Analysis of the Influence of Convolutional Layer in Deep Convolutional Neural Network on SAR Target Recognition

Wei Qu[1(✉)], Gang Yao[2], and Weigang Zhu[1]

[1] Space Engineering University, Beijing, People's Republic of China
quweistar@163.com
[2] Beijing Institute of Tracking and Telecommunication Technology, Beijing, People's Republic of China

Abstract. As a frontier hot spot in the current image processing field, deep learning has unparalleled superiority in feature extraction. Deep learning uses deep network structure to perform layer-by-layer nonlinear transformation, which can achieve the approximation of complex functions. From low-level to high-level, the representation of features becomes more and more abstract, and the more essential the original data is described. Aiming at the problem of SAR image target detection, this paper studies the influence of the number of convolution kernels, the size of the convolution kernel and the number of convolution layers in the deep convolutional neural network on SAR target recognition.

Keywords: Deep convolutional neural network · SAR · Target recognition

1 Introduction

Synthetic Aperture Radar (SAR), as a branch of microwave remote sensing, obtains ground object information through the interaction between electromagnetic waves and various media. Compared with optical remote sensing, it has all-weather and all-weather detection. Reconnaissance capabilities. In addition, SAR obtains large-area high-resolution radar images through the use of range-wise pulse compression technology and azimuth synthetic aperture technology, which can provide strong support for military intelligence acquisition, precision guidance, and strike effect evaluation [1, 2].

From the perspective of SAR image target detection and recognition algorithms, its essence is based on the acquisition of image features and the design of feature usage rules. For example, the most commonly used target detection algorithm based on Constant False Alarm Rate (CFAR) [3] mainly uses the gray-scale features of SAR images [6]. The feature extraction rules of these methods are often manually designed. In practical applications, when the amount of data is too large and the data is complex, the features extracted in this way are usually not representative and cannot represent the uniqueness between different types of data. This limits the accuracy of detection and recognition.

S. Shi et al. (Eds.): AICON 2020, LNICST 356, pp. 404–413, 2021.
https://doi.org/10.1007/978-3-030-69066-3_35

2 SAR Image Target Detection Method Based on Deep Convolutional Neural Network

The SAR target detection method based on the deep convolutional neural network (CNN) constructs a structure containing multiple deep neural networks to transform the original data into a higher-level and more abstract expression using a combination of nonlinear mapping relationships. Compared with the method based on pattern classification Methods such as support vector machines and traditional neural networks have the ability to extract features autonomously. Compared with deep neural network architectures such as restricted Boltzmann machines and autoencoders, CNN has more advantages in image classification and recognition, and has achieved more achievements. With the widespread recognition of CNN in optical image processing in recent years [7], it has also set off a research boom in the field of SAR target recognition. Morgan et al. [8] directly used the typical CNN structure to obtain good recognition performance, but did not consider the impact of the network structure design on the SAR image target recognition performance, and its accuracy was not ideal. In order to further improve the recognition performance of CNN, scholars have made improvements in the design of the network structure. Tian Zhuangzhuang [9] and others first improved the classification ability of CNN by introducing a category separability measure in the error cost function, and finally used SVM to classify features. Zhao et al. [10] used the highway convolutional layer to reduce the data requirements of the network, and the recognition rate was further improved. Xu Feng [11] and others use a sparse convolutional architecture instead of a fully connected layer, which can avoid the overfitting problem caused by a small training set.

Deep Convolutional Neural Network (CNN) due to its powerful feature extraction capabilities, not only has made great achievements and widespread recognition in optical image processing, but also achieved a better recognition rate than traditional recognition methods in SAR target recognition. However, the accuracy, timeliness, and generalization of deep CNN in SAR target recognition need to be further explored, mainly as follows: 1) In terms of recognition accuracy, different activation functions, regularization functions, number of convolution kernels and Factors such as size have different effects on the recognition rate; 2) In terms of recognition timeliness, changing the convolutional layer structure parameter settings within a certain range is beneficial to the improvement of the recognition rate, and it also often increases the amount of network calculation and training time. It is not conducive to the timeliness of recognition; 3) In terms of recognition generalization, since there are multiple variants of the same type of target and the SAR image target has attitude sensitivity, the recognition network needs to extract more robust and generalized features.

