

Convective Clouds Extraction From Himawari-8 Satellite Images Based on Double-Stream Fully Convolutional Networks

Xiaodong Zhang, *Member, IEEE*, Tong Wang[✉], Guanzhou Chen[✉], *Member, IEEE*, Xiaoliang Tan, and Kun Zhu

Abstract—Auto-extraction of convective clouds is of great significance. Convective clouds often bring heavy rain, strong winds, and other disastrous weather. Early warning of convection can effectively reduce loss. Using remote sensing images, we can get large-scale cloud information, which provides many effective methods for convective clouds detection. In this letter, we proposed a novel method to extract convective clouds. We introduce a novel deep network using only 1×1 convolution (3ONet) to extract the spectral characteristics. We then combine a 3ONet with the symmetrical dense-shortcut deep fully convolutional networks (SDFCNs) with a double-stream fully convolutional network to extract convective clouds. In the experiment, we used 12 000 Himawari-8 satellite image patches to verify the proposed framework. Experimental results with 0.5882 mean intersection over union (mIOU) pointed out the proposed method can extract convective clouds effectively.

Index Terms—Convective clouds, deep learning, fully convolutional network (FCN), remote sensing.

I. INTRODUCTION

CONVECTIVE clouds detection and early warning is an important part of modern weather forecasting. They usually cause severe weather such as strong wind, hail, thunder and lightning, and heavy rainfall [1], [2]. Remote sensing is a powerful method for convective cloud detection. The development of meteorological satellites has brought convenience to convective clouds monitoring [3], [4]. The use of geostationary satellite data makes it possible to obtain large-scale, continuous-time convection information, and expands the convection extraction method [5], [6].

There have been many studies on convection based on geostationary satellite data. Bai *et al.* [7] detect clouds by employing the infrared (IR) brightness temperature (BT_{IR}) and IR brightness temperature difference (BT_{DIR}) threshold values. However, it is easy to extract some cirrus clouds, because their cloud top brightness temperatures are sometimes similar between convective clouds and cirrus clouds [8].

Manuscript received April 23, 2019; revised June 10, 2019; accepted June 29, 2019. Date of publication July 19, 2019; date of current version March 25, 2020. This work was supported in part by the LIESMARS Special Research Funding and in part by the Fundamental Research Funds for the Central Universities. (*Corresponding author: Guanzhou Chen.*)

X. Zhang, G. Chen, X. Tan, and K. Zhu are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zxdlmars@whu.edu.cn; cgz@whu.edu.cn; xl_tan@whu.edu.cn; zkun@whu.edu.cn).

T. Wang is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: kingcopper@whu.edu.cn).

Color versions of one or more of the figures in this letter are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LGRS.2019.2926402

Carvalho and Jones [9] used the maximum spatial correlation to track and match the mesoscale convection systems.

However, these traditional convective clouds extraction methods need hand-designed features and carefully selected thresholds for classification [7], [8]. Deep neural networks have incomparable advantages in automatic advanced feature extraction and image classification [10], [11]. Convective clouds extraction is essentially a semantic segmentation task. In recent years, with the development of deep learning, many networks for semantic segmentation such as fully convolutional network (FCN) [12], pyramid scene parsing network (PSPNet) [13], and SegNet [14] have appeared.

Deep learning is increasingly applied in remote sensing. Application of deep learning technology can well extract the features of remote sensing images [15] and further improve the classification accuracy of the images. Many scholars have introduced the FCN model into the field of remote sensing image semantic segmentation. Romero *et al.* [16] suggest that deep architectures significantly outperform single-layer variants, capturing increasing levels of abstraction and complexity throughout the feature hierarchy. Maggiori *et al.* [17] proposed a fully convolutional neural network, and speeded up the training and shortened the running time, while also improving the classification accuracy. Chen *et al.* [18] introduced a symmetrical dense-shortcut deep FCN (SDFCN) based on FCNs model and obtained a better performance on ISPRS Vaihingen and Potsdam data sets than the classic FCN-8s and SegNet models.

In this letter, we designed a double-stream FCN for the convective clouds extraction task based on the data of the Himawari-8 satellite. We proposed the 3ONet to extract the spectral features and used the SDFCN to extract the geometric features.

The contributions of this letter include two points.

- 1) *1 × 1 Conv*: We explored the role of 1×1 convolution and found its ability to extract spectral features. On this basis, we proposed a network consisting of only 1×1 convolution (3ONet).
- 2) *Double-Stream Network*: We proposed a double-stream FCN. We combine the spectral features extracted by 3ONet and the geometric features extracted by SDFCN to complete the semantic segmentation task.

