \times 16 \times 1)$
output	Real or Fake matrix 16×16	

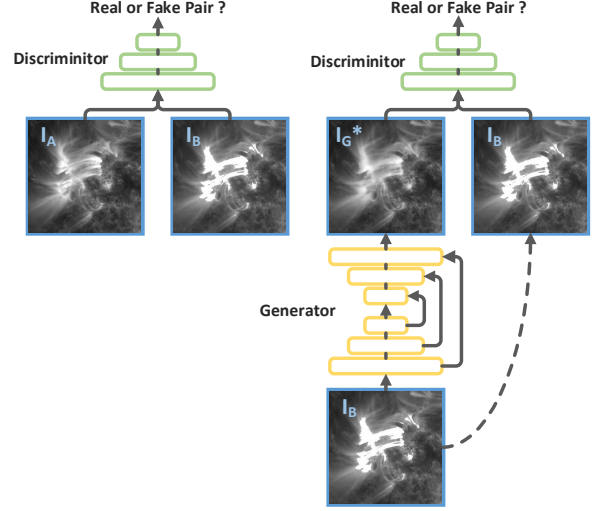


Figure 3: Training on adversarial cGAN loss.

instead of Convolution-BatchNorm-ReLU [33] of conventional Pix2Pix [29] in the proposed model. The former was proved to be more robust.

3.4. Loss Functions

It is extremely vital to design a loss function for modelings by using machine learning. Our task is to recover missing signal of OERs in an image, which mostly concerns fidelity of reconstructed signal, natural transition of pixel value. It has something in common with image inpainting, but not the same. Therefore, a new hybrid loss function is designed for OER recovery task as follows.

Given ground truth image I_A , degraded image I_B , initial binary mask I_C (1 for OERs) and I_C 's edge map I_D , as demonstrated in Fig. 1, the generator outputs $I_G = G(I_B)$. For training model, a hybrid loss function is defined, consisting of three components: 1) adversarial cGAN loss for high fidelity of reconstructed image I_G relative to ground-truth I_A , which is defined as:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}[\log D(I_B, I_A)] + \mathbb{E}[\log(1 - D(I_B, G(I_B)))], \quad (1)$$

where G and D represents the generator and discriminator, respectively, the objective of G is to minimize $D(I_B, G(I_B))$ so that $G(I_B)$ is more like a real instead of fake image, while D pursues the maximum of $D(I_B, I_A)$ to distinguish “fake” and “real” as far as possible. Figure 3 illustrates the process of the adversarial cGAN training, where the discriminator excises two tasks.

The left one is to discriminate if I_A and I_B are a pair of “real” and “fake”. The right other is to discriminate $G(I_B)$ is a “real” or “fake”, where the generator G is trained to produce fakes for deceiving the discriminator D , while D is trained to identify “fake” as far as possible.

2) L1 loss of masked region $I_B \odot I_C$ relative to $I_A \odot I_C$, targeting for accurate reconstruction of OER, which is defined as:

$$\mathcal{L}_1^m(G) = \mathbb{E}[\|(I_A - G(I_B)) \odot I_C\|_1]. \quad (2)$$

3) L1 loss of only edge mask, which could make the edges of OERs smooth to prevent artificial edges connecting OERs and non-OERs. It is defined as

$$\mathcal{L}_1^e(G) = \mathbb{E}[\|(I_A - G(I_B)) \odot I_D\|_1]. \quad (3)$$

In conclusion, the final optimization objective is given by

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda_1 \mathcal{L}_1^m(G) + \lambda_2 \mathcal{L}_1^e(G), \quad (4)$$

where λ_1 and λ_2 are the weights for combining the above three loss components. They are set to 0.1 in our experience.

4. Experimental Results

To evaluate the proposed mask-Pix2Pix model, it is compared with other state-of-the-art image inpainting algorithms. In addition, the contribution of each factor of the loss function to the overall performance is evaluated. These two experiments are carried out on our synthetic over-exposure images. Then, the pre-trained mask-Pix2Pix model are evaluated on real solar images.

