

EXPLORING CONVOLUTIONAL LSTM FOR POLSAR IMAGE CLASSIFICATION

Lei Wang¹, Xin Xu^{1*}, Hao Dong¹, Rong Gui¹, Rui Yang¹ and Fangling Pu^{1,2}

¹ School of Electronic Information, Wuhan University, Wuhan 430079, China;

² Collaborative Innovation Center of Geospatial Technology,
Wuhan University, Wuhan 430079, China;

ABSTRACT

Polarimetric synthetic aperture radar (PolSAR) image classification is one of the most important applications in PolSAR image processing. More and more deep learning methods are applied to PolSAR image classification. As we know, the polarimetric response of a target is related to the orientation of the target, but the features in rotation domain are not fully used in deep learning. We use a convolutional LSTM (ConvLSTM) along with a sequence of polarization coherent matrices in rotation domain for PolSAR image classification. First, nine different polarization orientation angles (POA) are used to generate nine polarization coherent matrices in rotation domain. Second, a deep learning model that stacked with multiple ConvLSTM layers and fully connected layers is proposed for classification. Finally, the sequence of polarization coherent matrices is fed into the ConvLSTM to classify PolSAR images. Experiments show that the classification results of ConvLSTM are better than the LeNet-5.

Index Terms— PolSAR, classification, convolutional LSTM, rotation domain

1. INTRODUCTION

PolSAR has the advantages of working under all weather conditions, large scope and certain penetrating capacity, so it is widely used in remote sensing areas. PolSAR image classification is one of the most important applications. It plays an important role in urban planning, agriculture, disaster prevention and so on.

Because of the superior performance of deep learning [1], many deep learning methods are applied to PolSAR image classification [2, 3, 4]. Compared with the traditional statistical modeling [5] and non-neural machine learning [6] methods, deep learning can achieve better performance [2, 4, 7]. Many deep learning methods use the Pauli coded features or polarization coherent matrix as the input [2, 3], but the hidden polarimetric features in rotation domain are not fully used. Scattering properties of a target is closely relate to the target orientation. The polarimetric features in rotation domain can provide discriminative information for land cover classification. Some deep learning methods [4] used the selected

polarimetric features proposed by the uniform polarimetric matrix rotation theory [8] for PolSAR image classification. These methods also achieved great results but the features from some other POAs were ignored. To make full use of the features in rotation domain, we rotate the POA from 0 to $2/\pi$ to generate a sequence of polarization coherent matrices. The sequence of polarization coherent matrices can be used as the input of a deep learning model. Because the input is sequence data, the long short-term memory (LSTM) [9, 10] models have more advantages than other deep learning models. PolSAR images have rich spatial information, but no spatial information is encoded by LSTM. [11] proposed a convolution LSTM (ConvLSTM) designed specifically for sequence data with spatial structure information. ConvLSTM is now widely used in video processing [12]. Hence, we use the ConvLSTM along with the polarization coherent matrices in rotation domain for PolSAR image classification.

The organization of this paper is as follows. Section 2 introduces the polarization coherent matrices in rotation domain and the ConvLSTM briefly. Then the structure of ConvLSTM for PolSAR classification is illustrated in detail. In section 3, the classification results of two real PolSAR images are evaluated. Finally, the conclusion is discussed in Section 4.

2. RELATED WORKS

2.1. Polarization coherent matrices in rotation domain

PolSAR image can be expressed with the 3×3 polarization coherent matrix T which is symmetrical. Then the polarization coherency matrix T with the rotation matrix $R_3(\theta)$ [8] becomes

$$T(\theta) = R_3(\theta)TR_3^H(\theta) \quad (1)$$

where θ denotes the POA.

The elements of $T(\theta)$ are shown in Equation 2 [8], where $Re[T_{ij}]$ and $Im[T_{ij}]$ are the real and imaginary parts of T_{ij} . We change the θ from 0 to $\pi/2$ with the step $\pi/18$, and then a sequence of polarization coherent matrices in rotation domain is generated. The generated 9 polarization coherent matrices can be regarded as the time steps of ConvLSTM.

$$\begin{aligned}
T_{11}(\theta) &= T_{11} \\
T_{12}(\theta) &= T_{12} \cos 2\theta + T_{13} \sin 2\theta \\
T_{13}(\theta) &= -T_{12} \sin 2\theta + T_{13} \cos 2\theta \\
T_{23}(\theta) &= \frac{1}{2}(T_{33} - T_{22}) \sin 4\theta + \operatorname{Re}[T_{23}] \cos 4\theta + j \operatorname{Im}[T_{23}] \\
T_{22}(\theta) &= T_{22} \cos^2 2\theta + T_{33} \sin^2 2\theta + \operatorname{Re}[T_{23}] \sin 4\theta \\
T_{33}(\theta) &= T_{22} \sin^2 2\theta + T_{33} \cos^2 2\theta - \operatorname{Re}[T_{23}] \sin 4\theta
\end{aligned} \tag{2}$$