Following this letter, we introduce the importance of spectral information in remote sensing images in Section II-A. The double-stream neural network, the 3ONet, and the SDFCN are introduced in Sections II-B and II-C. In Section III,

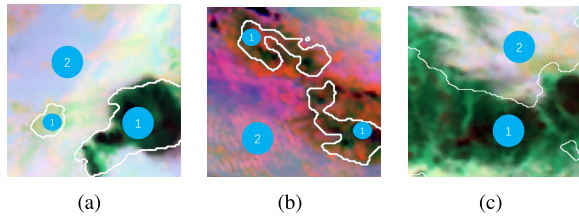


Fig. 1. (a)–(c) Data introduction: the false colors images with ground-truth boundary. Area 1 in the three pictures is convective clouds, and the non-convective region is area 2.

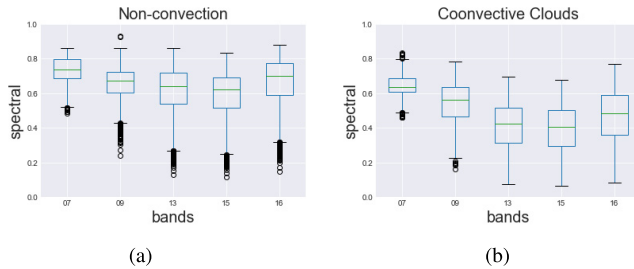


Fig. 2. Boxplot of statistical information of spectral response in (a) non-convective region and (b) convective clouds. The box represents a 50% range value, and the horizontal line in the box represents mean value. The horizontal line outside the box represents a 90% value, and the black circle represents outlier.

we present the data, experiment, results, and analysis. The conclusions and future work we conducted are discussed in Section IV.

II. METHODS AND NETWORK

A. Spectral Characteristics of Himawari-8 Images

For remote sensing images, especially multi-spectral images, the spectral characteristics must be considered when interpreting [15]. We select some samples for false color synthesis, as shown in Fig. 1. The image combined from false colors in the 7, 9, and 13 bands. The different hues between them indicate that spectral characteristics between the convective clouds and the non-convective region are different and that the spectral features are relatively significant.

Then, we calculate the spectral characteristics of the convective clouds and non-convective regions of all samples, and we draw the boxplots in Fig. 2. It is obvious that the spectral response of convection and non-convection is different in each band. Therefore, to classify and identify the remote sensing images in this task, we cannot blindly apply the traditional deep learning framework but consider the network sensitive to spectral information.

The so-called spectral feature in remote sensing essentially uses a plurality of bands of data to perform a certain function operation to obtain an index

$$z = f(x_1, x_2, \dots, x_n). \quad (1)$$

In formula (1), z represents a certain index, $x_1 - x_n$ represents the reflectance or response of each band, and f represents the established functional relationship.

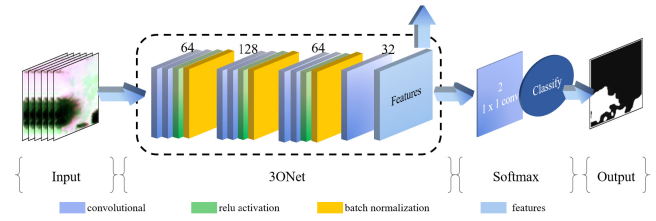


Fig. 3. 3ONet structure. Blue: convolutional layer. Number above: count of convolution kernels. Green: ReLU activation function. Orange: batch normalization layer. The final extracted spectral features are 32 dimensions.

The index responds well to characteristics in the spectral curve and has a strong correlation with certain surface parameters.

B. 3ONet

To extract spectral features that are not disturbed by spatial features, we adopt 1×1 convolutional layer as the fundamental structure in our networks. The essence of the 1×1 convolution kernel is to make a linear combination of the features in different channels. It is not affected by surrounding pixels, thus avoiding the effects of spatial features. We can also consider a 1×1 convolution as a full connection. Schalkoff [19] proved that enough neurons can fit any function. Therefore, we can use a deep 1×1 network to extract spectral features. In addition, the result of it is not affected by the geometry, but only related to the pixel's spectral characteristics.

Based on the considerations, we design a neural network consisting of only 1×1 convolution. The advantage is that a 1×1 convolution only calculates a single pixel, avoiding mixing with surrounding pixels. Therefore, we can separate the spectral information from the geometric information.