4.1. Implementation details

We apply our model to over-exposure training database which are artificially built from LSDO as described in subsection 3.2. For training, the input image size is 256×256 and the mini-batch size is set to 16. We train our model with ADAM optimizer [39] by setting $\beta_1 = 0.5$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. The initial learning rate is initialized to 0.0002, and then reduced to half every 100 epochs. We apply PyTorch on an NVIDIA Tesla P100 for model training.

4.2. Comparisons with state-of-the-art approaches

We compare the proposed Mask-Pix2Pix with a patch-based Planar Structure Guidance (PSG) [18] and a learning-based Pix2Pix network [29]. Figure 4 shows the comparisons on 3 samples, where it can be found that the proposed Mask-Pix2Pix can generate images with significantly better quality than other two benchmarks. Among these three compared methods, PSG [18] is the worst. It can be found that PSG does not work sufficiently for recovering OERs as shown in Fig. 4. The Pix2Pix [29] may underestimate the missing regions, still leaving over-exposure in the center of OERs after reconstruction. We quantitatively measure their performances using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [40]. The PSNR and SSIM are given below each reconstructed image in

Figure 4. The average PSNR and SSIM are also computed on the whole database (over 1600 samples), listed in Table 3. It can be observed that the proposed model outperforms other two benchmarks significantly, achieving up to 5dB relative to the Pix2Pix [29] and 15dB relative to the PSG [18].

4.3. Evaluation on each component of loss function

To evaluate contribution of each loss component, we perform the following three testings:

- 1) the adversarial loss and conventional L1 loss on whole image, i.e. $\mathcal{L}_{cGAN}(G, D) + \lambda_1 \mathcal{L}_1(G)$;
- 2) the adversarial loss and masked L1 loss, i.e. $\mathcal{L}_{cGAN}(G, D) + \lambda_1 \mathcal{L}_1^m(G)$;
- 3) the adversarial loss, masked L1 loss and edge mask loss/smoothness, i.e. $\mathcal{L}_{cGAN}(G, D) + \lambda_1 \mathcal{L}_1^m(G) + \lambda_2 \mathcal{L}_1^e(G)$.

Figure 5 illustrates three groups of experiences with different training losses. We can find that, despite the outputs of $\mathcal{L}_{cGAN} + \lambda_1 \mathcal{L}_1$ have smooth boundary between missing regions and normal regions, it fails to estimate accurate intensity and content of missing regions. Benefiting from the masked loss, $\mathcal{L}_{cGAN} + \lambda_1 \mathcal{L}_1^m$ can obtain better intensity and content estimation for missing regions, however, it yields artificial edges as illustrated in the first and third images in Figure 5. The network trained with the mixed $\mathcal{L}_{cGAN} + \lambda_1 \mathcal{L}_1^m + \lambda_2 \mathcal{L}_1^e$ losses outperforms the other two, successfully addressing all of the three criteria aforementioned in subsection 3.1.

Another improvement of the proposed model lies in the replacement of Convolution-BatchNorm/ReLU [33] in Pix2Pix [29] by Convolution-SwitchNorm-LReLU/ReLU [32], which is witnessed by the second and third rows of Table 3. In Table 3, the second row lists the average PSNR and SSIM of Pix2Pix with Convolution-BatchNorm-ReLU [33], while the third row is for Pix2Pix with Convolution-SwitchNorm-LReLU/ReLU [32]. It can be found that the third row is obviously better than the second row with respect to both PSNR and SSIM, which proves the effectiveness of Convolution-SwitchNorm-LReLU/ReLU.

4.4. Discussions and testing on real solar image

In real AIA image, over-exposure region is related to two effects: primary saturation (PS) and secondary saturation (SS) [13]. PS refers to the fact that, for intense incoming flux, CCD pixels lose their ability to accommodate additional charge. SS names the fact that PS causes charge to spill into their neighbors. Furthermore, over-exposure often accompany with diffraction, creating a star-like pattern in the images. However, there are no ground truth images paired with the over-exposure ones. Therefore, we train the mask-Pix2Pix model on synthesized database, where the saturated images cropped from the LSDO.

