



De-noising *SDO*/HMI Solar Magnetograms by Image Translation Method Based on Deep Learning

Eunsu Park^{1,2} , Yong-Jae Moon^{1,2} , Daye Lim^{1,2} , and Harim Lee^{1,2}

¹ Department of Astronomy and Space Science, College of Applied Science, Kyung Hee University, Yongin, 17104, Republic of Korea; moonjy@khu.ac.kr

² School of Space Research, Kyung Hee University, Yongin, 17104, Republic of Korea

Received 2019 October 18; revised 2020 February 7; accepted 2020 February 9; published 2020 February 26

Abstract

In astronomy, long-exposure observations are one of the important ways to improve signal-to-noise ratios (S/Ns). In this Letter, we apply a deep-learning model to de-noise solar magnetograms. This model is based on a deep convolutional generative adversarial network with a conditional loss for image-to-image translation from a single magnetogram (input) to a stacked magnetogram (target). For the input magnetogram, we use *Solar Dynamics Observatory* (*SDO*)/Helioseismic and Magnetic Imager (HMI) line-of-sight magnetograms at the center of the solar disk. For the target magnetogram, we make 21-frame-stacked magnetograms, taking into account solar rotation at the same position. We train a model using 7004 pairs of the input and target magnetograms from 2013 January to 2013 October. We then validate the model using 707 pairs from 2013 November and test the model using 736 pairs from 2013 December. Our results from this study are as follows. First, our model successfully de-noises *SDO*/HMI magnetograms, and the de-noised magnetograms from our model are mostly consistent with the target magnetograms. Second, the average noise level of the de-noised magnetograms is greatly reduced from 8.66 to 3.21 G, and it is consistent with that of the target magnetograms, 3.21 G. Third, the average pixel-to-pixel correlation coefficient value increases from 0.88 (input) to 0.94 (de-noised), which means that the de-noised magnetograms are more consistent with the target ones than the input ones. Our results can be applied to many scientific fields in which the integration of many frames (or long-exposure observations) are used to improve the S/N.

Unified Astronomy Thesaurus concepts: [Solar magnetic fields \(1503\)](#); [The Sun \(1693\)](#); [Astronomy data analysis \(1858\)](#)

1. Introduction

The simplest way to reduce noise is to take longer exposures (or stack) observations. The signal-to-noise ratio (S/N) is defined as the ratio of signal to noise. Photon noise increases in proportion to the square root of time, and signal increases in proportion to time. Thus, high S/Ns can be obtained by increasing the exposure time or stacking observations (Birney et al. 2006). For example, the *Hubble Space Telescope* has generated ultra-deep field images by using long exposure times of up to several days (Beckwith et al. 2006). However, this method has the disadvantage of taking a lot of time to reduce noise from observed data.

There have been a few attempts to reduce the noise levels of solar magnetograms. Wang et al. (1995) and Chae et al. (2001) examined weak and small magnetic field structures such as solar intranetwork or small bipoles by integrating 4096 frames of magnetograms taken at Big Bear Solar observatory. Schrijver et al. (1997) stacked Michelson Doppler Imager (MDI) magnetograms to obtain a sequence of high-resolution magnetograms and to determine the polarity pattern on solar quiet region more clearly. DeForest (2017) presented a set of algorithms based on locally adaptive filters to reduce noise in astrophysical images including solar magnetograms.

The *Solar Dynamics Observatory* (*SDO*; Pesnell et al. 2012) is a spacecraft mission with three instruments that investigates how solar magnetic field is generated and structured, and how this stored magnetic energy is released into the heliosphere and geospace as solar wind, energetic particles, and variations in the solar irradiance (Pesnell et al. 2012). The Helioseismic and Magnetic Imager (HMI; Scherrer et al. 2012; Schou et al. 2012)

is an onboard instrument of the *SDO*, and is designed to measure the Doppler shift, intensity, and magnetic field at the solar photosphere; it has continuously observed the data near solar surface since 2010 (Scherrer et al. 2012; Schou et al. 2012). The HMI magnetogram is acquired from multiple narrow spectral bands. This instrument has two charge-coupled device (CCD) cameras: one is the “front camera” that observes 45 s cadence line-of-sight (LOS) magnetograms, and the other is the “side camera” that observes 720 s cadence vector magnetograms. Liu et al. (2012) calculated the average noise level of *SDO*/HMI magnetograms by assuming that the noise level in a magnetogram is the standard deviation of a best-fitted Gaussian fitting function of the histogram of magnetic flux densities. They reported that the average noise level of *SDO*/HMI 45 s magnetograms is about 10.2 G and that of 720 s magnetograms is about 6.3 G.

Recently, a deep neural network (DNN; Lecun et al. 2015) and machine-learning algorithm called “Deep Learning” has been developed. DNN is a kind of artificial neural network that has been developed to learn how humans think and recognize an object using their deep hierarchical layer structures. The convolutional neural network (CNN; Lecun et al. 1998) is the most popular deep-learning method in the field of image processing and computer vision. In general, CNN models consist of convolution filters that extract features from their data sets. The generative adversarial network (GAN; Goodfellow et al. 2014) is another popular deep-learning method used for several generative tasks. Generally, a GAN consists of two networks: one is generative network (generator) and the other is discriminative network (discriminator). The purpose of the generator is to generate realistic fake data, and

the purpose of the discriminator is to distinguish fake data from real data. Their purposes are adversarial to each other, which gives the appearance of competition between the two networks. Based on this adversarial training, we expect the generator to produce a substantial amount of realistic data that the discriminator cannot distinguish, which is the optimal state for the generator. The deep convolutional generative adversarial network (DCGAN; Radford et al. 2015) is a combined model of CNN and GAN used for stable training and output. Based on DCGAN, several methods have been suggested to solve the various types of image-generation tasks (e.g., Isola et al. 2016; Ledig et al. 2016). Kim et al. (2019) suggested a DCGAN model to generate solar magnetograms using *SDO*/AIA images, and then applied the model to *Solar Terrestrial Relations Observatory*/Extreme Ultraviolet Imager (*STEREO*/EUVI) images to produce solar farside magnetograms. Park et al. (2019) applied both CNN and DCGAN models to the generation of solar ultraviolet (UV) images from *SDO*/HMI LOS magnetograms, then compared the results from two models.