Fig. 3 shows the network we design. In this network, the size of all convolution kernels is 1×1 . We design a convolution block, which contains two convolutional layers, a rectified linear unit (ReLU) [20] layer, and a batch normalization [21] layer. In 3ONet, we superimpose three convolution blocks. In the end, we use a convolutional layer with 32 convolution kernels and treat its output as a spectral feature.

C. Double-Stream Network

In addition, spatial relationships should also be considered in remote sensing image interpretation. For example, convective clouds are also around convective clouds. In addition, there is an overlap between the spectral response of the convective clouds and the non-convection regions, as shown in Fig. 2. Therefore, spectral characteristics are not sufficient to accurately identify the convective clouds. We use a double-stream network to combine the spectral and geometric features. Fig. 4 shows its structure. The image data will follow in two streams.

We adopt the SDFCN [18] to extract the geometric features. Instead of the traditional FCN, we symmetry the encoding layer with the decoding layer and add a shortcut connection between them. It has been proven in [18] that this network has a good effect on the semantic segmentation task of

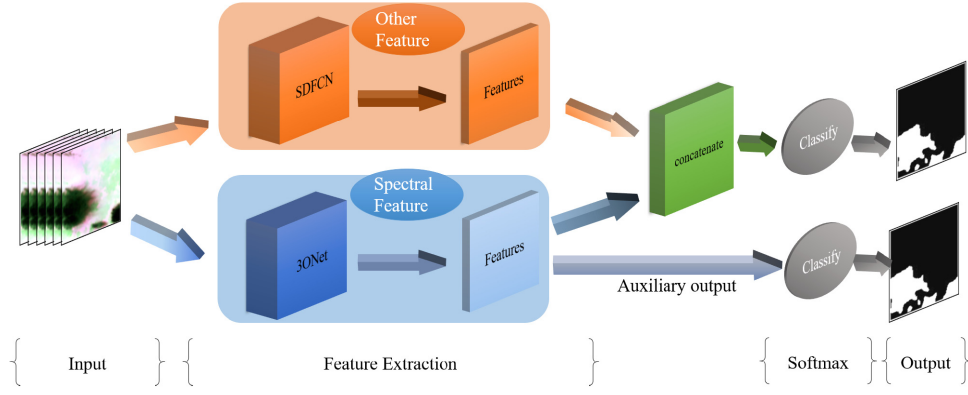


Fig. 4. Model structure. Blue arrow: data stream of spectral feature extraction. Orange arrow: data stream of other feature extraction, including geometric and texture features. Green arrow: combined feature. Gray arrow: classification using softmax.

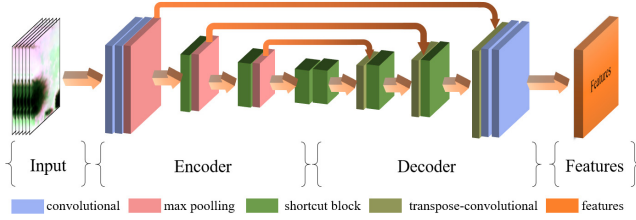


Fig. 5. SDFCN structure. SDFCN is stacked by encoders, decoders, and last softmax output layers. Encoders and decoders are symmetrically designed. Encoders are stacked by shortcut blocks and maxpooling layers, whereas decoders are stacked by transpose-convolutional layers and shortcut blocks.

remote sensing images. Fig. 5 shows the network's structure. The image data pass through the encoder layers and enter the decoder layers. The shortcut between the encoder and the decoder can help the network get more details of the image. Similar to 3ONet, we add a convolution layer containing 32 convolution cores behind the network and regard its output as geometric features. The window size of the convolution filter is 1×1 .

We regard the output of the 3ONet as spectral features. The two features are concatenated and combined through a convolution of 1×1 . In the classification layer, we use the softmax function to get the results.

In addition, we set up an auxiliary output to help 3ONet learn to extract spectral features. During the training, we conduct an experiment on the setting of its weight. The specific details will be shown in Section III.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Implementation of the Model

We use *TensorFlow* as the backend and implement it using *Keras*. We use the stochastic gradient descent (SGD) as the optimizer. Other training strategies used are as follows.