We further test our pre-trained mask-Pix2Pix model on the flare image of Sep. 10, 2017 X8.2. The results are illustrated in Fig. 7. The test result shows that, the pre-trained model recovers the intensity to some extent while fails to recover the fine structure of flare.

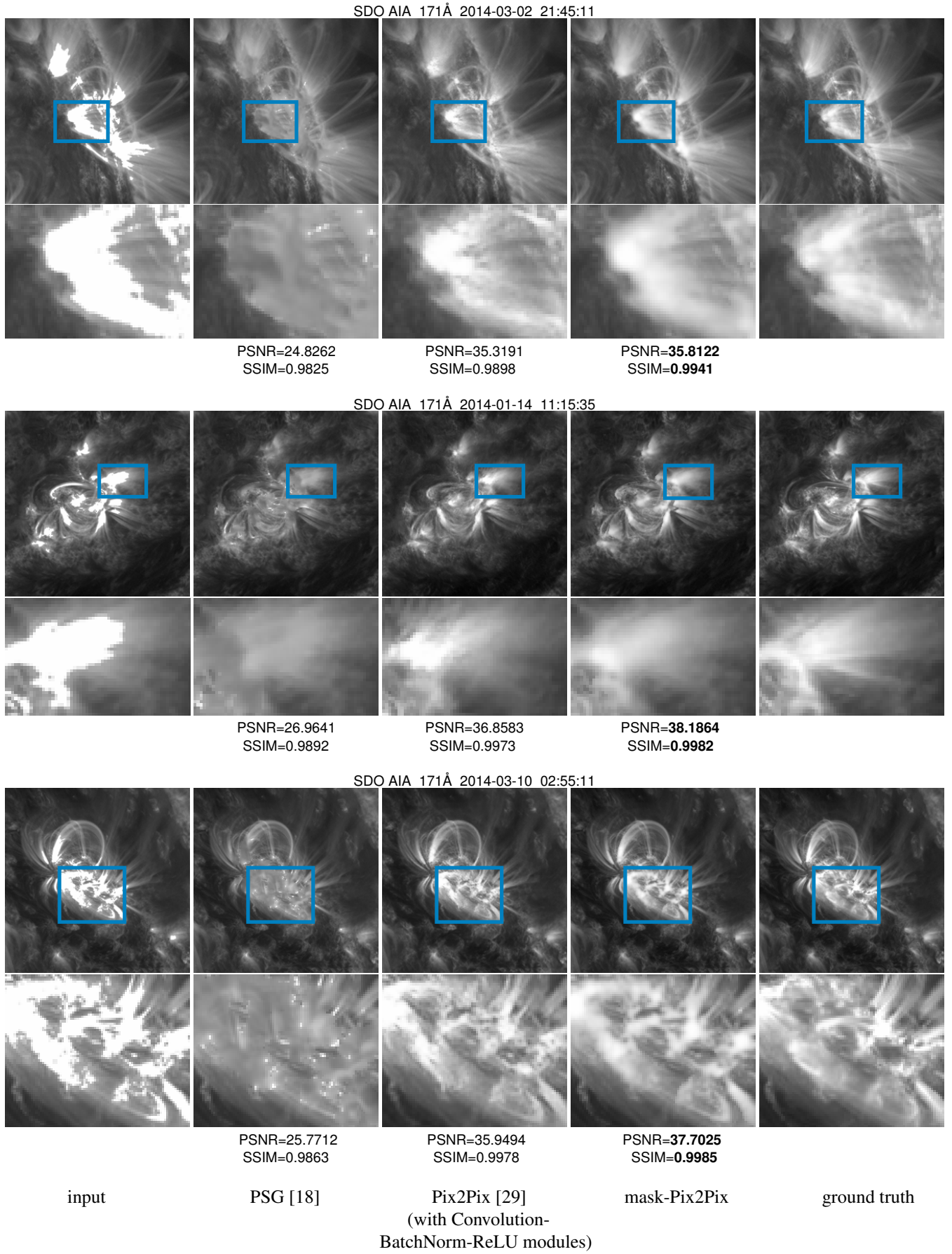


Figure 4: Comparisons between the proposed one and two state-of-the-art methods.

