Comparison of Real-valued and Complex-valued convolutional networks for TerraSAR-X Patch classification

Dušan Gleich[1], Danijel Šipoš[1], and Peter Planinšič[1]

Abstract – **This paper presents SAR patch categorization using real and complex valued (CV) deep Convolutional Networks (CNN) for categorization of Synthetic Aperture Radar (SAR) data. The categorization of Synthetic Aperture Radar (SAR) patches consists of feature extraction and classification. Over the past few years image categorization using deep learning became very popular, because it can handle large databases and has shown good recognition results. This paper presents deep convolutional networks for Synthetic Aperture Radar patch categorization. We have tested convolutional networks with20 layers. The CV-CNN consist in general of a real or complex valued input layer, output layer and one or more hidden layers. Hidden layers represent any combination of convolutional layers, pooling layers, activation functions, and are fully defined within complex valued domain. The custom database of patches was designed using 3 classes and parameters of CV-CNN were observed in order to achieve the best accuracy results.**

Keywords – **Synthetic Aperture Radar, Convolutional networks, patch categorization, deep learning.**

I. Introduction

Synthetic aperture radar (SAR) is an all-weather, night and day imaging system. SAR data sets are nowadays easily accessible from different airborne or spaceborne sources as it is Sentinel-1 and TerraSAR-X, which are mostly devoted to the wider scientific communities. Nowadays a high-resolution SAR images acquired from a spaceborne platforms can achieve resolution of 10 cm. SAR is particularly suitable for land cover classification, target detection, surveillance, land sliding, soil moisture etc. Because of scattering mechanism and speckle noise in SAR imagery, the interpretation and understanding of SAR images is different from visual photo analysis. The image understanding, and data interpretation of SAR data has been studied over the last few decades. The classical classification consists of statistical feature extraction and classification methods. Recently, classification using stacked neural networks (SNN) or shortly deep learning, which include deep belief network (DBN), convolutional neural network (CNN) and recurrent neural network (RNN) have shown very good recognition results using large databases with many categories.

By introducing a deep learning theory [1] for automatically learning features from a data sets using a multistage approach, a deep learning applications have become very attractive. A conventional neural networks and support vector machine (SVM) need feature extraction process to separate features within feature space, otherwise a deep learning does not need task-specific feature extractors and its because of this capable of learning features automatically from data sets. A deep neural network was applied to remote sensing image processing [2] and classification [3]. CNN as one of the typical deep learning models have achieved impressive performances in various fields [4]. It is difficult to separate classes within SAR images, because SAR images have complex scattering mechanisms and random speckle noise. In [5] a deep neutral network for SAR automatic target recognition (ATR) was applied and achieved the highest accuracy (more than 99%) on classification of ten-class targets compared with other cited methods at the time of this writing.

Some complex valued Convolutional Neural networks are not new and there has been many different approaches to complex-valued classification of real or complex valued problems [6]-[7]. Authors in [6] presented a variation of the CNN model with complex valued input and weights. The complex model as a generalization of the real model was proposed. The first investigation of Complex-Valued Convolutional Neural Networks (CV-CNN) for object recognition on Pol-SAR data was proposed in [8]. An architecture with only one single convolutional layer was used and showed promising results. A CV-CNN was presented in [9], by giving the full deduction of the gradient descent algorithm for training this type of networks. The comparison of CNN in the real valued domain - Real Valued Convolutional Neural Network (RV-CNN) for image classification was extended to the complex-valued domain. A deep complex valued CNN used for classification was applied to the Polarimetric SAR data. A comparison between a RV-CNN and CV-CNN was presented in [7].

This paper presents the deep convolutional network, which uses a real valued and complex valued approach. An influence of convolutional filters, number of layers of class recognition for complex and real valued SAR data was demonstrated.

