Semantic Segmentation of PolSAR Images Using Conditional Random Field Model Based on Deep Features

Tao Hu, Wei-hua Li, Xian-xiang Qin*

Information and Navigation College, Air Force Engineering University, Xi’an Shanxi 710077, China
hu1862965@163.com

Abstract. Aiming at the problem that the representation ability of traditional features is weakly, this paper proposes a semantic segmentation method based on deep convolutional neural network and conditional random field. The pre-trained VGG-Net-16 model is used to extract more powerful image features, and then the semantic segmentation of images is achieved through the efficient use of multiple features and context information by conditional random fields. The experimental results show that compared with the three methods using traditional classical features, the method achieves the highest overall classification accuracy and Kappa coefficient, indicating that VGG-Net-16 can extract more effective features.

1. Introduction

Polarimetric synthetic aperture radar (PolSAR) provides rich information of earth surface under all-weather and day-and-night conditions. PolSAR semantic segmentation, which can be a level of final product for end users and can also be a fundamental procedure to support other applications, receives plenty of researches[1-3]. Traditional methods for semantic segmentation of polarimetric SAR images are mainly achieved through target decomposition and statistical distribution. There are many methods for decomposing polarization data, such as Cloude decomposition [4], Freeman decomposition [5]. Statistical distribution models mainly include Wishart distribution [6], K distribution [7], and so on. Lee combined the objective decomposition and distribution models and then proposed a method that effectively improves semantic segmentation accuracy [6]. However, this kind of method does not consider the context information of the image and is susceptible to speckle noise. Therefore, many researchers began to pay attention to the semantic segmentation method using contextual information [1, 8]. The literature [1] uses context information through the CRF on the basis of the fusion of polarization features, and can obtain good results of regional consistency. The features used by the above methods mainly include combinational transformation based on polarization matrix, feature parameters and texture features based on the target decomposition theory [9]. These features are usually designed for specific problems and have a high degree of reliance on prior knowledge. In many cases, their ability to represent is often unsatisfactory. Therefore, how to extract more expressive features is the key way to improve the performance of image semantic segmentation methods.

Deep learning technology has made great progress in dealing with various computer vision tasks. Various deep neural network models have been proposed successively [10, 11], of which the Convolutional Neural Network (CNN) is most widely used in image processing. Widely. For the semantic segmentation of images, some scholars have designed a CNN model that directly achieves the semantic segmentation of optical images and has demonstrated excellent performance. Considering the
commonality among different types of images, it can be considered that a CNN front-end network trained with a large amount of data can be used as an effective model for image feature extraction.

Based on this and taking into account the advantages of multi-features and context information utilization of CRF, this paper proposes an image semantic segmentation method that combines pre-training CNN and Conditional Random Field (CRF) models. Firstly, using the classical CNN model-VGG-Net-16 model to extract the deep features of the image, and then use CRF to effectively use multiple features and contextual information to complete the semantic segmentation of the image.

2. Deep CRF model

In recent years, with the rapid development of data sets and the dramatic increase in computer performance, many high-performance convolutional neural network deep architectures have emerged, such as AlexNet [11], VGG-Net [12], and ResNet [14], etc. Each convolution layer of convolutional neural network can obtain different feature expressions of the input image, which avoids the complex feature extraction process in the traditional semantic segmentation algorithm.

Aiming at the problem that the traditional image semantic segmentation method is limited by the incapability of artificial feature representation, we proposed an image semantic segmentation method based on deep CRF model. It uses VGG-Net-16 [12] to extract the image depth feature and train CRF model to achieve image semantic segmentation.

2.1 Polarimetric SAR Image Preprocessing

For polarimetric SAR data, each pixel is represented by the 9-dimensional vector of the T matrix as follows:

$$V = \{ T_{11}, T_{22}, T_{33}, \text{real}(T_{12}), \text{imag}(T_{12}), \text{real}(T_{13}), \text{imag}(T_{13}), \text{real}(T_{23}), \text{imag}(T_{23}) \}$$

(1)

Among them, the real and imag are respectively the realistic and imaginary parts. In order to reduce the influence of speckle noise, the experimental data was processed by Lee filter [15]. In this paper, a pre-trained VGG-Net-16 model is used as a feature extractor on an ordinary optical image. It requires the input as a three-channel normalized image, so the filtered PolSAR image is implemented using the method of [10]. Normalize, then perform Principal Component Analysis (PCA) transform [19] to reduce the dimension of the data, and use the first three principal components as the input image for feature extraction.