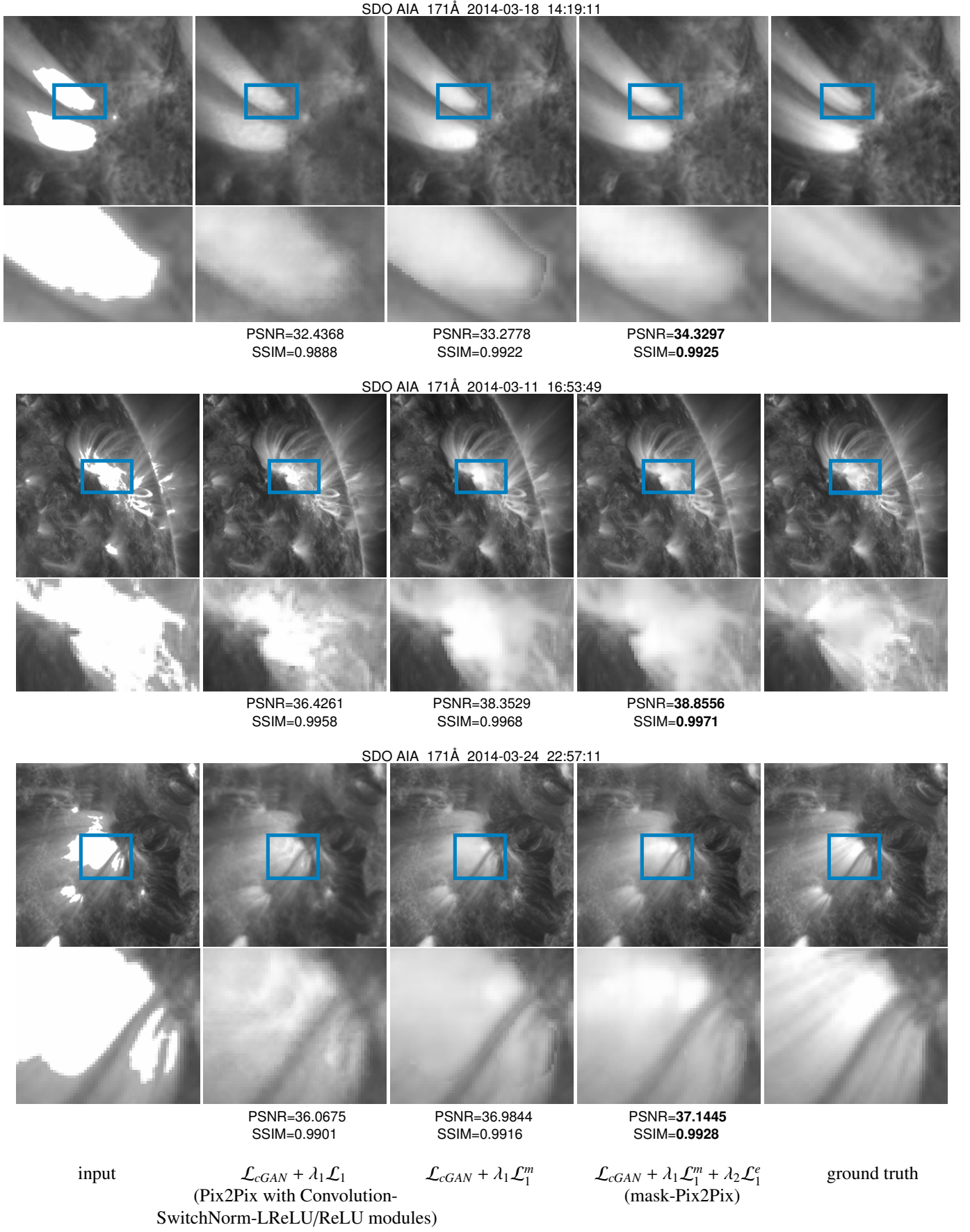


Figure 5: Evaluation on the hybrid loss function (contribution of each component is evaluated).

