Complex-valued U-Net for PolSAR Image Semantic Segmentation

Ao Liu1, Lingjuan Yu1*, Zhaoxin Zeng1, Xiaochun Xie2, Yuting Guo1, Qiqi Shao1

1School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou City, Jiangxi Province, 341000, China
2School of Physics and Electronic Information, Gannan Normal University, Ganzhou, Jiangxi Province, 341000, China.

* Corresponding author: yulingjuan@jxust.edu.cn

Abstract. As an image semantic segmentation network, U-Net has the advantage of simple structure which is suitable for semantic segmentation of PolSAR images with small datasets. However, the original U-Net is a real-valued (RV) network, whose input must be RV. If it is directly used in the segmentation of PolSAR image, the complex-valued (CV) input must be converted into RV, which results in the loss of information. In this paper, a CV U-Net, which is mathematically strict, is proposed for semantic segmentation of PolSAR images. Considering that the PolSAR dataset is small, the structure and parameters of CV U-Net are furtherly simplified based on the original U-Net to prevent overfitting. Experimental results on the Flevoland dataset show that the proposed CV U-Net has better segmentation performance than the original RV U-Net and some other semantic segmentation networks.

1. Introduction
In recent years, with the development of deep learning, image semantic segmentation has achieved remarkable results. The commonly used optical image semantic segmentation networks are FCN [1], U-Net [2], PSPNet [3], DeepLab [4-6], etc. Among them, U-Net has a simple structure, which is suitable for the semantic segmentation of remote sensing images with small datasets. Zhang et al. designed three U-Nets, namely, Normal-U-Net, Residual-U-Net, and Inception-U-Net, and combined them to form a strong model for building extraction. Experimental results showed that the combined model achieved a remarkable performance [7]. Chen et al. improved both U-Net and FCN to form two new modules respectively, and then fused these two modules for island remote sensing image segmentation. Experimental results showed that this network obtained good segmentation results [8].

Except for the optical remote sensing images, U-Net is also applied in the semantic segmentation of SAR remote sensing images [9]. Unlike the optical remote sensing images, SAR images contain both amplitude and phase information. However, the original U-Net is a real-valued (RV) network, and the input of this network is also RV. If the amplitude information of SAR images is taken as the input of the real-valued (RV) network, the phase information will be lost. In order to make full use of amplitude and phase information, Chen et al. proposed a complex-valued (CV) SegNet to achieve the segmentation of building targets in SAR images [10,11]. However, this network mainly uses real and imaginary channels to form a CV semantic segmentation network, and the CV pooling, CV activation function, and CV batch normalization are not strict complex-valued operations in the mathematical sense. In this paper, a CV U-Net, which is mathematically strict, is proposed for semantic segmentation of PolSAR images.
The input of this network is the CV coherence matrix of PolSAR data. Considering that the PolSAR data set is small, the structure and parameters of CV U-Net are furtherly simplified based on the original U-Net to prevent overfitting during the training of the network.

The rest of this paper is organized as follows. Section 2 illustrates the framework of the proposed CV U-Net and presents the mathematical operation involved in this network. Experimental results on the Flevoland dataset and discussion are shown in Section 3. Section 4 concludes this paper.

2. METHODOLOGY

In this section, the framework of CV U-Net is illustrated in detail, and then the mathematical operations involved in the proposed network are presented.

2.1. Framework of CV U-Net

Based on the original U-Net, the framework of CV U-Net is shown in Figure 1. It mainly includes two parts: encoder and decoder. The encoder can be divided into three convolution blocks. The first convolution block includes two 3×3 CV convolution operations with stride 1 and one 3×3 CV convolution operation with stride 2. The second convolution block is the same as that of the first convolution block. The third convolution block includes three 3×3 convolution operations with stride 1, and one 3×3 CV convolution operation with stride 2. Correspondingly, the decoder also includes three deconvolution blocks. The first deconvolution block includes one upsampling and CV convolution operation, one copy and concatenates operation, and three 3×3 CV convolution operations with stride 1. The second deconvolution block includes one upsampling and CV convolution operation, one copy and concatenates operation, and two 3×3 CV convolution operations with stride 1. The third deconvolution block includes the same structure as the second deconvolution module and one additional 1×1 CV convolution. After each CV convolution in the encoder and decoder, an activation function (i.e. CV ReLU) is used in increasing the nonlinearity of the CV network [12], and CV batch normalization is used in accelerating the convergence speed and reducing the risk of gradient dispersion [13].