There have been a few attempts to apply deep learning to the de-noising of solar data. Díaz Baso et al. (2019) developed a CNN model to de-noise solar data and applied it to pairs of synthetic magnetograms from simulations with and without noise. They also applied their model to pairs of magnetograms and their deconvolved ones by Swedish 1 meter Solar Telescope. From both applications they obtained de-noised magnetograms with much less noise.

In this Letter, we apply a DCGAN model to the de-noising of solar magnetograms using real observation data sets. For the data sets, we make pairs of original *SDO*/HMI LOS magnetograms the input, and 21-frame stacked ones the target; the stacked ones have much lower noise levels. The model outputs are de-noised magnetograms and they are compared with the target magnetograms. This Letter is organized as follows. The data will be described in Section 2, and the model in Section 3. Results are given in Section 4, and a brief summary is presented in Section 5.

2. Data

We use *SDO*/HMI LOS 45 s magnetograms from 2013 January to 2013 December. For the input magnetogram, we select a patch at the center of solar disk with size of 256 by 256 (about $\pm 76''8$). For the target magnetogram, we integrate 21 magnetograms that include 10 frames before and 10 frames after the input magnetogram considering solar rotation. The stacked magnetogram has an approximately 15 minute exposure time. We use magnetograms with a range of -100 G for minimum and 100 G for maximum. As a result, we make 8447 pairs of input and target magnetograms with a 1 hr cadence. Then we separate our data sets into training, validation, and test in chronological order. We select 707 pairs from 2013 November for the validation data set, 736 pairs from 2013 December for the test data set, and the remaining 7004 pairs for training data set.

3. Method

Our model is based on the model by Park et al. (2019). They modified the model of Isola et al. (2016), who suggested a general purposed solution based on a conditional generative adversarial network (cGAN; Mirza & Osindero 2014) and

DCGAN to resolve the image-to-image translation problems. Several authors suggested that the results from DCGAN models could be more realistic than those from CNN models (Isola et al. 2016; Ledig et al. 2016). Park et al. (2019) also reported that the generated solar UV images from DCGAN models are clearer than those from CNN models in most passbands, so we follow them. More details about our model and codes are available at our GitHub repository.³

Figure 1 shows the main structure of our model based on DCGAN. The purpose of the generator (G) is to generate target-like magnetograms (de-noised, hereafter) using input magnetograms. The purpose of the discriminator (D) is to distinguish pairs of the input magnetograms and target ones, called “Real Pair,” and pairs of the input ones and de-noised ones, called “Fake Pair.” To train our model, we use two loss functions: one is L1 loss (mean absolute error, L_1), which is given by

$$L_1(G) = \frac{1}{N} \sum_i^N |M_i^T - M_i^D|, \quad (1)$$

where i is a pixel number, M^I , M^T , and M^D are the input, target, and de-noised magnetograms, respectively. The generator tries to minimize the L_1 , which means that the generator trains itself to minimize the difference between the M^T , and M^D . The other is cGAN loss (L_{cGAN}), which is given by

$$L_{\text{cGAN}}(G, D) = \log(D(M^I, M^T)) + \log(1 - D(M^I, M^D)), \quad (2)$$

where G is the generator, D is the discriminator, $D(M^I, M^T)$ is the probability calculated by the discriminator using the Real Pair, and $D(M^I, M^D)$ is the probability calculated by the discriminator using the Fake Pair. The discriminator tries to maximize the L_{cGAN} to well distinguish between the Real Pair and the Fake Pair. On the other hand, the generator tries to minimize the L_{cGAN} to make the discriminator difficult to distinguish between the Real Pair and the Fake Pair. We expect the L_{cGAN} contributes to generating realistic de-noised magnetograms. The final loss function is given by

$$G^* = \operatorname{argmin}_G \max_D L_{\text{cGAN}}(G, D) + \lambda L_1(G), \quad (3)$$

where λ is the relative weight of the L_{cGAN} and the L_1 . In this work, we used 100 for the relative weight, like Isola et al. (2016). To minimize or maximize the losses, we use the adaptive momentum estimation solver (Kingma & Ba 2014) as an optimizer for both the discriminator and the generator. We save the generator in every 10,000 iterations, so we acquire 50 generator networks while the generator and the discriminator are alternatively trained for 500,000 iterations. Here one iteration refers to when one pair of images is trained in our model. In the validation step, we compare the target magnetograms with the de-noised ones by the 50 generators using the validation data set, and then we select the best model among the saved 50 generators. In the test step, we estimate the model performances of the selected generator in the validation step.