Table 3: Average PSNR and SSIM comparisons between our model and other two benchmarks.

Methods	Loss functions	Modules	Average PSNR	Average SSIM
PSG [18]			24.6219	0.9810
Pix2Pix [29]	$\mathcal{L}_{cGAN} + \lambda_1 \mathcal{L}_1^{[a]}$	Convolution-BatchNorm-ReLU	34.7763	0.9891
Pix2Pix	$\mathcal{L}_{cGAN} + \lambda_1 \mathcal{L}_1^{[a]}$	Convolution-SwitchNorm-LReLU/ReLU	35.9939	0.9918
mask Pix2Pix	$\mathcal{L}_{cGAN} + \lambda_1 \mathcal{L}_1^m [a]$	Convolution-SwitchNorm-LReLU/ReLU	37.4334	0.9937
mask Pix2Pix	$\mathcal{L}_{cGAN} + \lambda_1 \mathcal{L}_1^m + \lambda_2 \mathcal{L}_1^e [b]$	Convolution-SwitchNorm-LReLU/ReLU	39.6931	0.9985

^[a] $\lambda_1 = 0.1$.

^[b] $\lambda_1 = \lambda_2 = 0.1$.

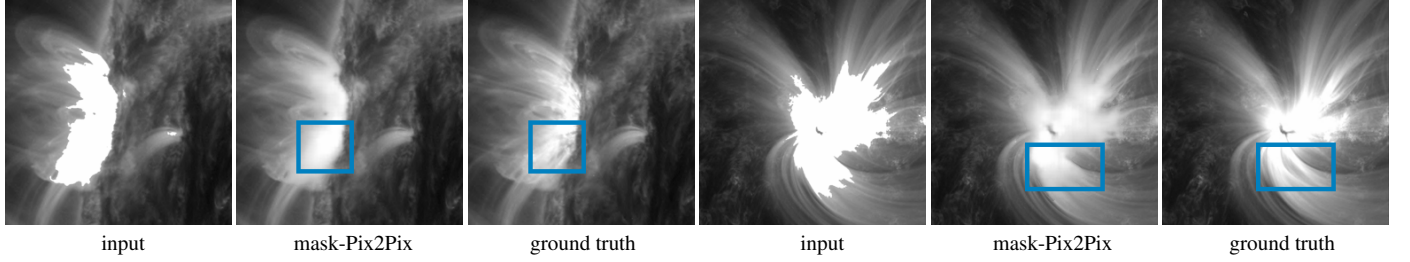


Figure 6: Fail examples of our model.

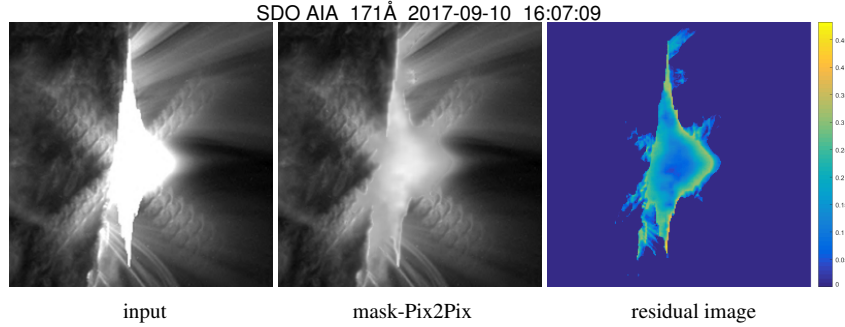


Figure 7: Testing on the real image.