For each pixel, the polarimetric feature can be defined as $\{T_{11}(\theta), T_{22}(\theta), T_{33}(\theta), \operatorname{Re}[T_{12}(\theta)], \operatorname{Im}[T_{12}(\theta)], \operatorname{Re}[T_{13}(\theta)], \operatorname{Im}[T_{13}(\theta)], \operatorname{Re}[T_{23}(\theta)], \operatorname{Im}[T_{23}(\theta)]\}$. For pixel-by-pixel classification, each pixel is fully represented by a local patch defined by a neighborhood window of size $w \times w$. Because the local patch is generated from the 9 polarization coherency matrices in rotation domain, the dim of the patch is $9 \times 9 \times w \times w$, where the first 9 is the number of time steps and the second 9 is the number of channels. We use the patch as the input of ConvLSTM (See Figure 1).

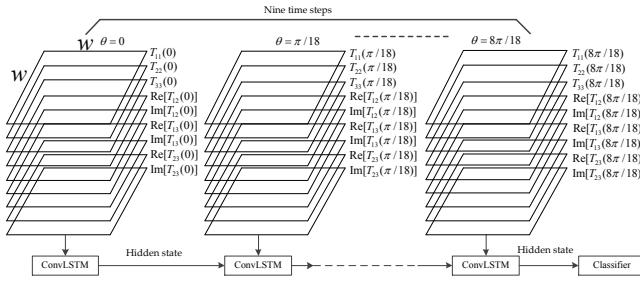


Fig. 1. The input of ConvLSTM.

2.2. Convolutional LSTM

LSTM as a special recurrent neural network (RNN) structure has proven stable and powerful for modeling long-range dependencies in various studies [9, 10], but it contains too much redundancy for spatial data. To address this problem, ConvLSTM replaces the fully connected gate layers of the LSTM with convolutional layers. ConvLSTM is capable of encoding spatio-temporal information in its memory cell. The key equations of ConvLSTM [11] are shown in below, where '*' denotes the convolution operation and 'o' denotes the Hadamard product:

$$\begin{aligned}
i_t &= \sigma(W_{xi} * \chi_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\
f_t &= \sigma(W_{xf} * \chi_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \\
C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * \chi_t + W_{hc} * H_{t-1} + b_c) \\
o_t &= \sigma(W_{xo} * \chi_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \\
H_t &= o_t \circ \tanh(C_t)
\end{aligned} \tag{3}$$

PolSAR images are full of spatial information, so ConvLSTM is suitable for PolSAR image classification using the sequence of polarization coherent matrices.

2.3. ConvLSTM for PolSAR image classification

Figure 2 is the structure of ConvLSTM for PolSAR classification, where T is the number of time steps, C is the number of channels, w is the size of local patches introduced in section 2.1 and class_num is the number of land covers. In this paper, T is 9, C is 9 and w is 15. The proposed ConvLSTM model has four ConvLSTM layers, four batch normalization layers, two fully connected layers and a softmax layer.

The sequence of polarization coherent matrices is input to multiple ConvLSTM layers directly. The hidden state of the fourth ConvLSTM layer in the final time step contains the representation of the input. This representation is then applied to two fully connected layers. Finally the output feature of the second fully connected layer is input to the softmax layer for classification.

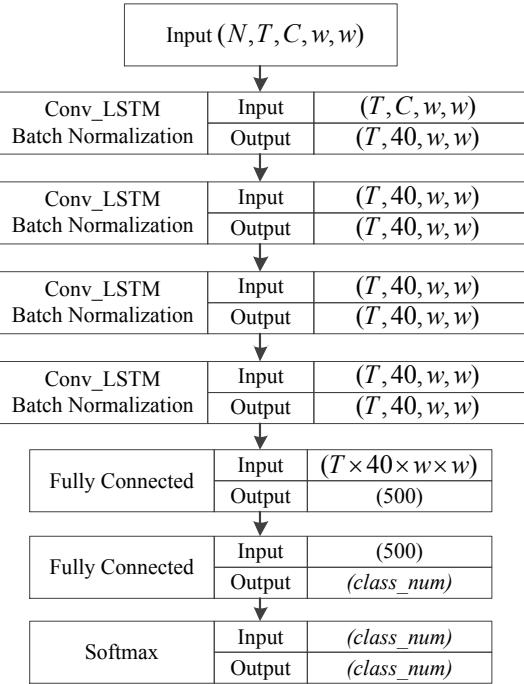


Fig. 2. The structure of ConvLSTM for PolSAR Classification.

2.4. Materials

Two real PolSAR datasets are used to verify the performance of ConvLSTM. The detailed data information is presented as follows.