In response to the above problems, this paper studies the SAR target recognition method based on deep CNN. First, based on the classic convolutional network Lenet-5, it is improved to construct the basic structure of SAR target recognition network (SAR-Lenet), and the influence of activation function and regularization function on recognition performance is analyzed; then, each volume The effect of the number and size of the convolution kernel of the buildup layer on the recognition performance is comparatively analyzed; again, in order to explore the impact of the number of network

convolutional layers on the recognition performance, the basic network is deformed by the layer replacement idea, and the number of convolutional layers is increased. While keeping the training time of the network basically unchanged; finally, optimize the SAR-Lenet network structure for the generalization of the network, use the dense block network structure to merge multiple convolutional layer features, and use the convolutional layer to replace the fully connected layer to increase The sparseness of the network, and at the same time, the optimized network (SAR-Net) is trained with data augmentation and expansion training set to further improve the generalization of the recognition network.

The convolutional layer is the main component of CNN, and different settings of its structural parameters will have different effects on the recognition network. The convolutional layer structure parameters refer to the number of convolution kernels, the size of convolution kernels, and the number of convolution layers. From the SAR-Lenet convolution layer structure parameters, it follows the parameter settings of Lenet-5, and these parameter settings are based on handwritten numbers. The recognition requirements may not meet the SAR target recognition requirements. Therefore, through comparative experiments, this paper explores the influence of the number of convolution kernels, the size of convolution kernels and the number of convolution layers in the SAR-Lenet convolution layer on recognition performance, and provides a basis for the subsequent parameter setting of the convolution layer in the target recognition network.

3 The Effect of the Number of Convolution Kernels on Recognition Performance

In the setting of the number of SAR-Lenet convolution kernels, keep the size of the convolution kernels of the three layers unchanged, keep the number of convolution kernels of the three layers the same, and set the number of convolution kernels to 8, 16, 32, 64, 128, respectively and 256, where A1–A6 represent the network with six different convolution kernel configurations. The time loss of the network convolution layer increases by 4 times. At the same time, A7, A8 and A9 are set on the basis of six orders of magnitude. The network parameters of nine different convolution kernels are set as shown in Table 1. In the experiment, the SOC data set is used for training and testing. During the training process, the recognition rate curve of the test set is shown in Fig. 1, and the network training time consumption is shown in Fig. 2.

Table 1. Networks with different numbers of convolution kernels

Layer	Kernel number	Step	A1	A2	A3	A4	A5	A6	A7	A8	A9
Conv1	7 × 7	2	8	16	32	64	128	256	64	128	32
Conv2	5 × 5	1	8	16	32	64	128	256	128	64	64
Conv3	3 × 3	1	8	16	32	64	128	256	128	128	128

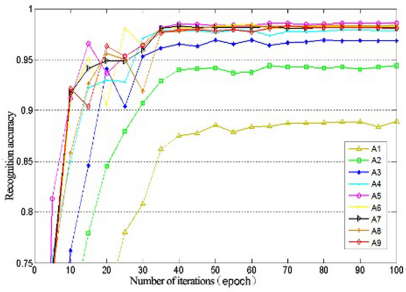

Fig. 1. Network recognition rate curve

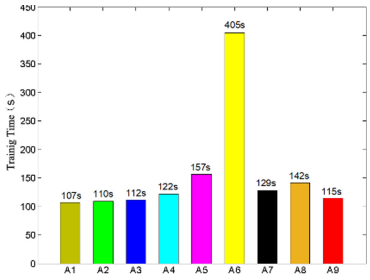

Fig. 2. Network training time consumption graph.

Comparing the recognition rate curves of A1, A2, and A3, it can be seen that as the number of convolution kernels increases, the recognition rate of the network increases greatly. From A3 to A5, the recognition rate increases less and less, and the accuracy increases by 1. % Or less, and the recognition rate from A5 to A6 has dropped. The results of the six networks show that the number of convolution kernels in the A1, A2, and A3 networks is not enough to meet the network's demand for the number of feature maps. The increase in the number of feature maps in the A3, A4, and A5 networks weakens the recognition rate. Recognition performance tends to be saturated. A6, due to the excessive number of feature maps, easily leads to network overfitting, and the recognition rate drops instead.