³ https://github.com/eunsu-park/solar_magnetogram_denoising

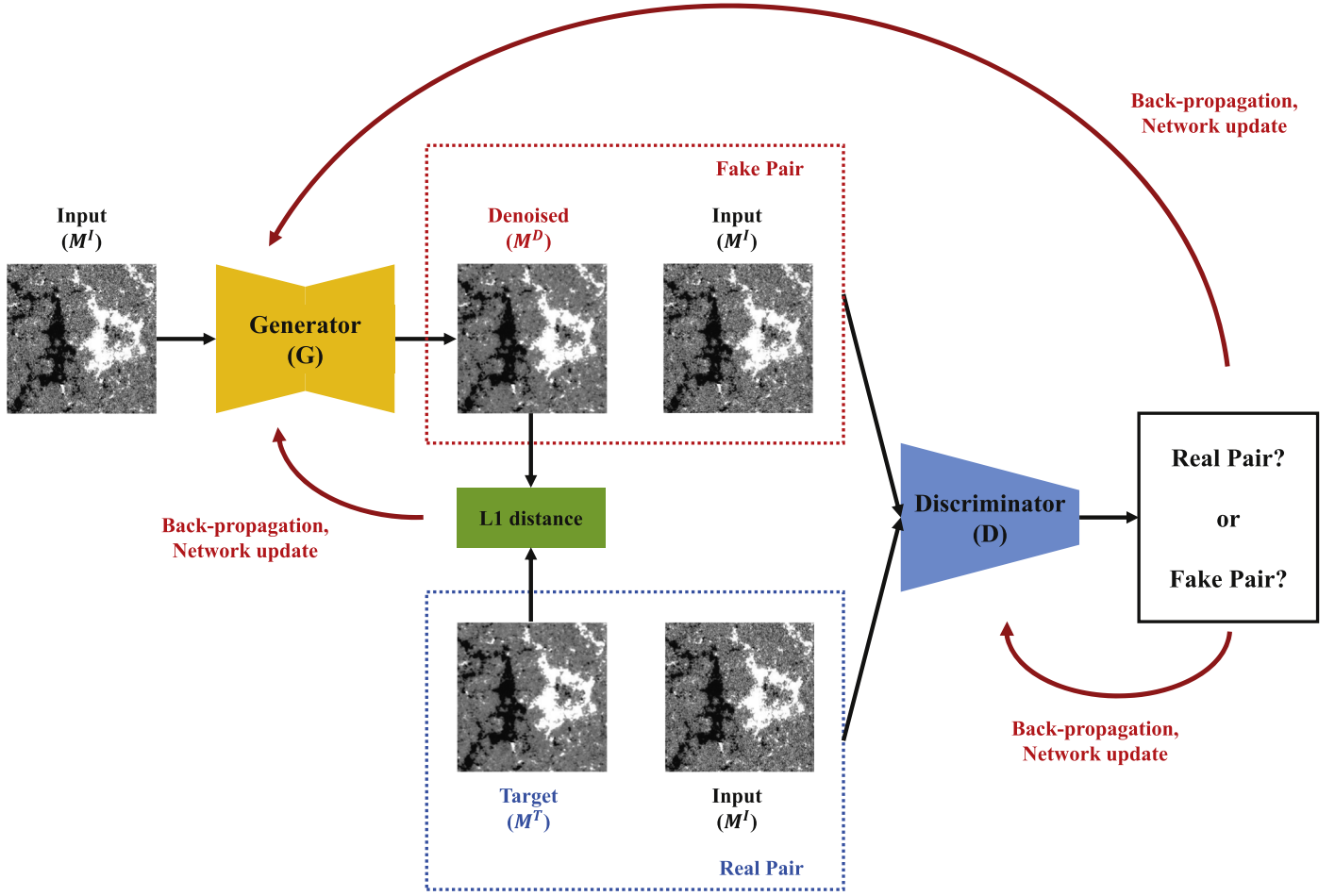


Figure 1. Flowchart and structures of our proposed model. G is the generator, D is the discriminator, M^I is an input magnetogram, M^T is a target magnetogram, and M^D is a de-noised magnetogram by the generator. The blue box is a Real Pair (M^I , M^T), and the red box is a Fake Pair (M^I , M^D).

4. Results and Discussion

Figure 2 shows input, target, and de-noised magnetograms (output), and the difference between target and de-noised magnetograms for three specific regions. As shown in Figure 2, the input magnetograms are quite noisy but the target and de-noised ones are much less noisy. Impressively, the de-noised magnetograms are quite consistent with the target ones, which is also evident in the difference image between them.

To test our results, we calculate the noise levels of the input, target, and de-noised magnetograms by applying a Gaussian fitting to the histogram of magnetic flux densities. We assume the standard deviation of the Gaussian fitting as the noise level of magnetograms (Liu et al. 2004, 2012). Figure 3 shows histograms of magnetic flux densities for the input, target, and de-noised magnetograms in Figure 2. As shown in Figure 3, the histograms for the de-noised magnetograms are similar to those of the target ones, and their noise levels are almost the same. Table 1 shows the average values of the noise levels for validation and test sets. Our model significantly reduces the average noise level from 8.66 to 3.21 G, which is comparable to that of target magnetograms, 3.21 G.

In addition to the noise levels, we calculate five types of metrics between target magnetograms and de-noised (or input) ones. The first metric is the pixel CC (higher is better). The second is the RE (smaller is better) of the total unsigned

magnetic flux (TUMF, Φ_i), which is given by

$$RE_i = (\Phi_i^{\text{Denoised}} - \Phi_i^{\text{Target}}) / \Phi_i^{\text{Target}}, \quad (4)$$

where i is a serial number of test samples. This value corresponds to the overestimation ($RE_i > 0$) or underestimation ($RE_i < 0$) that our method attributes to the TUMF. The third metric is the linear fitting of TUMF between the target magnetograms (Φ^T) and the de-noised ones (Φ^D), which is given by $\Phi^D = A\Phi^T + B$. The fourth metric is the NMSE (smaller is better) of the magnetic field (B_j) given by

$$NMSE_i = \sum (B_j^{\text{Denoised}} - B_j^{\text{Target}})^2 / \sum (B_j^{\text{Target}})^2, \quad (5)$$

where i is a serial number of test samples and j is a pixel number. The last metric is the peak S/N (higher is less noisy), which is used as a quality measurement between a original image and a compressed image, given by

$$\text{peakS/N}_i = 20 \log_{10} \left(\frac{\text{MAX}_i}{\sqrt{\text{MSE}_i}} \right), \quad (6)$$

where i is a serial number of test samples, MAX_i is the length of data range, and MSE_i is the mean squared error between target and input (or de-noised) magnetograms. The peak S/N value becomes higher when an image is less noisy and becomes zero when an image has no noise.