- 1) **Learning Rate:** We use the power decay function $lr = lr_{base} \times (1 - (\text{iter}/\text{iter}_{max}))^{power}$ to reduce the learning rate with the increase in training times. We set the lr_{base} to 0.01 and the *power* to be 0.9.
- 2) **Data:** Since the data format is uint16 type, the theoretical range is 0–65 535. Then, the data can be normalized to the range of 0–1 by dividing the data by 65 535 [formula (2)]. In addition, we also use local-minmax

TABLE I
BAND CHARACTERISTICS OF HIMAWARI-8/AHI

Band	7	9	13	15	16
Central wavelength (μm)	3.8853	6.9410	10.4073	12.3806	13.2807

Equatorial resolution of the five bands is 2km.

normalization [formula (3)] and global-minmax normalization [formula (4)] methods for data normalization. We compared the results with non-normalization. In addition, we used data improvement to flip the image randomly horizontally and vertically

$$N1_i^t = I_i^t / 65535 \quad (2)$$

$$N2_i^t = (I_i^t - \text{Min}_i^t) / (\text{Max}_i^t - \text{Min}_i^t) \quad (3)$$

$$N3_i^t = (I_i^t - \text{Min}^t) / (\text{Max}^t - \text{Min}^t). \quad (4)$$

In (2)–(4), I denotes the original image, subscript i means the i th image, and the superscript t means the t th band. $N1$ – $N3$, respectively, means the 0–65 535, local-minmax, and global-minmax normalized images.

- 3) **Auxiliary Loss Weight:** For auxiliary losses in a double-stream network, we set the weight to 0.4.
- 4) **Sample Balance:** We adopt the method of under-sampling negative samples. Since the samples have been divided into 100×100 pixel size, the specific method is to filter out samples with less than 1076 pixels in the convection area.

During the training, we used different loss functions, including Dice-loss, which is based on the Sørensen-Dice similarity coefficient for measuring the overlap between two segmented images [22], and categorical cross entropy loss (CE-loss), and Dice + CE-loss. We also performed experimental tests.

B. Data-Himawari-8/AHI

Himawari-8 [23]: Himawari-8 is a new geostationary meteorological satellite operated by the Japan Meteorological Agency (JMA). The satellite is located at 140.7°E and observes the Earth from 80°E to 160°W between 60°N and 60°S . The advanced Himawari imager (AHI) is an optical

TABLE II
NORMALIZATION AND LOSS FUNCTION EXPERIMENT RESULTS

Methods	FCN-8s		PSPNet		SDFCN		3ONet		Double-stream	
	κ	mIOU	κ	mIOU	κ	mIOU	κ	mIOU	κ	mIOU
ce+none	0	0.4130	0.3173	0.5335	0.0001	0.4112	0.3219	0.5330	0.3865	0.5628
dice+none	0	0.4111	0.2665	0.5169	0	0.4111	0.0626	0.1818	0.2688	0.4700
dice+ce+none	0	0.3330	0.3776	0.5468	0	0.4111	0.2441	0.3967	0.0826	0.2086
ce+n1	0	0.4111	0.2211	0.4917	0.0080	0.4140	0.2149	0.3643	0.3550	0.5040
dice+n1	0	0.4111	0.0793	0.2052	0.1201	0.2566	0.0249	0.1277	0.3657	0.5391
dice+ce+n1	0	0.4111	0.3203	0.5344	0.1517	0.4692	0	0.4111	0	0.4111
ce+n2	0.2317	0.4855	0.2537	0.4905	0.1706	0.4738	0.2224	0.4536	0.2727	0.5091
dice+n2	0	0.4111	0.2724	0.4977	0.2353	0.4951	0.1961	0.4756	0.2419	0.4856
dice+ce+n2	0.2090	0.4842	0.1919	0.4714	0.3169	0.5266	0.2525	0.4852	0.2521	0.5029
ce+n3	0.4157	0.5701	0.1565	0.4708	0.3999	0.5421	0.4277	0.5849	0.4247	0.5882
dice+n3	0	0.4111	0.3572	0.5573	0.3324	0.5028	0.3500	0.5532	0.4149	0.5776
dice+ce+n3	0.4073	0.5638	0.2078	0.4924	0.4243	0.5879	0.4273	0.5713	0.2618	0.4243

n1 0-65535 normalization

n2 local-minmax normalization

n3 global-minmax normalization

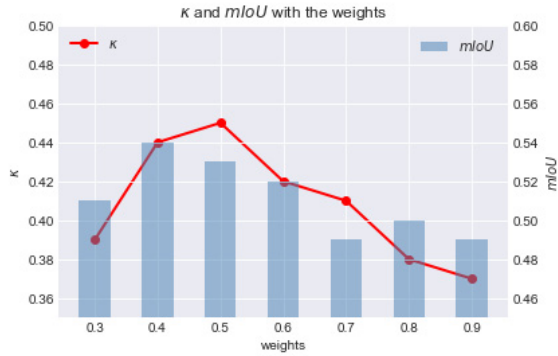


Fig. 6. κ and mIOU with weights: the relationship between the κ coefficient and the weight, and the relationship between mIOU and weight by using different auxiliary loss weights for network training. The x -axis means the auxiliary loss weights.

