Complex-Valued Convolutional Neural Networks for Object Detection in PolSAR data

Ronny Hänsch, Olaf Hellwich
Computer Vision and Remote Sensing, Technische Universität Berlin, Germany
FR3-1, Franklinstr. 28/29, 10587 Berlin
phone +49 30 314 73107
rhaensch@fpk.tu-berlin.de

Abstract

Detection methods for generic object categories, which are more sophisticated than pixel-wise classification, have been rarely introduced to polarimetric synthetic aperture radar (PolSAR) images. Despite the great success in other computer vision applications, the transfer to PolSAR data has been delayed due to the different statistical properties. This paper provides a first investigation of Complex-Valued Convolutional Neural Networks (CC-NN) for object recognition in PolSAR data. Although an architecture with only one single convolutional layer was used, the results are already superior to those obtained by a standard complex-valued neural network.

1 Introduction

Polarimetric Synthetic Aperture Radar (PolSAR) data measures the signal backscattered from the ground illuminated by an coherent polarized microwave. The characteristics of this echo depend on several properties like moisture, roughness and geometry and are therefore highly correlated with specific object categories. To automatically find those objects is one of the major research topics concerned with PolSAR related research. Most approaches are designed for real-valued data and require a projection of complex-valued PolSAR data to the real domain. Those projections can be made for example by using polarimetric decompositions theorems (like for example Freeman-Durden-decomposition [2]) or feature extraction methods especially designed for PolSAR images. Whatever the choice may be, it introduces a dependency of the classification result to this more or less arbitrarily projection. That is why in [3] the usage of Complex-Valued Multi-Layer Perceptrons (CV-MLPs) was proposed for pixel-wise classification of different land use in PolSAR data. Although real-valued MLPs are well known for several decades, their extension to the complex domain was delayed and their capabilities to solve classification tasks seldom investigated. The reasons are certain problems, which arise due to specific characteristics of complex-valued functions and their derivatives (see **Section 2** for more details).

A lot of research was concerned with pixel-wise classification of PolSAR imagery. Not only the afore mentioned work on CV-MLPs ([3]), but also a lot of other - meanwhile - standard approaches, like for example [8]. The spatial relationships between adjacent pixels or even an object oriented analysis, as done in computer vision research of optical near range data, was addressed seldomly and only in recent years by class specific algorithms. Convolutional Neural Networks (C-NNs) are one possible approach to exploit spatial relationships in structured image data in a very generic manner. The image is convolved by different trainable kernels in several layers and the output of those kernels serves as input to a standard MLP, which tries to relate it to object specific labels. Neither the unchanged image data only, nor pre-defined and possible sub-optimal feature detectors are used, but the feature extraction stage is part of the optimization process.

This paper proposes the usage of Complex-Valued Convolutional Neural Networks (CC-NNs) for classification in PolSAR data. The following two sections briefly review specific properties of CV-MLPs as well as C-NNs. In **Section 4** preliminary results of a CC-NN with one convolutional layer are presented and discussed. It outperforms the approach presented in [3] which uses standard CV-MLPs. In addition, the sensitivity of the learned kernels to different aspects of the PolSAR signal is shown.

2 Complex-Valued Multi-Layer Perceptrons

This section is going to concentrate on the most basic knowledge about CV-MLPs to enable the reader to grasp the differences to real-valued MLPs and understand the extensions made to the C-NNs explained in **Section 3**.

Attempts to develop learning rules for CV-MLPs can for

ISBN 978-3-8007-3272-2

example be found in [9, 4, 5], the influence of different error functions was studied in [10], and applications to classification tasks were investigated in [12, 3].

Functions of sigmoid shape are the common choice of activation functions for real-valued MLPs. They have the advantage, that they are bounded and differentiable at every point. As stated by Liouville's theorem constant functions are the one and only functional class, which fulfills both criteria at the same time within the complex domain. That is why CV-MLPs have to use either bounded but not analytical functions like $split-tanh$-function given by

$$f(z) = \tanh(\Re z) + i \tanh(\Im z) \tag{1}$$

where $\Re z$ and $\Im z$ are the real and imaginary parts of the complex number z, respectively, or analytical but not bounded functions. In the latter case special care has to be taken before and during training to avoid the net input approach the singular points [5].