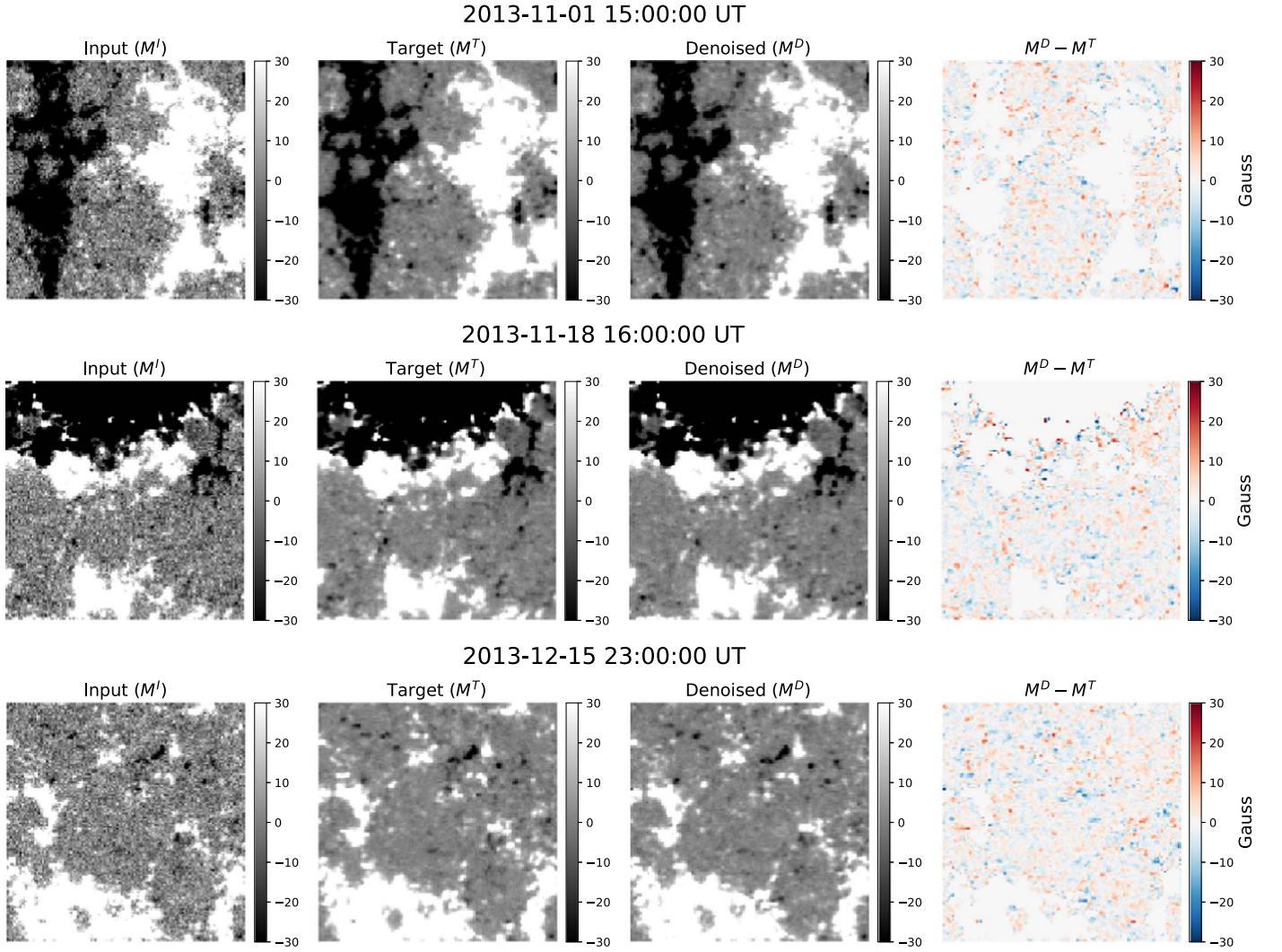


Figure 2. Three examples of comparisons between input, target, and de-noised magnetograms. The first column represents input magnetograms, the second column represents target magnetograms, the third column represents de-noised magnetograms from our model, and the last column represents the difference maps between de-noised and target magnetograms.

Table 1

The Average Noise Levels, Pixel-to-pixel Correlation Coefficient (Pixel CC), Relative Error (RE) of the Total Unsigned Magnetic Flux (TUMF), Linear Fitting of TUMF, Normalized Mean Squared Error (NMSE), and Peak S/N of Validation and Test Data Sets

		Input	Target	De-noised (ours)	Median	Gaussian	Bilateral
Noise Level	Validation	8.74	3.24	3.24	4.61	4.30	4.41
	Test	8.66	3.21	3.21	4.57	4.27	4.36
Pixel CC	Validation	0.88	1	0.94	0.93	0.95	0.94
	Test	0.88	1	0.94	0.93	0.95	0.94
RE	Validation	0.515	0	0.001	0.013	0.041	0.053
	Test	0.529	0	0.001	0.012	0.043	0.053
Linear Fitting (1e20 Mx)	Validation	$0.92\Phi^T + 4.55$	Φ^T	$0.99\Phi^T + 0.08$	$0.98\Phi^T + 0.26$	$0.97\Phi^T + 0.56$	$0.97\Phi^T + 0.60$
	Test	$0.90\Phi^T + 4.66$	Φ^T	$0.99\Phi^T + 0.08$	$0.96\Phi^T + 0.36$	$0.96\Phi^T + 0.58$	$0.96\Phi^T + 0.67$
NMSE	Validation	0.31 (0.23)	0	0.12 (0.07)	0.13 (0.09)	0.09 (0.06)	0.11 (0.08)
	Test	0.31 (0.23)	0	0.12 (0.07)	0.13 (0.09)	0.09 (0.06)	0.12 (0.08)
Peak S/N	Validation	28.49	100	32.53	32.03	33.36	32.59
	Test	28.53	100	32.62	32.17	33.53	32.72

Note. The value in between parentheses corresponds to NMSE for pixels larger than the noise level.