radiometer on board the Himawari-8. Its observation frequencies are every 10 min for the full disk and every 2.5 min for the area adjacent to Japan. The AHI has 16 spectral bands from visible to IR wavelengths. The spatial resolution of the IR bands is 2 km. The details of the data are shown in Table I. We only used some of the AHI's mid-IR data.

In addition to this, solar elevation angle data are also available, which can be used for solar elevation correction. We clip the satellite data to patch with a size of 100×100 pixels. There were 10000 patches in the training sample and 2000 patches in the verification set.

C. Results and Analysis

We trained and verified VGG-FCN-8s, PSPNet, SDFCN, 3ONet, and double-stream networks. Because of the sample imbalance, the pixels of convective clouds are about 5% of the total. If the classifier divides all regions into non-convective regions, then the overall accuracy (OA) reaches 0.95. But this does not mean that the classifier has high performance. Therefore, we did not use OA to evaluate the classification effect. We used the mean intersection over union (mIOU) [24] and κ coefficient, which can resist the imbalance of the sample to some extent [25] to judge the networks.

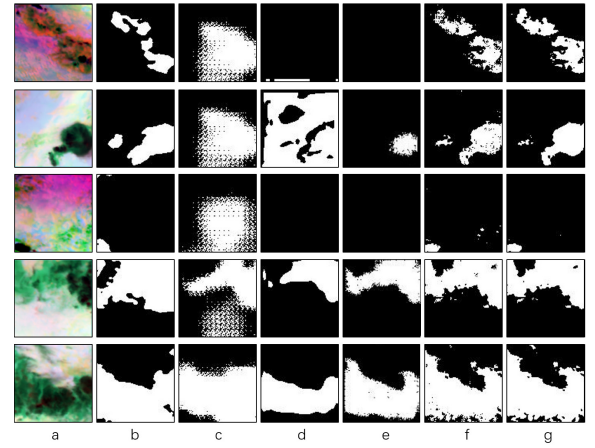


Fig. 7. (a) False color composite image, and the band used is 7/9/13. (b) Ground truth. (c) Prediction result of the FCN-8s. (d) PSPNet prediction result. (e) SDFCN prediction result. (f) 3ONet result. (g) Result of double-stream network. The white area in (b)–(g) is the convection area, and the black is the non-convection.

1) *Auxiliary Loss*: We tested the weight of the auxiliary loss and the test value is 0.3–0.9 at 0.1 interval. The results are shown in Fig. 6. From the κ coefficient, the result with the weight of 0.5 is the best, but 0.4 is the best from the perspective of mIOU. We had comprehensively considered both. Considering that κ coefficient has a small difference between 0.4 and 0.5 and mIOU has a large peak value at 0.4, we chose 0.4 as the weight of the final auxiliary loss.

2) *Normalization and Loss Function*: We used different loss functions and normalization methods to train each model. The test results are shown in Table II. It can be seen from the table that the normalization method has a great influence on the results of the model. For each model, the global-minmax normalization has the best results. Different networks have different best loss functions, and there is no obvious regularity between them. In addition, in the comparison between different models, we can see that 3ONet and SDFCN perform well in this task, and the double-stream network we designed has a better performance than others in terms of mIOU.

Some examples of model results are shown in Fig. 7.

D. Discussion

The 3ONet we designed excelled in this mission. Considering that it only uses 1×1 convolution, we could think the network extracts the spectral features of the AHI image. In addition, our double-stream FCN also completed the task of combining spectral characteristics with geometric characteristics.

In the example, the convective clouds extracted by the model have noise and small holes. We should put some post-processing such as morphological filtering to make the results smoother.

IV. CONCLUSION

In this letter, we proposed a double-stream network combining 3ONet and SDFCN to extract convective clouds using data of Himawari-8/AHI. We verified different normalization methods and loss functions. Our 3ONet and double-stream network are superior to FCN-8s and PSPNet in terms of mIOU and κ coefficients. In addition, we found that the normalization method has a great influence on the results. Global-minmax normalization can effectively improve the mIOU and κ coefficient of the model results.