2.2 Extraction Deep Features

The VGG-Net-16 model requires the input image size to be 224×224. First, a 12*12 block is taken centered on each pixel of the pre-processed experimental image, and then each block is subjected to bilinear interpolation. Sampling to 224 × 224, using VGG-Net-16 model to extract the depth features as the features of the center pixel.

2.3 CRF model establishment

CRF is a probabilistic graph model, which can be expressed as an undirected graph model $G = \{V, E\}$, where $V$ represents a set of nodes in the graph and $E$ represents a set of undirected edges between nodes. The CRF models the posterior distribution with a Gibbs distribution in the form of

$$P(y | x, \theta) = \frac{1}{Z(x, \theta)} \exp(-\sum_i E(y, x, \theta))$$

(2)
where $\theta = \{\theta_{i}, \theta_{j}\}$ are parameters and $Z$ means the normalization. In a second-order CRF, the energy can be written as

$$E(y, x, \theta) = \sum_{i \in \Omega} \psi_{i}(x_{i}, y_{i}, \theta_{i}) + \sum_{i \in \Omega, j \notin \Omega} \psi_{ij}(x_{i}, x_{j}, y_{i}, y_{j}, \theta_{ij})$$  \hspace{1cm} (3)

where $\Omega$ is the set of all nodes, $N_{i}$ the neighborhood of node $i$, $\psi_{i}$ and $\psi_{ij}$ are the unary and pairwise potentials, both of which depend on the observations as well as the parameter $\theta$.

In this paper, a generalized linear model is used to define the unary potential function [17]

$$\psi_{i}(x_{i}, y_{i}, \theta_{i}) = \exp\{y_{i} \theta_{i}^{T} g(x_{i})\}$$  \hspace{1cm} (4)

where $\theta_{i}$ are the weight corresponding to the feature contained in the feature vector, the value of which is determined in the parameter estimation. $g(x)$ indicates the features of the $i$th point. The pairwise potential function can be expressed as

$$\psi_{ij}(x_{i}, x_{j}, y_{i}, y_{j}, \theta_{ij}) = \exp\{y_{i} \theta_{ij}^{T} u_{ij}(x)\}$$  \hspace{1cm} (5)

$$u_{ij}(x) = |g_{ij}(x) - g_{ij}(x)|$$  \hspace{1cm} (6)

Among them, $u_{ij}(x)$ is the joint feature vector, which represents the influence of the difference between the feature vectors on the label, and $\theta_{ij}$ is the weight of the vector $u_{ij}(x)$.

After determining the CRF model, it is necessary to estimate the model parameters in the training stage. This paper uses the TRW algorithm [17] to estimate the model parameters. Parameter estimation is an iterative process. We can avoid long-term non-convergence by setting the maximum number of iterations. After the parameters are determined, the optimal solution $y^{*}$ that maximizes the posterior probability in equation (2), that is, $y^{*} = \arg \max_{y} \log P(y|x, \theta)$, is searched for under a fixed parameter, thereby determining the category label of each pixel and achieving image semantic segmentation.

3 Experiments

3.1 Flevoland

In order to verify the effectiveness of the proposed algorithm, three classical methods are compared with the textual methods in the experiment: CRF classification based on the features of Cloude decomposition and Freeman decomposition (abbreviated as method 1); Freeman decomposition and CRF classification of features derived from the diagonal elements of the covariance matrix (abbreviated as method 2); the CRF classification of features obtained by serially merging the features of the above two methods (abbreviated as method 3). Table 1 shows the features used in the comparison method.

| Cloude   | Freeman | Covariance matrix diagonal elements |
|---|---|---|
| H, $\alpha$, A, $\lambda_{1}$, $\lambda_{2}$, $\lambda_{3}$ | Ps, Pd, Pv | C11, C22, C33 |

In this paper, we choose to extract the features of FC6 (full connection layer) layer of VGG-Net-16 model for comparison experiments. The effect of different convolutional layer features on the performance of the algorithm will be analyzed in Section 3.2. The feature extraction of the method in this paper is completed on the MatConvNet [18] deep learning platform. All four methods use the same CRF model. The maximum number of iterations in the parameter estimation process is set to 1000. Four
groups of experiments selected the same training and test samples for comparison. The comprehensive evaluation index of classification performance is Overall Accuracy (OA) and Kappa coefficient.