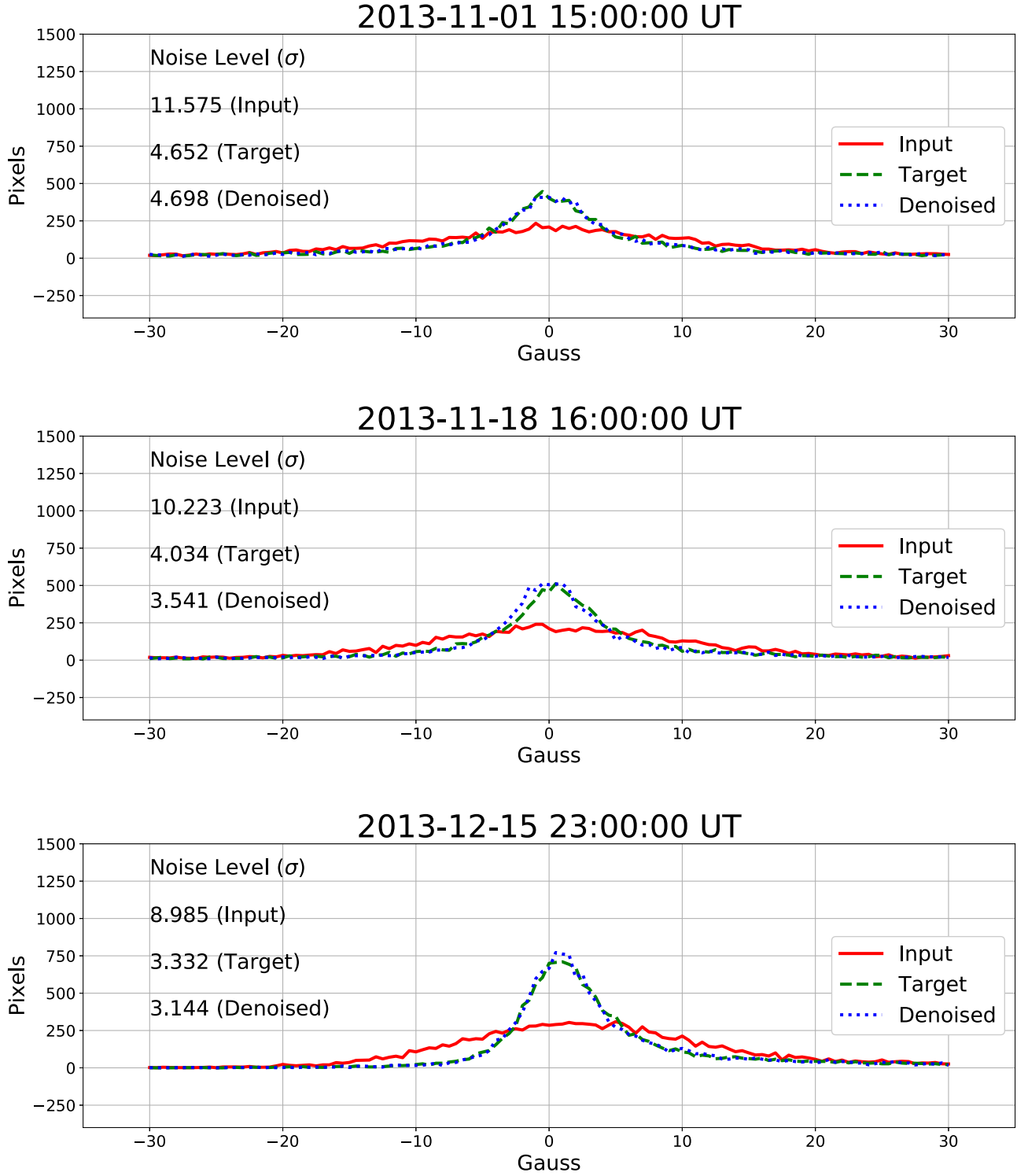


Figure 3. Three examples of histograms of magnetic flux densities from input, target, and de-noised magnetograms. The red lines represent the histograms of input magnetograms, the green lines represent the histograms of target magnetograms, and the blue lines represent the histograms of de-noised magnetograms from our model. The dates and times are the same as those of Figure 2.

Table 1 shows the average values of the five metrics between target magnetograms and de-noised (input) ones. The average CC value increases from 0.88 (input) to 0.94 (de-noised), which means that the de-noised magnetograms are more consistent with the target ones than the input ones. The average RE value of the de-noised magnetograms greatly decreases from 0.529 to 0.001. In view of RE, our model slightly

overestimates the TUMF but the error is quite small, about 0.1%. The linear fitting of TUMF between the target magnetograms and the de-noised ones is very close to the perfect line. The average NMSE value decreases from 0.31 to 0.12. The NMSE is sensitive to pixels having small magnetic flux densities. The average NMSE value, when we consider only the area higher than the noise level, is 0.07, which