A general learning rule was for example derived in [3]. The authors used it together with $split-tanh$-activation function and the complex quadratic error function (both explained by the same work) for classification of land use in PolSAR data:

$$w_{ji}^l(t+1) = w_{ji}^l(t) - \mu \sum_{\alpha=1}^{P} \delta_i^l \cdot y_j^{l-1} \tag{2}$$

$$\text{where } \mu = \frac{2 \cdot \tilde{\mu}}{P} \text{ and} \tag{3}$$

$$\delta_i^l = \begin{cases} \frac{\partial err}{\partial y_i^l} \cdot \frac{\partial y_i^l}{\partial h_i^l} + \frac{\partial err}{\partial y_i^{l*}} \cdot \frac{\partial y_i^{l*}}{\partial h_i^l} & , l = L \\ \sum_{r=1}^{N_{l+1}} \left(\delta_r^{l+1} w_{ir}^{l+1*} \frac{\partial y_i^l}{\partial h_i^l} + \delta_r^{l+1*} w_{ir}^{l+1} \frac{\partial y_i^{l*}}{\partial h_i^l} \right) & , l < L \end{cases} \tag{4}$$

3 Convolutional Neural Networks

Convolutional Neural Networks (C-NNs) are hierarchical multi-layered neural networks, which are inspired by biological research results on the visual cortex of mammals. They were introduced in [7] and successfully applied in several computer vision tasks, like optical character recognition [11] or face detection [6, 1].

Kernel based feature extraction is a widely used approach in a lot of computer vision applications. In most cases it is an ad hoc decision which kernel is used to calculate descriptive features. C-NNs use the output of different kernel-functions as input for the classification task, too. However, instead of applying pre-defined task specific feature operators before the classification process, they include the feature extraction into the optimization problem. During the training stage the error is not only backpropagated within the net itself, but even into the kernel. Therefore, no fixed a priori choice for a specific kernel function is made, but the net choose those kernels, which are most suitable for the specific task. Furthermore, little or even no pre-processing is required.

The architecture of a C-NN consists of several convolutional and subsequent sub-sampling layers. The sub-sampling reduces the computational effort during training and application. A common MLP follows the last convolutional/sub-sampling layer. A detailed description of a possible architecture of C-NN can be found for example in [1].

The one and only extension this work makes to standard C-NN is to use complex-valued weights and inputs in all layers. Therefore, complex domain specific learning as discussed for example in [3] has to be used.

4 Results and Discussion

4.1 Data

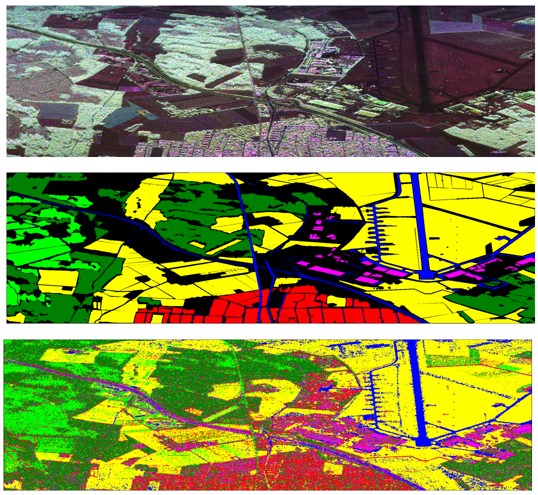

Figure 1: Top: PolSAR data of Oberpfaffenhofen (E-SAR, DLR); Middle: Ground truth data (no label (black), forest (dark green), low vegetation (light green), field (yellow), settlement (red), large building/industrial area (pink), street/railway (blue)); Bottom: Classification result of CC-NN

A color representation of an PolSAR image obtained by the E-SAR sensor (DLR) over Oberpfaffenhofen is shown at the top of **Figure 1**. Each pixel of this image represents a complex-valued Hermitian sample covariance matrix $\mathbf{C}$

$$\mathbf{C} = \frac{1}{n} \sum_{i}^{n} \mathbf{s}_i \cdot \mathbf{s}_i^* \tag{5}$$

$$\tag{6}$$

where n is the number of samples used for estimation and $\mathbf{s}$ is the lexicographic scattering vector

$$\mathbf{s} = (S_{HH}, \sqrt{2} S_{HV}, S_{VV})^T \tag{7}$$

ISBN 978-3-8007-3272-2

S_{TR} is an element of the scattering matrix **S** measured by the PolSAR sensor. The subscripts denote horizontal H and vertical V polarization during transmission T and receiving R.