5. Conclusions

We propose a mask-Pix2Pix network with a hybrid loss function to optimize the recovery of OERs in solar images. The proposed model can robustly handle all kinds of OERs regarding shape, size and location in an image. In addition, the evaluations on each component of loss function indicate that the two incremental components beyond conventional cGAN have important contributions to the proposed model. Moreover, the investigation on Convolution-SwitchNorm-LReLU/ReLU also demonstrates that better performance can be achieved comparing to conventional Convolution-BatchNorm-ReLU. However, one limitation of our mode is that it may fail on highly textured areas, and large OERs, as illustrated in Figure 6, which will be further investigated in our future work.

References

- [1] K. Shibata, Y. Ishido, L. W. Acton, K. T. Strong, T. Hirayama, Y. Uchida, A. H. McAllister, R. Matsumoto, S. Tsuneta, T. Shimizu, et al., Observations of x-ray jets with the yohkoh soft x-ray telescope, Publications of the Astronomical Society of Japan 44 (1992) L173–L179.
- [2] F. Clette, J. P. Delaboudiniere, K. P. Dere, P. Cugnon, E. S. Team, EIT: The extreme ultraviolet imaging telescope, 1995.
- [3] J.-P. Delaboudiniere, G. Artzner, J. Brunaud, A. H. Gabriel, J. Hochedez, F. Millier, X. Song, B. Au, K. Dere, R. Howard, et al., Eit: extreme-ultraviolet imaging telescope for the soho mission, in: The SOHO Mission, Springer, 1995, pp. 291–312.
- [4] V. Domingo, B. Fleck, A. I. Poland, The soho mission: an overview, Solar Physics 162 (1-2) (1995) 1–37.
- [5] B. Handy, L. Acton, C. Kankelborg, C. Wolfson, D. Akin, M. Bruner, R. Carvalho, R. Catura, R. Chevalier, D. Duncan, et al., The transition region and coronal explorer, Solar Physics 187 (2) (1999) 229–260.
- [6] R. Howard, J. Moses, A. Vourlidas, J. Newmark, D. Socker, S. Plunkett, C. Korendyke, J. W. Cook, A. Hurley, J. Davila, W. T. Thompson, O. St. Cyr, E. Mentzell, K. Mehalick, J. R. Lemen, J. P. Wuelser, D. W. Duncan, T. Tarbell, C. J. Wolfson, T. Carter, Sun earth connection coronal and heliospheric investigation (secchi), Space Science Review, v.136, 67-115 (2008) 136. doi:10.1007/s11214-008-9341-4.
- [7] M. L. Kaiser, T. Kucera, J. Davila, O. S. Cyr, M. Guhathakurta, E. Christian, The stereo mission: An introduction, Space Science Reviews 136 (1-4) (2008) 5–16.
- [8] T. Kosugi, K. Matsuzaki, T. Sakao, T. Shimizu, Y. Sone, S. Tachikawa, T. Hashimoto, K. Minesugi, A. Ohnishi, T. Yamada, et al., The hinode