HMI magnetogram 2017-09-05 00:00 UT

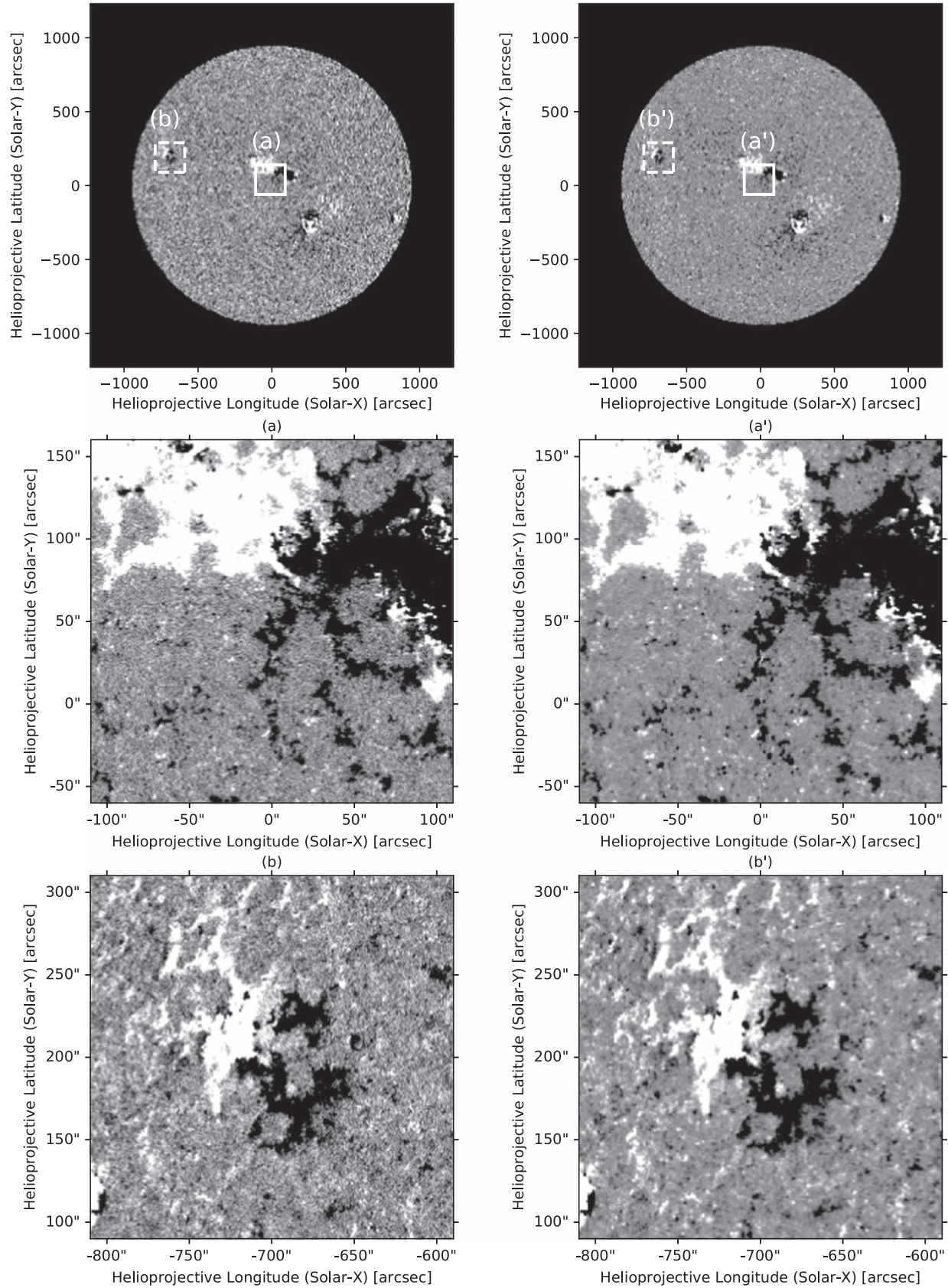


Figure 4. Application of our model to a full-disk *SDO*/HMI magnetogram at 2017 September 5 00:00 UT. The first column represents original *SDO*/HMI magnetograms, and the second column represents de-noised magnetograms from our model. The noise levels of (a), (a'), (b), and (b') are 10.20, 4.05, 11.79, and 4.91 G, respectively.

becomes much smaller. The average Peak S/N value between the target magnetograms and the de-noised ones is 32.62 dB, while the value between the target ones and the input ones is 28.53 dB. In conclusion, all four metric values are greatly improved when we consider the de-noised magnetograms generated by our deep-learning model.

We compare our results with conventional smoothing methods such as median, Gaussian, and bilateral methods.⁴ Optional parameters of these methods are determined by looking for the best correlations with target images in the training data sets. More details about these methods and their optional parameters are described in the [Appendix](#). Overall metric scores from these methods are similar to those from our model. However, the average noise level of the de-noised magnetograms is noticeably smaller than those from these methods. The linear-fitting results of our method are better than those of the other methods. A visual comparison among the methods is given in the [Appendix](#).

We have made our model by training the magnetograms taken at the solar center. To look for a possible application of our model, we apply the model to two different regions: the solar center and near the limb. Figure 4 shows an example of the application to full-disk *SDO*/HMI magnetogram at 00:00 UT on 2017 September 5. The first region, denoted by (a) and (a') in Figure 4, is located at solar center, but is four times wider area than our data sets. The second region, denoted by (b) and (b') in Figure 4, is located near the limb, but is four times wider than our data sets. A careful comparison between the input magnetogram and the de-noised one for two regions shows that noise signals in the input ones are successfully removed in the de-noised ones. The noise level in the first region decreases from 10.20 to 4.05 G, and in the second region from 11.79 to 4.91 G.

5. Conclusion and Summary

In this Letter, we have applied a deep-learning method based on DCGAN to the de-noising of solar magnetograms. We have selected 8447 pairs of *SDO*/HMI magnetograms as the input and their corresponding stacked magnetograms as the target. We have separated our data sets into training, validation, and test sets in chronological order. We have trained our model using 7004 pairs from 2013 January to 2013 October. Then we have validated the model using 707 pairs from 2013 November, and tested 736 pairs from 2013 December.