1. RADARSAT-2 Flevoland dataset

The RADARSAT-2 (RS-2) Flevoland spaceborne dataset is acquired by the C-band RADARSAT-2 PolSAR system at

fine quad-pol mode. It is over Flevoland in the Netherlands, with an image size of 1400×1200 pixels. The spatial resolution is 12 m. A total of four classes are identified, consisting of water, forest, farmland, and buildings. Figure 3(a) shows the Pauli RGB image. Figure 3(b) shows the ground truth map, which was manually created based on very high resolution optical images.

2. AIRSAR Flevoland dataset

The airborne dataset is the NASA/JPL AIRSAR L-band four-look fully polarimetric data. The Pauli RGB image is shown in Figure 3(c). This scene also covers over Flevoland, the Netherlands, with an image size of 750×1024 pixels and a resolution of 12×6 m. Since [14], this dataset is widely used in land use classification with the well-established ground truth map, which is shown in Figure 3(d). A total of 11 classes are identified, consisting of eight crop classes, and three other classes of bare soil, water, and forest.

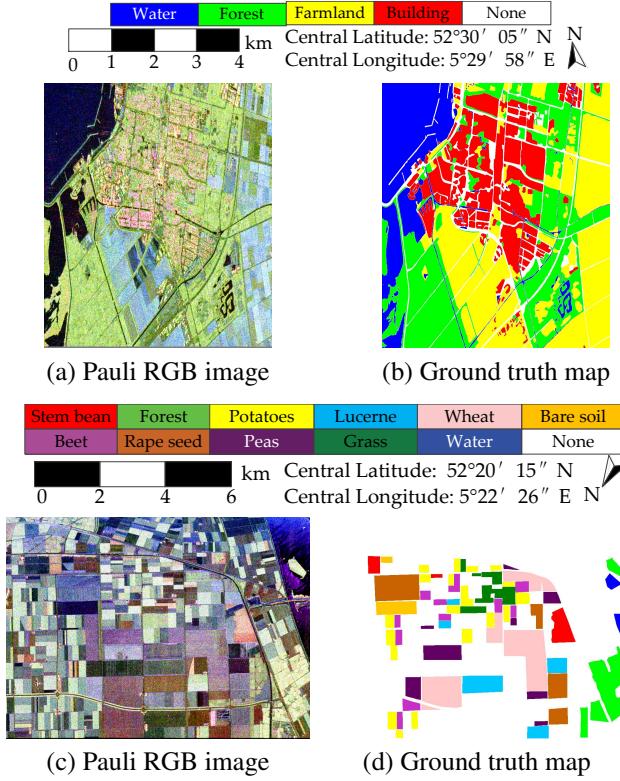


Fig. 3. The Pauli RGB images and ground truth maps of RS-2 Flevoland and AIRSAR Flevoland.

For an entire PolSAR image, the sliding window of $w \times w$ is used to generate a significant amount of image patches which can serve as the training samples. For each dataset, the training samples are randomly selected from those generated patches. For RS-2 Flevoland, the number of total patches is 1,680,000 and the number of training patches is 50,400. For AIRSAR Flevoland, the number of total patches is 768,000 and the number of training patches is 38,390.

Table 1. The classification accuracies of RS-2 Flevoland (%).

Land cover	SVM	LeNet-5	LeNet-5-S	ConvLSTM
Water	95.98	96.00	98.57	98.53
Forest	84.01	89.54	87.88	88.70
Building	82.78	89.93	90.31	91.44
Farmland	87.38	90.18	90.75	91.95
OA	86.96	90.96	91.20	92.06

3. RESULTS

The support vector machine(SVM) and LeNet-5[13] are used for comparison. The input of SVM is the original polarization coherent matrix T with the size of $9 \times w \times w$. No data in rotation domain is used. The LeNet-5 has two kinds of inputs. The first kind of input is the same as SVM. The second kind of input is the sequence of polarization coherent matrices, which is the same as ConvLSTM. The nine polarization coherent matrices are stacked to generated a new data matrix with the size of $81 \times w \times w$. The LeNet-5 with sequence input is called LeNet-5-S for short. The overall accuracy (OA) is used to judge the performances of the methods.

3.1. Results of RADARSAT-2 Flevoland

The Figure 4 and Table 1 are the classification results and accuracies of RS-2 Flevoland. The OA of SVM is much lower than the other methods. When the input of LeNet-5 has no rotation domain data, the OA is lower then the OA of LeNet-5-S. The OA of ConvLSTM is the highest.