- (solar-b) mission: an overview, in: *The Hinode Mission*, Springer, 2007, pp. 5–19.
- [9] J. R. Lemen, D. J. Akin, P. F. Boerner, C. Chou, J. F. Drake, D. W. Duncan, C. G. Edwards, F. M. Friedlaender, G. F. Heyman, N. E. Hurlburt, et al., The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo), in: *The Solar Dynamics Observatory*, Springer, 2011, pp. 17–40.
- [10] W. D. Pesnell, B. J. Thompson, P. C. Chamberlin, The solar dynamics observatory (sdo), in: *The Solar Dynamics Observatory*, Springer, 2011, pp. 3–15.
- [11] P. Boerner, C. Edwards, J. Lemen, A. Rausch, C. Schrijver, R. Shine, L. Shing, R. Stern, T. Tarbell, C. J. Wolfson, et al., Initial calibration of the atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo), in: *The Solar Dynamics Observatory*, Springer, 2011, pp. 41–66.
- [12] P. H. Scherrer, J. Schou, R. Bush, A. Kosovichev, R. Bogart, J. Hoeksema, Y. Liu, T. Duvall, J. Zhao, C. Schrijver, et al., The helioseismic and magnetic imager (hmi) investigation for the solar dynamics observatory (sdo), *Solar Physics* 275 (1-2) (2012) 207–227.
- [13] S. Guastavino, M. Piana, A. M. Massone, R. Schwartz, F. Benvenuto, Desaturating euv observations of solar flaring storms, *arXiv preprint arXiv:1904.04211*.
- [14] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, J. Verdera, Filling-in by joint interpolation of vector fields and gray levels.
- [15] S. Esedoglu, J. Shen, Digital inpainting based on the mumford–shah–euler image model, *European Journal of Applied Mathematics* 13 (4) (2002) 353–370.
- [16] C. Barnes, E. Shechtman, A. Finkelstein, D. B. Goldman, Patchmatch: A randomized correspondence algorithm for structural image editing, in: *ACM Transactions on Graphics (ToG)*, Vol. 28, ACM, 2009, p. 24.
- [17] S. Darabi, E. Shechtman, C. Barnes, D. B. Goldman, P. Sen, Image melding: Combining inconsistent images using patch-based synthesis., *ACM Trans. Graph.* 31 (4) (2012) 82–1.
- [18] J.-B. Huang, S. B. Kang, N. Ahuja, J. Kopf, Image completion using planar structure guidance, *ACM Transactions on graphics (TOG)* 33 (4) (2014) 129.
- [19] K. He, J. Sun, Image completion approaches using the statistics of similar patches, *IEEE transactions on pattern analysis and machine intelligence* 36 (12) (2014) 2423–2435.
- [20] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, *arXiv preprint arXiv:1511.07122*.
- [21] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, A. A. Efros, Context encoders: Feature learning by inpainting, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2536–2544.
- [22] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, H. Li, High-resolution image inpainting using multi-scale neural patch synthesis, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6721–6729.
- [23] S. Iizuka, E. Simo-Serra, H. Ishikawa, Globally and locally consistent image completion, *ACM Transactions on Graphics (ToG)* 36 (4) (2017) 107.
- [24] Y. Song, C. Yang, Z. Lin, X. Liu, Q. Huang, H. Li, C.-C. Jay Kuo, Contextual-based image inpainting: Infer, match, and translate, in: *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.
- [25] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, B. Catanzaro, Image inpainting for irregular holes using partial convolutions, in: *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 85–100.
- [26] Z. Yan, X. Li, M. Li, W. Zuo, S. Shan, Shift-net: Image inpainting via deep feature rearrangement, in: *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 1–17.
- [27] Q. Xiao, G. Li, Q. Chen, Deep inception generative network for cognitive image inpainting, *arXiv preprint arXiv:1812.01458*.
- [28] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, M. Ebrahimi, Edgeconnect: Generative image inpainting with adversarial edge learning, *arXiv preprint arXiv:1901.00212*.
- [29] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [30] M. Mirza, S. Osindero, Conditional generative adversarial nets, *arXiv preprint arXiv:1411.1784*.
- [31] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [32] P. Luo, J. Ren, Z. Peng, Differentiable learning-to-normalize via switchable normalization, *arXiv preprint arXiv:1806.10779*.
- [33] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv preprint arXiv:1502.03167*.
- [34] J. Lei Ba, J. R. Kiros, G. E. Hinton, Layer normalization, *arXiv preprint arXiv:1607.06450*.
- [35] D. Ulyanov, A. Vedaldi, V. Lempitsky, Instance normalization: The missing ingredient for fast stylization, *arXiv preprint arXiv:1607.08022*.
- [36] A. Kucuk, J. M. Banda, R. A. Angryk, A large-scale solar dynamics observatory image dataset for computer vision applications, *Scientific data* 4 (2017) 170096.
- [37] P. Rez, M. Gangnet, A. Blake, Poisson image editing., *Acm Transactions on Graphics* 22 (3) (2003) 313–318.
- [38] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*.
- [39] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, *international conference on learning representations*.
- [40] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, et al., Image quality assessment: from error visibility to structural similarity, *IEEE transactions on image processing* 13 (4) (2004) 600–612.