The main results of this study are as follows. First, our model successfully generates the de-noised *SDO*/HMI magnetograms, and the de-noised magnetograms are much more consistent with the target magnetograms than the input ones. Second, our model greatly reduces the noise levels of the input magnetograms. The average noise level of the de-noised magnetograms is 3.21 G, which is quite lower than that of the input ones (8.66), and is consistent with that of the target magnetograms, 3.21 G. It is also noted that the average noise level of the de-noised magnetograms is even lower than that of *SDO*/HMI 720 s magnetograms calculated by Liu et al. (2012), 6.3 G. Third, all five metric values (CC, RE, linear fitting, NMSE, and peak S/N) of the de-noised magnetograms are much better than those of the input ones. Fourth, we applied the trained model to a full-disk *SDO*/HMI magnetogram to show a

possibility of the application of our model from the solar center to solar limb. Then we found that the application is quite successful in that the noise level of the de-noised magnetogram is greatly improved.

In this Letter, we have demonstrated that a deep-learning model based on DCGAN can be used to generate the de-noised magnetograms by training many pairs of single and stacked ones. Our de-noised magnetograms can be used for several studies on small magnetic structures; canceling magnetic features, magnetic flux emergence, solar surface motions, magnetic turbulence at the photosphere, and so on (Livi et al. 1985; Schrijver et al. 1997; Chae et al. 2001; Abramenko 2018). This idea can be applied to many astronomical areas because S/Ns are not large due to insufficient photons. There are a few necessary conditions to apply this model to data. First, there should be enough data sets for training and test. Second, the integration of frames (or long-exposure observations) has to be successfully made. Third, there should be little significant motions of features during the integration. As a good example of application, our preliminary results show that this method is successfully applied to make Sloan Digital Sky Survey images de-noised (Park et al. 2019). Furthermore, our method can be applied in many scientific fields in which the integration of many frames are used to improve the S/N.

This work was supported by the BK21 plus program through the National Research Foundation (NRF) funded by the Ministry of Education of Korea, the Basic Science Research Program through the NRF funded by the Ministry of Education (NRF-2013M1A3A3A02042232, NRF-2016R1A2B4013131, NRF-2019R1A2C1002634, NRF-2019R1C1C1004778, NRF-2020R1C1C1003892), the Korea Astronomy and Space Science Institute (KASI) under the R&D program “Study on the Determination of Coronal Physical Quantities using Solar Multi-wavelength Images (project No. 2019-1-850-02)” supervised by the Ministry of Science and ICT, and Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP; 2018-0-01422, Study on analysis and prediction technique of solar flares). We thank the numerous team members who have contributed to the success of the *SDO* mission. We acknowledge the community effort devoted to the development of the following open-source packages that were used in this work: NumPy (numpy.org), Keras (keras.io), TensorFlow (tensorflow.org), and SunPy (sunpy.org).

Appendix Smoothing Methods

The result of our median smoothing for a given pixel p is given by

$$M[I]_p = \text{median}(I_q \in S), \quad (7)$$

where I_q is an intensity at a pixel q in window size S . The size of the median filter is set to be 3 by comparing the results of training data sets.

The results of our Gaussian smoothing for a given pixel p is given by

$$G[I]_p = \sum_{q \in S} G_\sigma(\|p - q\|) I_q, \quad (8)$$

where G_σ is a Gaussian function with σ , and I_q is an intensity at pixel q in window size S . The size of the Gaussian filter and σ