As shown in Figure 1, the input of the encoder is a 6-channel CV coherence matrix. The output of the first convolution block is a 32-channel 32×32 feature map; the output of the second convolution
block is a 64-channel 16×16 feature map; the output of the third convolution block is a 128-channel 8×8 feature map. In the decoder, the output of the first deconvolution block is a 256-channel 16×16 reconstructed feature map; the output of the second convolution block is a 128-channel 32×32 reconstructed feature map; the output of the third deconvolution block is 64-channel 64×64 reconstructed image. Finally, the magnitude and softmax operation is implemented to obtain the final segmentation result of the PolSAR image.

2.2. Related mathematical operations

CV convolution with any stride \( s (s \geq 1) \): Because the input feature maps of the \( l \)-th \((l \geq 1)\) layer are also the output feature maps of the \((l-1)\)-th layer, the input feature map of the \( l \)-th CV convolution layer is denoted as \( O_{l-1}^{-1} \) \((i=1,2,\ldots, I)\), where \( I \) is the number of CV input channels. The CV convolution kernel with size \( K_1 \times K_2 \) is denoted as \( W_{ji} \) \((j=1,2,\ldots, J)\), where \( J \) is the number of CV convolution kernels. Then the \( j \)-th CV output feature map \( O_{l}^{j} \) can be obtained by,

\[
O_{l}^{j}(x, y) = \sigma \left( \sum_{i=1}^{I} \sum_{u=0}^{K_{1}-1} \sum_{v=0}^{K_{2}-1} W_{ji}^{l}(u, v) O_{l-1}^{-1}(x s + u, y s + v) + b_{j}^{l} \right) = \sigma \left( Z^{l}(x, y) \right)
\]

where \( Z \) represents the convolution result, \((x, y)\) represents the location of one pixel in a feature map, and \((u, v)\) represents the location of one pixel in a convolutional kernel. \( \sigma(*) \) represents the CV activation function. \( b_{j}^{l} \) is the \( j \)-th CV bias.

Upsampling: The upsampling operation is always used in enlarging the input feature maps. Take the \( l \)-th layer into consideration. Suppose each pixel of the input feature map is upsampled into a patch with size \( G \times G \), then the output feature map can be calculated by,

\[
O_{l}^{i}(xG + m, yG + n) = O_{l-1}^{-1}(x, y)
\]

where \( m \in [0, G-1] \), \( n \in [0, G-1] \).

Magnitude operation: Magnitude operation is used in converting CV numbers into RV numbers before softmax classification of CV U-Net. Suppose the number of categories is \( K \), then the number of input channels of magnitude operation is also \( K \). The magnitude of the \( k \)-th \((k=1, 2, \ldots, K)\) output channel can be calculated by,

\[
O_{k}^{l}(x, y) = \sqrt{\left| R \left( O_{k}^{l-1}(x, y) \right) \right|^2 + \left| I \left( O_{k}^{l-1}(x, y) \right) \right|^2}
\]

Softmax operation: After the magnitude operation, softmax operation can be implemented. The output of the softmax classifier can be calculated by,

\[
p_{k}(x, y) = \frac{\exp \left( O_{k}^{l}(x, y) \right)}{\sum_{m=1}^{K} \exp \left( O_{m}^{l}(x, y) \right)}
\]

where \( p_{k}(x, y) \) denotes the probability belonging to the \( k \)-th class for one pixel located in \((x, y)\).

3. Experimental results & Discussion

In this section, the experimental data set is introduced firstly. Then the data preprocessing is given. Finally, experiments are carried out on the Flevoland dataset and the results are also analyzed.