Comparing the A7 and A8 networks, the total number of convolution kernels is the same, but the number of conv1 and conv1 feature maps is reversed. Because SAR-Lenet adds a pooling layer to each convolutional layer to reduce the dimensionality, the size of the conv2 feature map It is half of conv1. Conv1 consumes more time than conv2 under the same number of convolution kernels. Therefore, the training time of A8 network is 13 s longer than that of A7 network, and the recognition rate of the two is basically the same. Comparing A9 and A4, the number of A9 convolution kernels is 32 more than that of A4, but the training time is 7 s less.

4 The Effect of Convolution Kernel Size on Recognition Performance

The size of the convolution kernel represents the size of the receptive field. In the performance study of the size of the convolution kernel, based on the A5 network with the highest recognition rate and the A1 network with the lowest recognition rate, the convolution kernel size is selected from 3, 5, 7, 9 Four sizes, four networks are obtained based on the A5 network, namely B1, B2, B3, B4; based on the A1 network, six networks are obtained, namely B5, B6, B7, B8, B9 and B10; The size of the convolution kernel for each network is shown in Table 2 and 3. In the experiment, the SOC data set was used for training and testing. During the training process, the recognition rate curves of the test set are shown in Fig. 3 and Fig. 5. The network training time consumption is shown in Fig. 4 and Fig. 6.

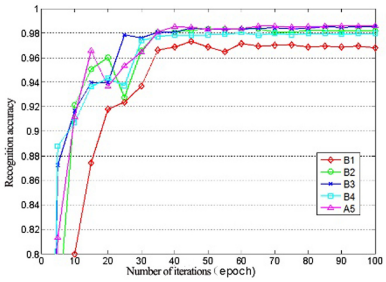

Fig. 3. Network recognition rate curve

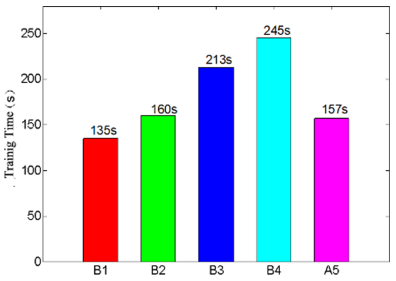

Fig. 4. Network training time consumption graph

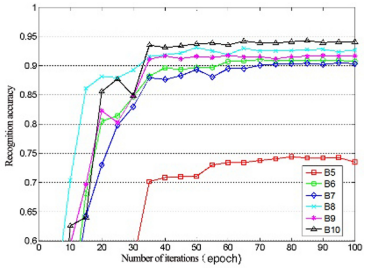

Fig. 5. Network recognition rate curve

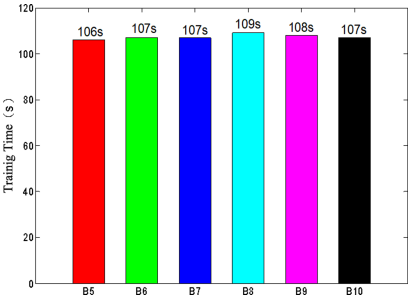

Fig. 6. Network recognition rate curve

Table 2. Networks with different convolution kernel sizes

Layer	Kernel size	Step	B1	B2	B3	B4
Conv1	128	2	3 × 3	5 × 5	7 × 7	9 × 9
Conv2	128	1	3 × 3	5 × 5	7 × 7	9 × 9
Conv3	128	1	3 × 3	5 × 5	7 × 7	9 × 9

Table 3. Networks with different convolution kernel sizes

Layer	Kernel size	Step	B5	B6	B7	B8	B9	B10
Conv1	8	2	3 × 3	5 × 5	7 × 7	9 × 9	5 × 5	9 × 9
Conv2	8	1	3 × 3	5 × 5	7 × 7	9 × 9	7 × 7	7 × 7
Conv3	8	1	3 × 3	5 × 5	7 × 7	9 × 9	9 × 9	5 × 5

From the above experimental results, when the number of convolution kernels is set to 128, since the number of network convolution kernels is basically saturated, the size of the convolution kernel has little effect on recognition performance, and the network training time increases as the size of the convolution kernel increases.; Comparing the five networks of A5 and B1–B4, A5 has the best recognition performance, indicating

that different convolution kernel size combinations are more conducive to the network extracting features of different scales, which is conducive to the improvement of network recognition rate, and A5 also accounts for timeliness excellent.