II. Real-valued convolutional neural networks

Convolutional neural network models are represented by utilizing various layers of the neural networks. A convolutional neural network consists of input layer followed by optional sub-sampling and regularization layers and ending in fully connected layers. The input into convolutional neural network

[1]Authors are with University of Maribor, Faculty of electrical Engineering and Computer Science, Koroška cesta 46, 2000 Maribor, Slovenia, E-mail: dusan.gleich@um.si.

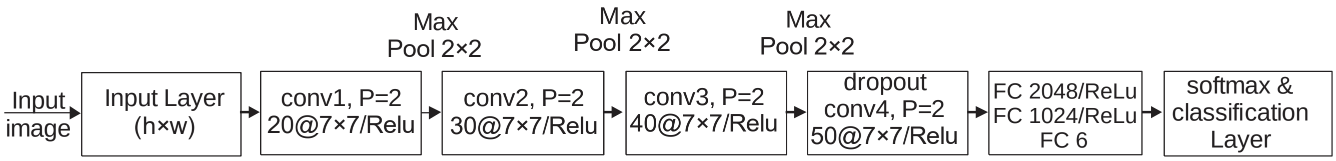

Fig. 1. Convolutional neural network with 20 layers.

is an image which is processed by several filters. The response of the filter is obtained by the convolution, therefore the convolutional response encodes the input and determine features by reducing dimensionality.

The responses of the filters represent inputs to non saturating activation function, which can drastically accelerate learning process. This is done by special rectified linear units (ReLU), which are involved in saturating nonlinearities process. Those functions were applied after every convolutional and fully connected layer. The final layer uses softmax activation to maximize the multinomial logistic regression objective. The result of filtering is usually sub-sampled in order to further reduce dimension of the features makes invariant features to translation. When the max-pool subsampling layers are applied a 2 × 2 max-pool layer divides the output into a set 2×2 cells, which are not overlapped. The maximum activated filter response is recorder into each cell. In this way the input dimensions is reduced by 2 and produced features are invariant to object translations.

The architecture of convolutional neural network is depicted in Fig. 1, which is composed of four convolution layers and three max pooling layers. Each of the first three convolution layers is followed by a max pooling layer, with a pooling size of 2 × 2 and a stride of 2 pixels. The ReLU nonlinearity is applied to every hidden convolution layer. The input image was filtered by 20 convolution filters of size 7×7 in the first convolution layer, resulting in 20 feature maps. The first pooling layer's outputs are sent into the second convolution layer, which has a convolution filter size of 7 × 7, leading to 30 feature maps. The filter size of the third convolution layer is 7 × 7, producing 40 feature maps of size. The fourth convolution layer includes 50 feature maps with a convolution filter size of 7 × 7, which brings out 50 feature maps. The dropout regularization technique is used before fourth convolutional layer. The ReLU nonlinearity was applied after each convolutional layer and after fourth convolutional layer, fully connected layer with dimension of 2048 was inserted before fully connected layers with 1024 and 6 units, respectively. After fully connected layers with 2048 and 1024 units, ReLU layer was applied. At the end of this convolutional network, softmax and classification layers were used. Within this convolutional network a dropout layer was used, which changes architecture and reduces over-fitting. The idea is to connect convolutional and fully connected layers so that hidden neuron outputs are deactivated with probability p during training. This probability was set to p = 0.5. The drop out layer reduces the co-adaptation of neuron. The dropout forces neurons to provide more robust contributions with combination of arbitrary active neuron. The set of neurons is changed randomly in the every epoch and the over-fitting is reduced by 1/(1−p), if it is compared with the network structure without dropout layer.

III. Structure of Complex-Valued Convolutional Neural Networks

In this section the CV-CNN architecture is introduced. Same as with the normal RV-CNN, the CV-CNN is also based on the 2D multichannel input or co-called channel maps. However, the main difference between the two is that each channel array value in CV-CNN is represented in the complex domain as well as hidden layers, which are convolutional filters, pooling filter and activation functions with complex valued inputs and outputs.