3.1. Flevoland dataset

The Flevoland dataset is an L-band full PolSAR dataset, which is acquired by the NASA/Jet Propulsion Laboratory AIRSAR platform in 1989 during the MAESTRO-1 Campaign. It is widely used as a benchmark data for PolSAR image classification. The RGB image formed by Pauli decomposition is
shown in Figure 2(a), and its size is 1024×750. The ground truth and the legend are shown in Figure 2(b). There is a total of 15 categories of land covers, including stembeans, peas, forest, lucerne, three types of wheat, beet, potatoes, bare soil, grass, rapeseed, barley, water, and a small number of buildings.

3.2. Data preprocessing
The coherency matrix $\mathbf{T}$ of Flevoland dataset is chosen as the input of CV U-Net. It is a 6-channel data which can be written as:

$$\{T_{11}, T_{22}, T_{33}, T_{12}, T_{23}, T_{13}\}$$

(5)

where each channel with size 1024×750. For RV networks, another 6-channel data is used as the input, which is also related with the matrix $\mathbf{T}$ [14].

The data preprocessing steps are as follows. The first step is to expand 6-channel input data to be with size 1024×832 in mirror mode. The second step is to cut the expanded 6-channel data by 64 sliding windows without overlapping. The third step is to randomly select 40% of the 64×64 samples as training samples and the remaining 60% as testing samples. The last step is to expand the number of training samples by enlarging and rotating.

![Figure 2. Flevoland dataset: (a) RGB image based on Pauli decomposition, (b) Ground truth and legend](image)

3.3 Experimental results and discussion
To verify the effectiveness of CV U-Net, other four classic RV networks are used for comparison. They are FCN [1], PSPNet [3], DeepLabv3+ [6], and U-Net [2]. The semantic segmentation results by using five networks are shown in Figure 3(a)-(e) respectively. It is worth noting that there are segmentation results of both training and testing samples in these figures. In addition, the intersection over union (IOU), mean intersection over union (MIOU), overall accuracy (OA), and mean pixel accuracy (MPA) is used in evaluating the segmentation performance of testing samples, which are listed in Table 1. As shown in Figure 3, wheat 1 in the red oval area is misclassified into rapeseed or wheat 2 by using the RV network, and grass in the black box is misclassified into lucerne or wheat 3 by using the RV network. However, these regions are correctly classified by using the proposed CV network. The reason is that the different phase information of land covers is helpful in the segmentation. Furthermore, land covers in the blue oval area shown in Figure 3(d) and (e) are different. CV U-Net can obtain stronger regional continuity than U-Net.

From Table 1, the IOU of wheat 1 obtained by each RV network is lower than 80%, while it is 91.03% by using CV U-Net. Besides, the IOU of grass is very low by using RV networks. The maximum of IOU is 57.91%, while the minimum is 0.81%. However, the IOU of grass obtained by the CV network is 98.38%. These conclusions are consistent with those obtained from the figures. In addition, MIOU, OA, and MPA obtained by CV U-Net are 96.72%, 99.27%, and 98.25% respectively. They are higher than those obtained by RV networks. Therefore, the segmentation performance of CV U-Net is better than that of the other four RV networks.
Figure 3. Segmentation results of Flevoland dataset: (a) FCN, (b) PSPNet, (c) DeepLabv3+, (d) U-Net, (e) CV U-Net