Since CC-NNs belong to the group of supervised learning schemes, ground truth is needed in addition to the data itself during the learning phase. The PolSAR image was therefore manually and independently labeled by different humans to derive several label maps. The final label map, used during training and shown at the middle of **Figure 1**, was derived by using only pixels which were assigned with consistent labels in all label maps. All other pixels were excluded during learning and marked as black. The other colors code different kinds of land use, namely forest (dark green), low vegetation (light green), field (yellow), settlement (red), large building/industrial area (pink), and street/railway (blue).

4.2 Classification Performance

The aim of this paper is to investigate the general applicability of this learning scheme to classification of PolSAR data. Therefore, the number of convolutional layers was restricted to one in this study. The computational burden is acceptable for this simple net topology. Subsequent subsampling layers would only impede the interpretability without being much advantageous for the final classification task.

Nevertheless, the generalization performance of an convolutional neural network depends mainly on the number of convolutional kernels and the topology of the subsequent MLP. Therefore different numbers of neurons within the convolutional layer, as well as different numbers of hidden layers and neurons per hidden layer were investigated.

It is important to note, that beside the implicit speckle reduction by calculating sample covariance matrices through spatial averaging, no further speckle reduction or other noise suppression techniques were applied. The small window size of 2×3 for calculating sample covariance matrices result in image data with still strongly developed speckle.

Table 1: Convolutional CV-MLP test error for different net architectures

	5	10	50
6	$54\% \pm 24\%$	$35\% \pm 13\%$	$25\% \pm 2\%$
10-6	$41\% \pm 12\%$	$30\% \pm 4\%$	$24\% \pm 2\%$
50-6	$42\% \pm 10\%$	$28\% \pm 3\%$	$24\% \pm 2\%$
10-10-6	$38\% \pm 7\%$	$31\% \pm 5\%$	$26\% \pm 1\%$
50-50-6	$31\% \pm 4\%$	$26\% \pm 1\%$	$24\% \pm 1\%$

The classification results are summarized in **Table 1**. Each row denotes different topologies ranging from a single layer net (with six output neurons due to the six-class problem) to a net with two hidden layers each with fifty neurons. The columns mean different numbers of neurons in the single convolutional layer. The net was trained with four-thousand randomly drawn data points per class (about 2% of available pixels), where the rest was used to estimate the test error. The $split - tanh$-function (Eq.1) and the complex quadratic error function were used as activation and error function, respectively. Each cell contains mean and standard deviation of the test error estimated by five runs. Although the image data is of course stationary, the net was trained by online learning (weight update after each sample). This learning scheme speeds up the training process significantly compared to batch learning. Furthermore it introduces some kind of statistical fluctuations, which can help to overcome local minima. The learning converged (beside the minor fluctuations due to online learning) usually within twenty epochs. The receptive field of neurons within the convolutional layers was set to a size of 3×3.

As expected the test error usually decreases with more hidden neurons. A more complex topology lead to a better performance as well. However, the largest decrease of error can be achieved by using more features, meaning more neurons within the convolutional layer. The bottom of **Figure 1** shows a visual representation of the classification result.

Table 2: CV-MLP test error for different net architectures

	CV-MLP
6	$59\% \pm 5\%$
10-6	$36\% \pm 5\%$
50-6	$34\% \pm 3\%$
10-10-6	$33\% \pm 1\%$
50-50-6	$34\% \pm 3\%$

A CV-MLP was evaluated on the same data set with identical performance measurements. The results under usage of the same activation and error function are summarized in **Table 2**, again for different net topologies. As can be seen the introduction of contextual knowledge as well as the usage of learned convolutional features is able to significantly improve the performance if compared to a standard CV-MLP. Furthermore, the results of the proposed convolutional neural network are comparable to result obtained by CV-MLPs with previously performed speckle reduction, ranging from $19.5\% \pm 0.6\%$ to $23.7\% \pm 0.5\%$ for different filters.

4.3 Learned Convolutional Kernels

Its difficult to interpret the learnt kernel-functions visually, since they are multi-dimensional complex-valued functions defined over a spatial neighborhood of complex-

ISBN 978-3-8007-3272-2

valued sample covariance matrices. Therefore, **Figure 2** visualizes the activation of several neurons of the convolutional layer. As can be seen the activation differs greatly for different neurons. While some of them seem to specialize on local line features of different orientations, others try to capture polarimetric information.