The convolutional layer indicates convolution between the sliding window and complex input patch, where the former acts as a filter bank. The result is a matrix where each output value is calculated with a complex dot product sum of the corresponding window and an input patch. Multiple different filter banks are used to search for different features of the specific region in the input patch. This output data can also be interpreted as output maps, which are then further connected to a nonlinear sigmoid or tanh activation function. In this particular approach the sigmoid function was used with the purpose of generating complex feature maps. The convolution result, with included previous layers outputs, can now be described as $O_i^{(l+1)} \in \Im^{W_2 \times H_2 \times I}$ and is calculated by

$$
\begin{aligned}
O_i^{(l+1)} &= \frac{1}{1+exp(-\Re V_i^{(l+1)})} + j\frac{1}{1+exp(-\Im(V_i^{(l+1)}))} \\
V_i^{(l+1)} &= w_{ik}^{(l+1)} \star O_k^{(l)} + b_i^{(l+1)} \qquad (1)
\end{aligned}
$$

where the filter banks are described with $\omega_{ik}^{(l)} \in \Im^{F\times F\times K\times I}$, input feature maps with $O_i^{(l)} \in \Im^{W_1\times H_1\times K}$ and the bias $b_i^{(l+1)} \in \Im^{I}$. Variable l represents the number of current layers, whereas * represents a convolution operator. $V_i^{(l+1)}$ represents weighted sum of inputs to the ith output feature maps in the layer l +1. The convolutional layer is determined by a number of feature maps I, filter size $F \times F \times K$, stride S and zero-padding P. The purpose of the pooling layer is to down-sample the patch resolution. Therefore, it is also known as a subsampling layer. With this task a spatial invariance is achieved, making the network insensitive to small shifts or distortions [11]. This is mostly realized with the subsampling or max pooling function. The subsampling function averages values of the window, where the max pooling function takes the maximum value of the window. An average function was used and restructured for the complex domain as followed

$$O_i^{(l+1)}(x,y) = \underset{u,v=0,\dots,g-1}{\text{ave}} O_i^{(l)}(x \cdot s + u, y \cdot s + v) \quad (2)$$

where *g* is the pooling size and *s* is the stride. (*x*,*y*) indicates the feature map location of *i*th units $O_i^{(l+1)}(x,y)$. The fully connected layer handles the classification after we calculate the final output of several convolutional and pooling layers. For the CV-CNN multiple fully connected layers have been used to connect each neuron with all the neurons in the previous layer [11]. The output can be described as

$$O_i^{(l+1)} = f(\Re(V_i^{(l+1)})) + jf(\Im(V_i^{(l+1)})) \quad (3)$$

$$V_i^{(l+1)} = \sum_{k=1}^{K} \omega_{ik}^{(l+1)} \cdot O_k^{(l)} + b_i^{(l+1)} \quad (4)$$

where *K* represents the number of neurons in the *l*th fully connected layer. In the final stage the result of multiple neurons is inserted to the output layer which then encodes the specific patch classes. Output of this layer is represented as a vector, described with $1 * C$, where *C* defines the number of classes and also represents the length of a vector. Vector also has to be a complex value, therefore scalar 1 from the previous equation has to be replaced with $(1+1 * j)$. The patch belongs to the class whose distance to vector value location is the shortest.

IV. Experimental results

4.1. Dataset

The custom database was designed, which consisted of 10 Single Look Complex data acquired in the Spotlight mode using TerraSAR-X satellite. Data were acquired over different urban, agriculture and forest areas with different incidence angles. A SAR patch data base was designed by an expert, which manually selected several patches per class, where a ground truth was known. 3 classes were selected: C01 Urban areas, C02 Forest, C03 River. 2000 patches per category were selected for training and 1000 patches for testing. The sample of SAR patches of all 3 categories are depicted in Fig. 2. In Figure samples of classes are depicted horizontally, starting with C1, C2 and C3.

4.2. CNN configuration

We designed 3 different configurations of CNN and changed patch size from 12×12, 24×24 and 48×48. The input was a complex valued image, single polarized in complex-valued format. For real valued approach we used only detected or amplitude data.