Table 1. Segmentation performance of Flevoland dataset

| Methods/Class | FCN  | PSPNet | DeepLabv3+ | U-Net | CV U-Net |
|---------------|------|--------|------------|-------|----------|
| Stembeans     | 85.84| 87.69  | 83.23      | **94.15** | 90.25    |
| Peas          | 73.45| 74.95  | 59.72      | 93.80 | **95.87** |
| Forest        | 75.22| 76.40  | 71.28      | **95.60** | 94.29    |
| Lucerne       | 75.10| 71.68  | 77.82      | 91.89 | **99.13** |
| Wheat 1       | 77.85| 76.86  | 79.31      | 78.66 | **91.03** |
| Beet          | 76.58| 47.51  | 49.24      | 91.84 | **98.47** |
| Potatoes      | 67.02| 68.67  | 66.28      | 95.85 | **96.49** |
| Bare soil     | 43.70| 39.87  | 73.11      | 95.03 | **99.78** |
| Grass         | 57.91| 47.48  | 26.02      | 0.81  | **98.38** |
| Rapeseed      | 43.15| 38.52  | 44.46      | 58.03 | **97.71** |
| Barley        | 97.44| 91.26  | 83.90      | 80.93 | **99.09** |
| Wheat 2       | 46.85| 35.06  | 51.80      | 51.82 | **98.55** |
| Wheat 3       | 86.40| 70.58  | 71.50      | 84.64 | **95.01** |
| Water         | 91.94| 88.06  | 93.57      | 98.10 | 96.38    |
| Buildings     | 55.19| 63.00  | 47.16      | 67.32 | **97.35** |
| MIOU (%)      | 71.80| 67.21  | 67.16      | 79.90 | **96.72** |
| OA (%)        | 93.36| 94.85  | 94.32      | 97.76 | **99.27** |
| MPA (%)       | 82.12| 76.74  | 80.22      | 85.37 | **98.25** |

4. Conclusions
In this paper, a CV U-Net network is proposed for PolSAR image segmentation. The framework of CV U-Net also includes an encoder and decoder, which is the same as the original U-Net. However, the structures and parameters of CV U-Net are simplified based on the original U-Net. The mathematical operations of CV U-Net mainly include CV convolution, upsampling, magnitude operation, and softmax operation, which are presented in detail respectively. Experimental results on the Flevoland dataset show
that CV U-Net can obtain better segmentation performance than the classic RV FCN, PSPNet, DeepLabv3+, and U-Net.

**Acknowledgment**

This work was supported by the National Natural Science Foundation of China (61501210), and the Science and Technology Project of Jiangxi Provincial Education Department (190459).

**References**

[1] Long, J., Shelhamer, E., and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition. Boston. pp. 3431-3440.

[2] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[3] Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017) Pyramid Scene Parsing Network. In Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu. pp. 2881-2890.

[4] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2014) Semantic image segmentation with deep convolutional nets and fully connected CRFs. Computer Science, 2014(4):357-361.

[5] Chen, L.C., Papandreou, G., Kokkinos I, et al. (2018) DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4):834-848.

[6] Chen, L. C., Papandreou, G., Schroff, F., and Adam, H. (2017) Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

[7] Zhang W., Tang P., Zhao L. and Huang Q., (2019) A Comparative Study of U-Nets with Various Convolution Components for Building Extraction, Joint Urban Remote Sensing Event (JURSE), Vannes, pp. 1-4.

[8] Chen T., Wang H., Liu H., Du J. and Wu P., (2020) An Island Remote Sensing Image Segmentation Algorithm Based on FC_U-Net Network," International Conference on Computer Vision, Image and Deep Learning (CVIDL), Chongqing, pp. 100-105.

[9] Pai, M., Mehrotra, V., Aiyar, S., Verma, U., and Pai, R. M. (2019) Automatic Segmentation of River and Land in SAR Images: A Deep Learning Approach. IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Sardinia, pp. 15-20.

[10] Chen, J. and Qiu, X., (2019) Equivalent Complex Valued Deep Semantic Segmentation Network for SAR Images, International Applied Computational Electromagnetics Society Symposium (ACES), Miami, pp. 1-2.

[11] Chen J, Peng L., Qiu X., et al. (2019) A 3D building reconstruction method for SAR images based on deep neural network. Scientia Sinica Informationis, 49(12):1606-1625.

[12] Yu, L., Hu, Y., Xie, X., Lin, Y., and Hong, W., (2020) Complex-valued full convolutional neural network for SAR target classification. IEEE Geoscience and Remote Sensing Letters, 17(10): 1752-1756.

[13] Trabelsi, C., Bilaniuk, O., Zhang, Y., et al. (2017) Deep complex networks. arXiv preprint arXiv:1705.09792.

[14] Zhou, Y., Wang, H., Xu, F., et al. (2016) Polarimetric SAR image classification using deep convolutional neural networks. IEEE Geoscience and Remote Sensing Letters, 13(12): 1935-1939.