When the number of convolution kernels is set to 8, the number of network convolution kernels is under-saturated, and the size of the convolution kernel has a more obvious impact on network recognition performance. Compare B5, B6, B7 and B8, as the size of the convolution kernel increases the recognition rate of the network is improved. It can be considered that increasing the scale of the convolution kernel to extract the features can effectively compensate for the impact of the insufficient number of feature maps; compare B10 and B5–B8 five networks, B10 improves the recognition rate by combining different convolutions and sizes; Comparing B9 and B10, the two networks both use different convolution kernel size combinations, but the recognition rate of B10 is higher. It can be considered that shallow features are suitable for large-size convolution kernel extraction, and high-level features need to be extracted by small-size convolution kernel; It can be seen from Fig. 6 that the training time consumption of the six networks is relatively small. The main reason is that the number of network convolution kernels is small. The calculation difference between the networks is smaller than the total calculation, but it can still be seen the size of the convolution kernel increases and the training time of the network increases.

5 The Effect of the Number of Convolutional Layers on Recognition Performance

In general, increasing the number of convolutional layers can improve the accuracy of recognition and easily increase the risk of network overfitting. This section continues to use SAR-Lenet as the basis to explore the influence of different numbers of convolutional layers on network recognition performance. In the experiment, the SOC data set is used for training and testing.

Using the number of convolution kernels of the A9 network, the network has a higher recognition rate and better timeliness. Therefore, the number of convolution kernels of the three convolution layers of SAR-Lenet is set to 32, 64, and 128 in sequence. Since the network's demand for feature maps is basically satisfied with this number of convolution kernels, the size of the convolution kernel has a small impact on the recognition performance. To facilitate subsequent research, the size of the convolution kernel adopts two scales of convolution kernel size. At the same time, in accordance with the principle of setting the larger size of the first layer, the convolution kernel sizes of the three convolution layers are set to 7, 5, and 5 in sequence. The SAR-Lenet corresponding to the number and size of the convolution kernel is referred to as network C here. The experimental results of network C in Table 4–7 show that the above parameter settings have a better recognition rate and further reduce network training time overhead.

In order to avoid the increase of network time training while increasing the number of convolutional layers, the idea of layer replacement is adopted to split a convolutional layer into two convolutional layers, and at the same time change the size of the convolution kernel or the number of convolution kernels in the replacement layer, and other the

convolutional layer remains unchanged. Since only the number or size of the convolution kernel of the replacement layer has been changed, without considering its impact on the network recognition performance, the network overhead can be kept basically unchanged. There are two main aspects to increase the number of convolutional layers:

Increase the number of convolutional layers while reducing the size of the convolution kernel of the replacement layer, and the number of convolution kernels remains unchanged. Specifically, one convolutional layer in convolution block 3 of the C network can be replaced with two layers, and then the size of the convolution kernel can be reduced. The number of convolution kernels of the two convolutional layers is 128. The specific replacement process is as follows:

$$64 \times 5^2 \times 128 \Rightarrow 64 \times 3^2 \times 128 + 128 \times 3^2 \times 128 \tag{1}$$

So as to get the C1 network. The convolution block 2 of the C network can also be replaced as follows:

$$32 \times 5^2 \times 64 \Rightarrow 32 \times 3^2 \times 64 + 64 \times 3^2 \times 64 \tag{2}$$

To get the C2 network. The C3 network can be obtained by combining the C1 network and the replaced convolution block of C2. Then continue to replace the convolution block 2 and convolution block 3 of the C3 network as follows:

$$32 \times 3^2 \times 64 + 64 \times 3^2 \times 64 \Rightarrow 32 \times 3^2 \times 64 + 64 \times 2^2 \times 64 + 64 \times 2^2 \times 64 \tag{3}$$

$$64 \times 3^2 \times 128 + 128 \times 3^2 \times 128 \Rightarrow \ 64 \times 3^2 \times 128 + 128 \times 2^2 \times 128 + 128 \times 2^2 \times 128 \tag{4}$$

To get the C4 network. In this way, C1 and C2 are obtained from C, C3 is obtained from C1 and C2, and C4 is obtained from C3.