The layers of a deep convolutional network were learned in the training phase. Three different neural networks were learned, one, by one for three different sizes of input patches. The parameters of a convolutional are depicted in Fig. 1. Within the training process 1000 samples for each class were used and convolutional network was learned using 100 epochs. The goal of the paper was to investigate a structure of a deep neural network and verify its performances.

Tables 1 and 2 report the over all accuracy of the presented methods for 3 classes using a single complex-valued HH polarized datasets. 1000 images were used for testing stage and average accuracy is reported. Tables 1 and 2 show that the best accuracy was achieved using a complex valued approach.

TABLE I
MEAN ACCURACY OF CLASS RECOGNITION IN % FOR PATCH SIZES OF 12×12, 24×24 AND 48×48 USING A COMPLEX VALUED APPROACH

Class	12 × 12	24 × 24	48 × 48
C1	82	84	73
C2	86	89	64
C3	84	87	65
Total(%)	84	86	67

TABLE II
MEAN ACCURACY OF CLASS RECOGNITION IN % FOR PRESENTED METHODS USING REAL VALUED APPROACH.

Class	12 × 12	24 × 24	48 × 48
C1	82	81	74
C2	63	87	57
C3	68	86	63
Total(%)	71	84	64

From Tables I and II we can conclude that the highest accuracy was obtained with the complex-valued convolutional network using patch size of 24×24 pixels, then 86% of average accuracy was achieved. The real valued convolutional network performed the best the patch size of 24×24, providing accuracy of 84%. The patch size of 12×12 pixels provided accuracy of 84% and 71 % for complex valued and real valued case. The patch size of 48×48 pixels did not provide good recognition results. Experimental results showed that there is still a lot of improvements possible in the structure of deep convolutional networks. The best possible results could be achieved by combining real valued and complex valued parts of the convolutional networks.

V. Conclusion

In this paper a real valued and a complex valued convolutional network for categorization of SAR data were compared using different patch sized. Complex valued are real valued convolutional networks are not the same, but they should be very carefully designed. The experimental results showed that complex valued convolutional networks can archive better results in recognition in comparison to the real valued convolutional networks.

References

[1] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol. 313, no. 5786, pp. 504–507, 2006.

[2] W. Diao, X. Sun, X. Zheng, F. Dou, H. Wang, and K. Fu, “Efficient saliency-based object detection in remote sensing images using deep belief networks,” IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 2, pp. 137–141, 2016.

[3] F. P. S. Luus, B. P. Salmon, F. van den Bergh, and B. T. J. Maharaj, “Multiview deep learning for land-use classification,” IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 12, pp. 2448–2452, 2015.

[4] X. Chen, S. Xiang, C. L. Liu, and C. H. Pan, “Vehicle detection in satellite images by hybrid deep convolutional neural networks,” IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 10, pp. 1797–1801, 2014.

[5] S. Chen, H. Wang, F. Xu, and Y. Q. Jin, “Target classification using the deep convolutional networks for sar images,” IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 8, pp. 4806–4817, 2016.

[6] N. Guberman, “On complex valued convolutional neural networks,” CoRR, vol. abs/1602.09046, 2016. [Online]. Available: http://arxiv.org/abs/1602.09046

[7] H. G. Zimmermann, A. Minin, and V. Kusherbaeva, “Comparison of the complex valued and real valued neural networks trained with gradient descent and random search algorithms,” in European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 2011.

[8] R. Haensch and O. Hellwich, “Complex-valued convolutional neural networks for object detection in Polsar data,” in 8th European Conference on Synthetic Aperture Radar, June 2010, pp. 1–4.

[9] C. A. Popa, “Complex-valued convolutional neural networks for real-valued image classification,” in 2017 International Joint Conference on Neural Networks (IJCNN), May 2017, pp. 816–822.

[10] M. Wilmanski, C. Kreucher, and A. Hero, “Complex input convolutional neural networks for wide angle sar atr,” in 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Dec 2016, pp. 1037– 1041.

[11] 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.