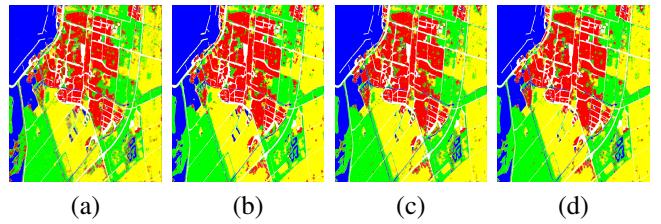


Fig. 4. Classification results of RS-2 Flevoland. (a) SVM. (b) LeNet-5. (c) LeNet-5 with sequence input. (d) ConvLSTM with sequence input.

3.2. Results of AIRSAR Flevoland

The classification results and accuracies of AIRSAR Flevoland are shown in Figure 5 and Table 2. The OA of SVM is the lowest. Except the forest and beet, the accuracies of other 9 land covers of ConvLSTM are the highest and ConvLSTM achieves the best classification result for AIRSAR Flevoland.

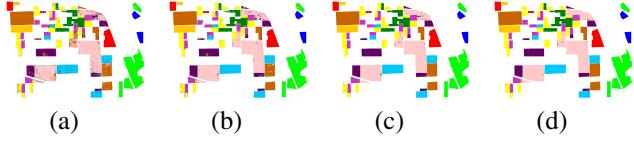


Fig. 5. Classification results of AIRSAR Flevoland. (a) SVM. (b) LeNet-5. (c) LeNet-5 with sequence input. (d) ConvLSTM with sequence input.

Table 2. The classification accuracies of AIRSAR Flevoland (%).

Land cover	SVM	LeNet-5	LeNet-5-S	ConvLSTM
Water	99.99	99.73	100.00	100.00
Forest	96.95	97.96	99.69	99.36
Stem bean	98.50	98.80	99.69	99.88
Potatoes	91.29	97.68	98.37	99.52
Lucerne	91.03	98.53	99.28	99.62
Wheat	86.24	97.23	98.12	98.73
Bare soil	98.67	99.01	99.97	100.00
Beet	92.85	91.93	98.93	98.46
Rape seed	93.08	97.02	98.99	99.44
Peas	82.35	93.79	99.26	99.48
Grass	93.03	94.32	99.42	99.47
OA	91.30	96.83	98.96	99.28

4. CONCLUSIONS

In this paper, the polarization coherent matrices in rotation domain are used as the input of ConvLSTM for PolSAR image classification. The results of ConvLSTM using the sequence of polarization coherent matrices are the best and surpass both of the SVM and LeNet-5. From the experiments we can get the following two conclusions. First, the polarization coherent matrices in rotation domain can improve the classification results. Second, ConvLSTM is powerful for modeling spatial sequence data and is also suitable for PolSAR image classification.

5. REFERENCES

- [1] G. E. Hinton, and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504-507, 2006.
- [2] F. Gao, T. Huang, J. Wang, J. Sun, A. Hussain, and E. Yang, "Dual-Branch Deep Convolution Neural Network for Polarimetric SAR Image Classification," *Applied Sciences*, vol. 7, 2017.
- [3] Z. Zhang, H. Wang, F. Xu, and Y. Q. Jin, "Complex-Valued Convolutional Neural Network and Its Application in Polarimetric SAR Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1-12, 2017.
- [4] S. W. Chen, and C. S. Tao, "Multi-temporal PolSAR crops classification using polarimetric-feature-driven deep convolutional neural network," *International Workshop on Remote Sensing with Intelligent Processing*, 2017.
- [5] H. Dong, X. Xu, H. Sui, F. Xu and J. Liu, "Copula-Based Joint Statistical Model for Polarimetric Features and Its Application in PolSAR Image Classification," *Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 5777-5789, 2017.
- [6] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, 2015, vol. 61, pp. 85-117.
- [7] J. L. Hu, L. Mou, A. Schmitt and X. X. Zhu, "FusioNet: A Two-Stream Convolutional Neural Network for Urban Scene Classification using PolSAR and Hyperspectral Data," *Urban Remote Sensing Event*, 2017.
- [8] S. W. Chen, X. S. Wang, and M. Sato, "Uniform Polarimetric Matrix Rotation Theory and Its Applications," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 52, pp. 4756-4770, 2014.
- [9] S. Hochreiter, and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, pp. 1735-1780, 1997.
- [10] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," *Neural Information Processing Systems*, vol. 4, pp. 3104-3112, 2014.
- [11] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," *Neural Information Processing Systems*, vol. 9199, pp. 802-810, 2015.
- [12] Z. Li, E. Gavves, M. Jain, and C. G. M. Snoek, "VideoLSTM Convolves, Attends and Flows for Action Recognition," *Computer Vision & Image Understanding*, 2016.
- [13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278-2324, 1998.
- [14] J. S. Lee, M. R. Grunes, and E. Pottier, "Quantitative comparison of classification capability: fully polarimetric versus dual and single-polarization SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 2343-2351, 2001,