Increase the number of convolutional layers while reducing the number of convolution kernels in the replacement layer, and the size of the convolution kernel remains unchanged. In this way, the number of convolution kernels of the convolution layer added to the convolution block 3 in the C network can be reduced from 128 to 44, and the number of convolution kernels of the second convolution layer is still 128, which ensures the input of the convolution block 3 And the number of output feature maps remains unchanged, and the specific replacements are as follows:

$$64 \times 5^2 \times 128 \Rightarrow 64 \times 5^2 \times 44 + 44 \times 5^2 \times 128 \tag{5}$$

In this way, D1 is obtained; also the convolution block 2 of the C network is replaced as follows:

$$32 \times 5^2 \times 64 \Rightarrow 32 \times 5^2 \times 22 + 22 \times 5^2 \times 64 \tag{6}$$

In this way, D2 is obtained, and D3 is also obtained by combining D1 and D2 networks. At the same time, the convolution block 3 and convolution block 2 of the C1 network can also be replaced as follows:

$$64 \times 3^2 \times 128 + 128 \times 3^2 \times 128 \Rightarrow 64 \times 3^2 \times 88 + 88 \times 3^2 \times 88 + 88 \times 3^2 \times 128 \tag{7}$$

$$32 \times 3^2 \times 64 + 64 \times 3^2 \times 64 \Rightarrow 32 \times 3^2 \times 44 + 44 \times 3^2 \times 44 + 44 \times 3^2 \times 64 \tag{8}$$

Get the D4 and D5 networks respectively, and combine the two to get the D6 network. See Table 4 for details of the above deformation in the C network. For the layer replacement of the above network, in order to control the size of the output feature map after each convolution operation, the sliding step size of convolution block 2 and convolution block 3 is S = 1, and for the 5 × 5 convolution kernel zero padding P = 2. For 3 × 3 convolution kernel zero-filling P = 1, for two consecutive 2 × 2 convolution kernels, the first convolution kernel zero-filling P = 0, the second convolution kernel zero-filling P = 1. This makes the size of the feature map output by each convolution block unchanged. The parameters in the convolution block in Table 4 are the number and size of the convolution kernel. The sliding step size of the first convolution block is 2, and the step size of the remaining two convolution blocks is 1. The pooling layer in the network remains unchanged, and there are three maximum pooling layers after three convolutional blocks.

Table 4. The relationship between convolutional layer parameters and recognition performance

Network	Convolution block 1	Convolution block 2	Convolution block 3	Number of convolutional layers	Training time (s)	Average recognition rate (%)
C	32 × 7 × 7/2	64 × 5 × 5/1	128 × 5 × 5/1	3	123	98.04
C1	32 × 7 × 7/2	128 × 5 × 5/1	(128 × 3 × 3) × 2/1	4	121	98.57
C2	32 × 7 × 7/2	(64 × 3 × 3) × 2/1	128 × 5 × 5/1	4	122	98.89
C3	32 × 7 × 7/2	(64 × 3 × 3) × 2/1	(128 × 3 × 3) × 2/1	5	121	99.30
C4	32 × 7 × 7/2	64 × 3 × 3 + (64 × 2 × 2) × 2/1	128 × 3 × 3 + (128 × 2 × 2) × 2/1	7	123	99.07
D1	32 × 7 × 7/2	128 × 5 × 5/1	44 × 5 × 5/1 + 128 × 5 × 5/1	4	124	98.54
D2	32 × 7 × 7/2	22 × 5 × 5/1 + 64 × 5 × 5/1	128 × 5 × 5/1	4	125	98.51
D3	32 × 7 × 7/2	22 × 5 × 5/1 + 64 × 5 × 5/1	44 × 5 × 5/1 + 128 × 5 × 5/1	5	124	99.13
D4	32 × 7 × 7/2	128 × 5 × 5/1	(88 × 3 × 3) × 2/1 + 128 × 3 × 3/1	5	125	98.81
D5	32 × 7 × 7/2	(44 × 3 × 3) × 2/1 + 64 × 3 × 3/1	128 × 5 × 5/1	5	122	99.04
D6	32 × 7 × 7/2	(44 × 3 × 3) × 2/1 + 64 × 3 × 3/1	(88 × 3 × 3) × 2/1 + 128 × 3 × 3/1	7	120	98.96