In the future, we will focus on the post-processing of the results to further improve the classification accuracy. We also consider transferring our network to fit other satellite data for wider applications. In addition, the volume of the network should be further reduced by knowledge distillation and other methods.

ACKNOWLEDGMENT

The authors would like to thank the developers of TensorFlow and Keras. They would also like to thank the Himawari-8 data provided by Piesat, Inc. The numerical calculations in this letter have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

REFERENCES

- [1] T. Fujita, D. L. Bradbury, C. Murino, and L. Hull, "A study of mesoscale cloud motions computed from ATS-I and terrestrial photographs," *Satell. Mesometeorolog. Res. Project, Res. Paper* 71, 1968.
- [2] K. Liang, H. Shi, P. Yang, and X. Zhao, "An integrated convective cloud detection method using FY-2 VISSR data," *Atmosphere*, vol. 8, no. 2, p. 42, 2017.
- [3] M. Baldauf, A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, "Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities," *Monthly Weather Rev.*, vol. 139, no. 12, pp. 3887–3905, 2011. doi: [10.1175/MWR-D-10-05013.1](https://doi.org/10.1175/MWR-D-10-05013.1).
- [4] Q. Long, Q. Chen, K. Gui, and Y. Zhang, "A case study of a heavy rain over the Southeastern Tibetan plateau," *Atmosphere*, vol. 7, no. 9, p. 118, Sep. 2016.
- [5] M. Wei, W. Xu, and T. Zuyu, "Geographic distribution and life cycle of mesoscale convective system in china and its vicinity," *Prog. Natural Sci.*, vol. 5, p. 9, 1997.
- [6] T. Z. W. H. W. Xu and M. Yu, "A survey of MESO- α -scale convective systems over China during 1995," *ACTA Meteorolog. Sinica*, vol. 2, pp. 39–50, Apr. 1998.
- [7] J. Bai, H. Wang, and Z. Tao, "Recognition and tracing of severe convective cloud from ir images of gms," *J. Tropical Meteor.*, vol. 3, no. 1, pp. 192–201, 1997.
- [8] D. Doelling, A. Gambheer, and J. Stassi, "Deep convective cloud calibration," in *Proc. 29th CERES Sci. Team Meeting*, Hampton, VA, USA, 2003, pp. 281–289.
- [9] L. M. V. Carvalho and C. Jones, "A satellite method to identify structural properties of mesoscale convective systems based on the maximum spatial correlation tracking technique (mascotte)," *J. Appl. Meteorol.*, vol. 40, no. 10, pp. 1683–1701, 2001. doi: [10.1175/1520-0450\(2001\)040<1683:ASMTIS>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1683:ASMTIS>2.0.CO;2).
- [10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, May 2015. doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [11] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, p. 1, Feb. 2015. doi: [10.1186/s40537-014-0007-7](https://doi.org/10.1186/s40537-014-0007-7).
- [12] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [13] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.
- [14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [15] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the Art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.
- [16] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349–1362, Mar. 2016.
- [17] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Fully convolutional neural networks for remote sensing image classification," in *Proc. IEEE IGARSS*, Jul. 2016, pp. 5071–5074.
- [18] G. Chen, X. Zhang, Q. Wang, F. Dai, Y. Gong, and K. Zhu, "Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1633–1644, May 2018.
- [19] R. J. Schalkoff, *Artificial Neural Networks*, vol. 1. New York, NY, USA: McGraw-Hill, 1997.
- [20] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8609–8613.
- [21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: <https://arxiv.org/abs/1502.03167>
- [22] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *CoRR*, vol. abs/1606.04797, pp. 565–571, Oct. 2016. [Online]. Available: <http://arxiv.org/abs/1606.04797>
- [23] K. Bessho *et al.*, "An introduction to Himawari-8/9—Japan's new-generation geostationary Meteorological Satellites," *J. Meteorol. Soc. Jpn. II*, vol. 94, no. 2, pp. 151–183, 2016.
- [24] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodríguez, "A review on deep learning techniques applied to semantic segmentation," *CoRR*, vol. abs/1704.06857, Apr. 2017. [Online]. Available: <http://arxiv.org/abs/1704.06857>
- [25] J. Carletta, "Assessing agreement on classification tasks: The kappa statistic," *Comput. Linguistics*, vol. 22, no. 2, pp. 249–254, Jun. 1996. [Online]. Available: <http://dl.acm.org/citation.cfm?id=230386.230390>