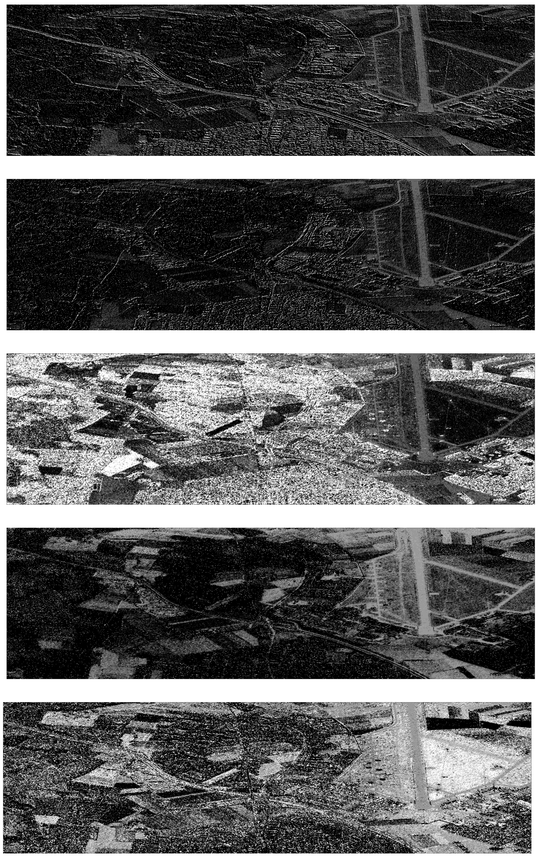

Figure 2: Activation of several neurons of the convolutional layer

5 Conclusion

The preliminary results presented by this paper show that CC-NN can be successfully utilized to classify PolSAR data. Although the net topology with only one convolutional layer is quite simple, the classification performance increased significantly if compared to CV-MLPs. Indeed, the achieved error rates are comparable to those acquired by applying a speckle reduction technique before classifying with a CV-MLP. Further improvements are expected, when the speckle is reduced before the classification by CC-NN as well as by usage of more convolutional layers. Furthermore, future work will include a more thoroughly study of the learnt kernel functions. In particular, the sensitivity for different polarimetric features like entropy, anisotropy and alpha angle should be investigated.

References

[1] Garcia, C.; Delakis, M.: *Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 11, 2004.

[2] Freeman, A.; Durden, S.L.: *A three-component scattering model for polarimetric SAR data*, IEEE Transactions on Geoscience and Remote Sensing, Vol. 36, No. 3, pp. 963-973, 1998.

[3] Hänsch, R.; Hellwich, O.: *Classification of Polarimetric SAR Data by Complex Valued Neural Networks*, ISPRS Workshop High-Resolution Earth Imaging for Geospatial Information, Hannover, Germany Volume XXXVIII, Part 1-4-7/W5, 2009.

[4] Hirose, A.: *Continuous complex-valued backpropagation learning*, Electronics Letters, Vol. 28, No. 20, pp. 1854-1855, 1992.

[5] Kim, T.; Adali, T.: *Approximation by fully complex multilayer perceptrons*, Neural computation Vol. 15, No. 7, pp. 1641-1666, 2003.

[6] Lawrence, S.; Gilles, C.; Tsai, A.; Back, A.: *Face Recognition: A Convolutional Neural Network Approach*, IEEE Trans. Neural Networks, Vol. 8, No. 1, pp. 98-113, 1997.

[7] LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P.: *Gradient-Based Learning Applied to Document Recognition*, Proc. IEEE, Vol. 86, No. 11, pp. 2278-2324, 1998.

[8] Lee, J.-S.; Grunes, M.R.; Kwok, R.: *Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution*, International Journal of Remote Sensing, Vol. 15, No. 11, pp. 2299-2311, 1994.

[9] Leung, H.; Haykin, S.: *The complex backpropagation algorithm*, IEEE Transactions on Signal Processing, Vol. 39, No. 9, pp. 2101-2104, 1991.

[10] Prashanth, A.; Kalra, P.K.; Vyas, N.S.: *Surface classification using ANN and complex-valued neural network*, Proceedings of the 9th International Conference on Neural Information Processing, ICONIP '02 Vol. 3, pp. 1094-1098, 2002.

[11] Wang, J.; Jean, J. *Multi-Resolution Neural Networks for Omnifont Character Recognition*, Proc. Intâl Conf. Neural Networks, Vol. 3, pp. 1588-1593, 1993.

[12] Yang, C-C.; Bose, N.K.: *Landmine detection and classification with complex-valued hybrid neural network using scattering parameters dataset*, IEEE transactions on neural networks, Vol. 16, No. 3, pp. 743-753, 2005.

ISBN 978-3-8007-3272-2