The experimental data in this paper is obtained by the AIRSAR system. Fig. 1(a) is a Pauli RGB synthesis diagram with a size of 1024 x 750 pixels. Includes 11 categories of crops: beans, forests, rapeseed, bare land, potatoes, sugar beets, wheat, peas, alfalfa, grasslands and waters. The ground truth map is shown in Fig. 1(b). Blank areas are unlabeled categories, 10% of marked data is selected for training, and all marked data are used as test data. The experimental results are shown in Fig. 1. Fig. 1 (c), (d), (e) are Method 1, Method 2, Method 3, Fig. 1 (f) is the method of this article.

![Fig. 1. Comparison of Flevoland data classification results](image)

Table 2. The classification accuracy of Flevoland data

| category       | Method 1 | Method 2 | Method 3 | Our Method |
|----------------|----------|----------|----------|------------|
| bean           | 0.971    | 0.833    | 0.967    | 0.977      |
| forest         | 0.759    | 0.940    | 0.733    | 0.920      |
| potatoes       | 0.680    | 0.840    | 0.821    | 0.988      |
| alfalfa        | 0.609    | 0.892    | 0.719    | 0.977      |
| wheat          | 0.934    | 0.881    | 0.864    | 0.800      |
| bare land      | 0.514    | 0.871    | 0.903    | 0.993      |
| sugar beet     | 0.913    | 0.903    | 0.895    | 0.893      |
| rapeseed       | 0.572    | 0.782    | 0.627    | 0.892      |
| pea            | 0.589    | 0.821    | 0.820    | 0.953      |
| grassland      | 0.962    | 0.774    | 0.838    | 0.952      |
| water          | 0.701    | 0.970    | 0.526    | 0.883      |
| OA             | 0.751    | 0.870    | 0.778    | 0.911      |
| Kappa          | 0.720    | 0.854    | 0.752    | 0.901      |

As can be seen from Fig. 1, the proposed method is significantly less misclassified than the other three methods of comparison. Table 2 shows the quantitative assessment data. It can be seen from the table that the method has also achieved the best overall classification accuracy of 0.911 and Kappa
coefficient of 0.901. The classification accuracy of all categories is above 0.8, and most of them are above 0.9. And the highest classification accuracy was achieved in alfalfa, wheat, sugar beet, rapeseed, pea, and grassland, which further illustrated the effectiveness of the method.

From the above experimental results, it can be seen that the characterization capability of high-dimensional features obtained by connecting multiple sets of features in series may be lower than that of low-dimensional features. This shows that the extracted high-dimensional features contain redundant uncorrelated information, which leads to the weakening of the classification ability of some feature vectors. This method has achieved the best classification results, indicating that CNN features have stronger representation ability than traditional features. Using CNN features can effectively improve the classification performance.

3.2 Comparison of feature layers
To compare which features in the VGG-Net-16 model, we extracted FC7, FC6, Conv5-3, Conv4-3, Conv3-3, Conv2-2, and Conv1-2 features and tested the Overall Accuracy. Table 3 shows the results of testing under the Flevoland data set.

| layer          | Conv1-2 | Conv2-2 | Conv3-3 | Conv4-3 | Conv5-3 | FC6  |
|----------------|---------|---------|---------|---------|---------|------|
| OA             | 0.751   | 0.822   | 0.861   | 0.869   | 0.875   | 0.911|
|                |         |         |         |         |         | 0.900|

From Table 3, it can be seen that under the Flevoland data, with the increase of the depth of the convolution layer, the classification accuracy is on the rise and reaches the highest on the FC6 layer. This is because the deeper features in the VGG-Net-16 model are more abstract and have a higher level of semantic information. The classification accuracy of the FC7 layer features is slightly lower than FC6. This is because the VGG-Net-16 model is trained on the ImageNet dataset. Although the FC7 layer features are more semantic, they are more in line with the training set's classification properties, so the classification accuracy is lower. Therefore, in the method of this paper, the VGG-Net-16 model feature extraction layer selects the FC6 layer.