⁴ https://cs.nyu.edu/fergus/teaching/comp_photo/5_image_proc_noise_bilateral.pdf

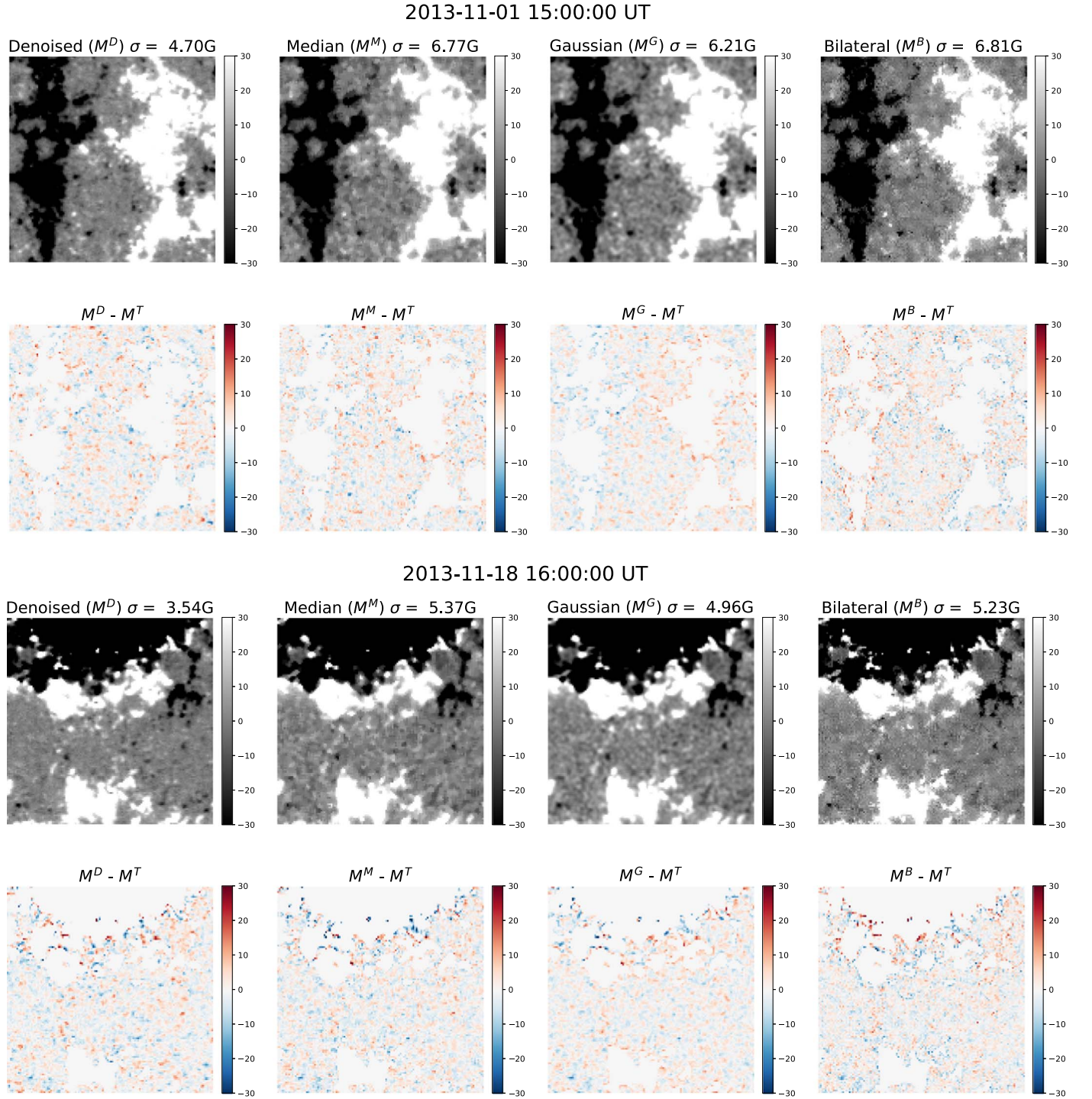


Figure 5. Two examples of the results of our model and three smoothing ones. The first column represents the results of our model, the second column represents the results of the median smoothing, the third column represents the results of the Gaussian smoothing and the last column represents the results of the bilateral smoothing. The first row represents the magnetograms from the methods, and the second row represents the difference maps between the results and target magnetograms.

are set to be 3 and 1 by comparing the results of training data sets.

The results of of bilateral smoothing for a give pixel p is given by

$$B[I]_p = \frac{1}{W_p} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I_p - I_q|) I_q, \quad (9)$$





where W_p is a normalized factor, σ_s is spatial extent of the kernel, σ_r is a minimum amplitude of an edge and I_q is an

intensity at pixel q in window size S . The size of the bilateral filter, σ_s , and σ_r are set to be 3, 10, and 10 by comparing the results of training data sets.

Figure 5 shows the results of our model and three smoothing ones for two examples. As shown in Figure 5, difference maps among these methods are similar to one another. Our visual inspection shows that the quality of magnetograms from our method are significant better than those from three smoothing methods. This is consistent with the fact that the noise levels of

the magnetograms from our model are smaller than those of the three smoothing methods.

ORCID iDs

Eunsu Park  <https://orcid.org/0000-0003-0969-286X>
 Yong-Jae Moon  <https://orcid.org/0000-0001-6216-6944>
 Daye Lim  <https://orcid.org/0000-0001-9914-9080>
 Harim Lee  <https://orcid.org/0000-0002-9300-8073>

References

- Abramenko, V. I. 2018, in IAU Symp. 340, Long-term Datasets for the Understanding of Solar and Stellar Magnetic Cycles, ed. D. Banerjee et al. (Cambridge: Cambridge Univ. Press), 281
- Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, *AJ*, 132, 1729
- Birney, D. S., Gonzalez, G., & Oesper, D. 2006, *Observational Astronomy* (2nd ed.; Cambridge: Cambridge Univ. Press)
- Chae, J., Martin, S. F., Yun, H. S., et al. 2001, *ApJ*, 548, 497
- DeForest, C. E. 2017, *ApJ*, 838, 155
- Díaz Baso, C. J., de la Cruz Rodríguez, J., & Danilovic, S. 2019, *A&A*, 629, A99
- Galvez, R., Fouhey, D. F., Jin, M., et al. 2019, *ApJS*, 242, 7
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., et al. 2014, in *Advances in Neural Information Processing Systems 27*, ed. Z. Ghahramani et al. (Red Hook, NY: Curran Associates, Inc.), 2672
- Isola, P., Zhu, J.-Y., Zhou, T., et al. 2016, in *IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2017*, http://openaccess.thecvf.com/content_cvpr_2017/html/Isola_Image-To-Image_Translation_With_CVPR_2017_paper.html
- Kim, T., Park, E., Lee, H., et al. 2019, *NatAs*, 3, 397
- Kingma, D. P., & Ba, J. 2014, in *3rd International Conf. on Learning Representations, ICLR 2015*, <https://dblp.org/rec/journals/corr/KingmaB14.bib>
- Lecun, Y., Bengio, Y., & Hinton, G. 2015, *Natur*, 521, 436
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, *Proc. IEEE*, 86, 2278
- Ledig, C., Theis, L., Huszar, F., et al. 2016, in *IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2017*, http://openaccess.thecvf.com/content_cvpr_2017/html/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.html
- Liu, Y., Hoeksema, J. T., Scherrer, P. H., et al. 2012, *SoPh*, 279, 295
- Liu, Y., Zhao, X., & Hoeksema, J. T. 2004, *SoPh*, 219, 39
- Livi, S. H. B., Wang, J., & Martin, S. F. 1985, *AuJPh*, 38, 855
- Mirza, M., & Osindero, S. 2014, arXiv:1411.1784
- Park, E., Moon, Y.-J., Lee, J.-Y., et al. 2019, *ApJL*, 884, L23
- Park, Y., Choi, Y.-Y., Moon, Y.-J., et al. 2019, *BKoAS*, 44, 54
- Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. 2012, *SoPh*, 275, 3
- Radford, A., Metz, L., & Chintala, S. 2015, in *4th International Conf. on Learning Representations, ICLR 2016*, <https://dblp.org/rec/journals/corr/RadfordMC15.bib>
- Scherrer, P. H., Schou, J., Bush, R. I., et al. 2012, *SoPh*, 275, 207
- Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, *SoPh*, 275, 229
- Schrijver, C. J., Title, A. M., van Ballegoijen, A. A., et al. 1997, *ApJ*, 487, 424
- Wang, J., Wang, H., Tang, F., et al. 1995, *SoPh*, 160, 277