It can be seen from the network training time that the layer replacement idea basically keeps the network time overhead unchanged. From the point of view of recognition rate, the recognition rate of the network after increasing the number of convolutional layers is higher than that of the C network, indicating that the C network is not saturated in the network depth. Increasing the network depth can effectively improve the recognition rate. Among them, the C3 network has a higher recognition rate. The C network has increased by 1.26%, which is the largest improvement. Comparing C3, C4 and D6, although the number of C4 and D6 convolutional layers is two more than C3, due to the saturation of recognition performance, the network has a certain degree of overfitting, and the recognition rate is instead Decrease; for networks with the same number of convolutional layers, the recognition rate of the C3 network is higher than that of the D3, D4, and D5 networks, and the recognition rate of the C1 and C2 networks is higher than that of D1 and D2, indicating that the combination of the number and size of different convolution kernels can identify the network It will also bring a certain impact; in the combination of convolutional layers, the recognition rate of C3 and D3 networks is higher than other networks, indicating that the matching method of convolutional blocks in this network structure is more suitable for target feature extraction.

6 Conclusion

Based on the above analysis, it can be seen that the number of different convolution kernels not only affects the recognition rate of the network, but also affects the timeliness of the network. When setting the number of convolution kernels, when the total number of convolution kernels in the network remains unchanged, the number of shallow convolution kernels can be appropriately reduced, and the number of deep convolution kernels can be appropriately increased, which will help improve the timeliness of the network. Combining the influence of the number and size of convolution kernels on the recognition performance of SAR-Lenet, it can be concluded that: 1) The number of convolution kernels has a greater impact on the recognition network performance than the size of the convolution kernel. Only when the number of convolution kernels is met, the recognition the rate is high; 2) When the number of convolution kernels is saturated, the size of the convolution kernel has little effect on the network performance. When the number of convolution kernels is insufficient, increasing the size of the convolution kernel is conducive to a very high recognition rate; 3) Different convolution kernels the size combination is more conducive to network feature extraction, and the scale of the first layer of convolution kernel is relatively large; 4) Increasing the number and size of the convolution kernel will increase the training time of the network. The number of SAR-Lenet convolutional layers is set to 5 and the C3 network structure has the highest recognition rate; when the number of convolution kernels tends to be saturated, increasing the number of convolution kernels does not improve the recognition rate, increasing the number of convolutional layers, the recognition rate can be It is further improved, but it is not that the more the number of convolutional layers, the better, it needs to be controlled.

References

1. Licheng, J., Xiangrong, Z., Biao, H., et al.: Intelligent SAR image processing and interpretation (2008)
2. Jianshe, S., Yongan, Z., Lihai, Y.: Synthetic Aperture Radar Image Understanding and Application. Science Press, Beijing (2008)
3. Performance analysis of CA-CFAR two approximate methods under linear detection
4. Gangyao, K., Gui, G., Yongmei, J.: Synthetic Aperture Radar Target Detection Theory, Algorithm and Application (2007)
5. Novak, L.M., Hesse, S.R.: On the performance of orderstatistics CFAR detectors. In: IEEE 25th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, USA, November 1991, pp. 835–840 (1991)
6. Wenqing, S., Yinghua, W., Hongwei, L.: Automatic region screening target detection algorithm for high-resolution SAR images. J. Electron. Inf. Technol. **38**(5), 1017–1025 (2016)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
8. Morgan, D.A.E.: Deep convolutional neural networks for ATR from SAR imagery. In: Proceedings of the Algorithms for Synthetic Aperture Radar Imagery XXII, Baltimore, MD, USA, 2015, vol. 23:94750F (2015)
9. Zhuangzhuang, T., Ronghui, Z., Jiemin, H., et al.: Research on SAR image target recognition based on convolutional neural network. J. Radars **5**(3), 320–325 (2016)
10. Lin, Z, Ji, K., Kang, M., et al.: Deep Convolutional Highway Unit Network for SAR target classification with limited labeled training data. IEEE Geosci. Remote Sens. Lett. (2017)
11. Chen, S., Wang, H., Xu, F., et al.: Target classification using the deep convolutional networks for SAR images. IEEE Trans. Geosci. Remote Sens. **54**(8), 4806–4817 (2016)