4 Conclusion
This paper presents a PolSAR image semantic segmentation method based on deep convolutional neural network and conditional random fields. This method uses the convolutional neural network to extract the depth features, and then realizes the PolSAR image semantic segmentation through the efficient use of conditional random fields for multiple features and context information. The experimental results show that the FC6 layer is the most effective feature extraction layer when using VGG-Net-16 model extraction features for image semantic segmentation. In addition, compared with the three methods using traditional classical features, the segmentation results with the highest precision are obtained in this paper, demonstrating the effectiveness of the proposed method.

Acknowledgement
This work is supported by National Science Foundation of China under Grant 41601436, 61403414 and 61703423, and in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2018JM4029.

References
[1] Yang W, ZHANG X, CHEN L et al. (2010) Semantic segmentation of Polarimetric SAR imagery using Conditional Random Fields. In: proceedings of the Geoscience and Remote Sensing Symposium. New York. pp. 1138-1142.
[2] Zhou Y, Wang H, Xu F, et al. (2017) Polarimetric SAR Image Classification Using Deep Convolutional Neural Networks. IEEE Geoscience & Remote Sensing Letters, Commun, 13: 1935-9.

[3] Chen S W, Tao C S. (2018) PolSAR Image Classification Using Polarimetric-Feature-Driven Deep Convolutional Neural Network. IEEE Geoscience & Remote Sensing Letters, Commun, 127: 1-5.

[4] Cloude S R, Pottier E. (1997) An Entropy Based Classification Scheme for Land Applications of Polarimetric SAR. IEEE Trans. Geosci. Remote Sens., Commun, 35:68-78.

[5] Zhao L W, Zhou X G, Jiang Y M, et al. (2007) Iterative Classification of Polarimetric SAR Image Based on the Freeman Decomposition and Scattering Entropy. Synthetic Aperture Radar. In: Asian and Pacific Conference on. IEEE, Japan. pp.473-476.

[6] Lee J S, Grunes M R. (1992) Classification of Multi-Look Polarimetric SAR Data Based on Complex Wishart Distribution. International Journal of Remote Sensing, Commun, 15:2299-2311.

[7] Beaulieu J M, Touzi R. (2004) Segmentation of Textured Polarimetric SAR Scenes by Likelihood Approximation. IEEE Transactions on Geoscience & Remote Sensing, Commun, 42:2063-2072.

[8] Wu Y, Ji K, Yu W, et al. (2008) Region-Based Classification of Polarimetric SAR Images Using Wishart MRF. IEEE Geoscience & Remote Sensing Letters, Commun, 5:668-672.

[9] Zheng Y, Li Z, Zhang C. (2016) A Hybrid Architecture Based on CNN for Image Semantic Annotation. IEEE, Commun, 15:348-356.

[10] Krizhevsky A, Sutskever I, Hinton G E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. International Conference on Neural Information Processing Systems. Curran Associates Inc. pp:1097-1105.

[11] Szegedy C, Liu W, Jia Y, et al. (2015) Going Deeper with Convolutions. IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp:1-9.

[12] Simonyan K, Zisserman A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science.

[13] Li S Z. (1995) Markov Random Field Modeling in Computer Vision. Springer Japan, pp. 788-796.

[14] He K, Zhang X, Ren S, et al. (2016) Deep Residual Learning for Image Recognition. In: Computer Vision and Pattern Recognition. IEEE, pp:770-778.

[15] Han P, Han B. (2018) Lee Filter of PolSAR Image Based on Typical Scattering Difference Index. Systems Engineering & Electronics, Commun, 77:1718-1723.

[16] Mika S, Smola A, Scholz M. (1999) Kernel PCA and De-noising in Feature Spaces. In: Conference on Advances in Neural Information Processing Systems II. pp:536-542.

[17] Domke J. (2013) Learning Graphical Model Parameters with Approximate Marginal Inference. IEEE Transactions on Pattern Analysis & Machine Intelligence, Commun, 35:2454-2467.

[18] Vedaldi A, Lenc K. (2014) MatConvNet: Convolutional Neural Networks for MATLAB. Commun, 17:689-692.