FULLY CONVOLUTIONAL NETWORKS FOR MULTI-TEMPORAL SAR IMAGE CLASSIFICATION

Adugna G. Mullissa, Claudio Persello, Valentyn Tolpekin

Faculty of Geoinformation Science and Earth Observation of the University of Twente
Enschede, the Netherlands
a.mullissa@utwente.nl

ABSTRACT

Classification of crop types from multi-temporal SAR data is a complex task because of the need to extract spatial and temporal features from images affected by speckle. Previous methods applied speckle filtering and then classification in two separate processing steps. This paper introduces fully convolutional networks (FCN) for pixel-wise classification of crops from multi-temporal SAR data. It applies speckle filtering and classification in a single framework. Furthermore, it also uses dilated kernels to increase the capability to learn long distance spatial dependencies. The proposed FCN was compared with patch-based convolutional neural network (CNN) and support vector machine (SVM) classifiers. The proposed method performed better when compared with the patch-based CNN and SVM.

***Index Terms*—** Fully convolutional networks, deep learning, SAR, Sentinel-1, Remote Sensing

1. INTRODUCTION

Agricultural crop mapping is an essential component to crop yield estimation. This has serious implications in economic policy making. In this regard, remote sensing image classification has been the technique of choice because it is cost effective and Sentinel data are freely accessible [1]. SAR data are particularly relevant for multi-temporal analysis of agricultural fields because of their all-weather imaging capability. Crop types can be distinguished from each other based on their backscattering characteristics, spatial features (texture) and temporal evolution of the radar reflectivity. Recently introduced deep learning methods such as convolutional neural networks (CNN) have been advantageous in these tasks because they can learn spatial features directly from the input image more effectively than standard techniques [2]. CNN have been effective in pixel-wise classification of remote sensing images.

In the literature, several studies have successfully applied standard CNN's [3] to SAR data in the context of semantic segmentation [4] [5]. The processes consisted of applying standard CNN to the data without speckle filtering [6]. Standard CNNs are patch based i.e. only the central pixels in the patch is labeled. This, however, results in redundant processing at inference time which leads to high computational time. To improve this, a fully convolutional networks (FCN) [7] is used. FCN's can be adapted to pixel-wise labelling of an input remote sensing image. Furthermore, both despeckling and semantic labeling of the data can be performed in one single processing framework.

In this paper, we design and apply FCN using dilated convolutions for the application of crop mapping from multi-temporal Sentinel-1 SAR images. We used an FCN with three convolutional layers using dilated kernels interleaved by non-linear activation functions. The dilated convolution was designed to improve the spatial contextual feature learning. In the designed network, SAR speckle filtering, spatial feature learning and semantic segmentation are performed within a single framework. The objective of the paper is as follows: 1) to evaluate the performance of FCN to classify a multi-temporal SAR series, 2) to evaluate the spatial contextual feature learning capability and 3) to investigate the suitability of the designed network for the application to crop type mapping.

2. METHODOLOGY

In this paper, we use a FCN that consists of filter banks interleaved by non-linear activations whose parameters are learned by minimizing a loss function. The main elements of the network are convolutional layers with dilated kernels. The weights of the layers are represented by a four dimensional arrays which dimensions are given by the kernel size, the number of channels being processed from the input image and the number of convolutional filters. For this purpose, we use a dilated filter bank to capture large range pixel dependency in the image. To minimize the number of parameters while keeping a large receptive field, instead of using convolutions with downsampling, we adopt dilated kernels as suggested by [7]. The dilated kernels capture large spatial

Layer	module type	dim	dilation	stride	pad
DK1	conv	$5 \times 5 \times 10 \times 16$	1	1	2
	lReLU				
DK2	conv	$5 \times 5 \times 16 \times 32$	2	1	4
	lReLU				
DK3	conv	$5 \times 5 \times 32 \times 32$	3	1	6
	lReLU				
class.	conv	$1 \times 1 \times 32 \times 2$	1	1	0
	softmax				

Table 1. Proposed FCN with three dilated convolutions (FCN-DK3).

texture patterns in the image by inserting zeros between filter cells.

The proposed network consists of three convolutional layers that use dilation, one classification layer and a softmax loss function. The convolutional layers use a 5×5 kernel with increasing dilation factor of one, two and three (Table 1). The convolutions are separated from each other by non-linear activations. We used a Leaky Rectified Linear Units (lReLU) with a leak factor of 0.1 [7]. The stride of the filter is set at 1 to avoid downsampling and a zero padding is applied to maintain the same dimension in the feature maps created from the input image. The parameters of the designed FCN are shown in Table 1. The networks used in this paper are implemented using the MatConvNet library version 1.0-beta-23 compiled with CUDA toolkit.

3. EXPERIMENTAL SETUP

3.1. Dataset

The proposed FCN is tested using 10 Sentinel-1 dual polarized ground range detected (GRD) products acquired over Flevoland, the Netherlands. The images have a nominal ground resolution of 10 meters. The Sentinel-1 sensor acquires data in C band for the dual polarimetric images (Table I). The test areas contain 1243×1393 pixels and cover 12.4 km $\times$ 13.9 km in range and azimuth directions, respectively. The Sentnel-1 image scene covers an entirely agricultural area (Figure 1). The acquisition dates and acquisition parameters are shown in Table 2. The ground reference data used in this paper was supplied in the form of ESRI shapefile by NEO B.V.

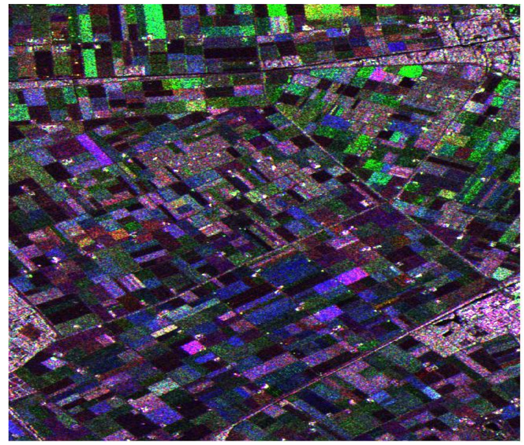

Fig. 1. False color composite image obtained by taking the images at different dates.

Parameter	Description
Polarization	HH, HV
Sensor	Sentinel-1A
Resolution	$10\text{m} \times 10\text{m}$
Incidence angle	39.35^0
Orbit	Ascending
Temporal baseline	12 days
Dates	May 30, 2016 - August 10, 2017
Number of images	10

Table 2. Acquisition parameters for the Sentinel-1 images.

3.2. Network training

The networks were trained using stochastic gradient descent (SGD) method. Batch normalization was used for every convolution layer. We finally trained DK3 by initializing the weights of the first two layers randomly. A multi-stage training was applied to minimize the training and validation errors. The networks were trained for 150 epochs with a learning rate of 10^{-4} and an additional 20 epochs with a rate of 10^{-5}. To apply this we used a training set of 1000 randomly selected labeled patches. A mini-batches of 32 samples and a weight decay factor of 5×10^{-4} was used. The trained network was finally applied to the test tile.

The training of the network in this paper was conducted by taking only the intensity image of the individual SAR image. The classification result from FCN-DK3 was compared with the state of the art classifiers: Support vector machines

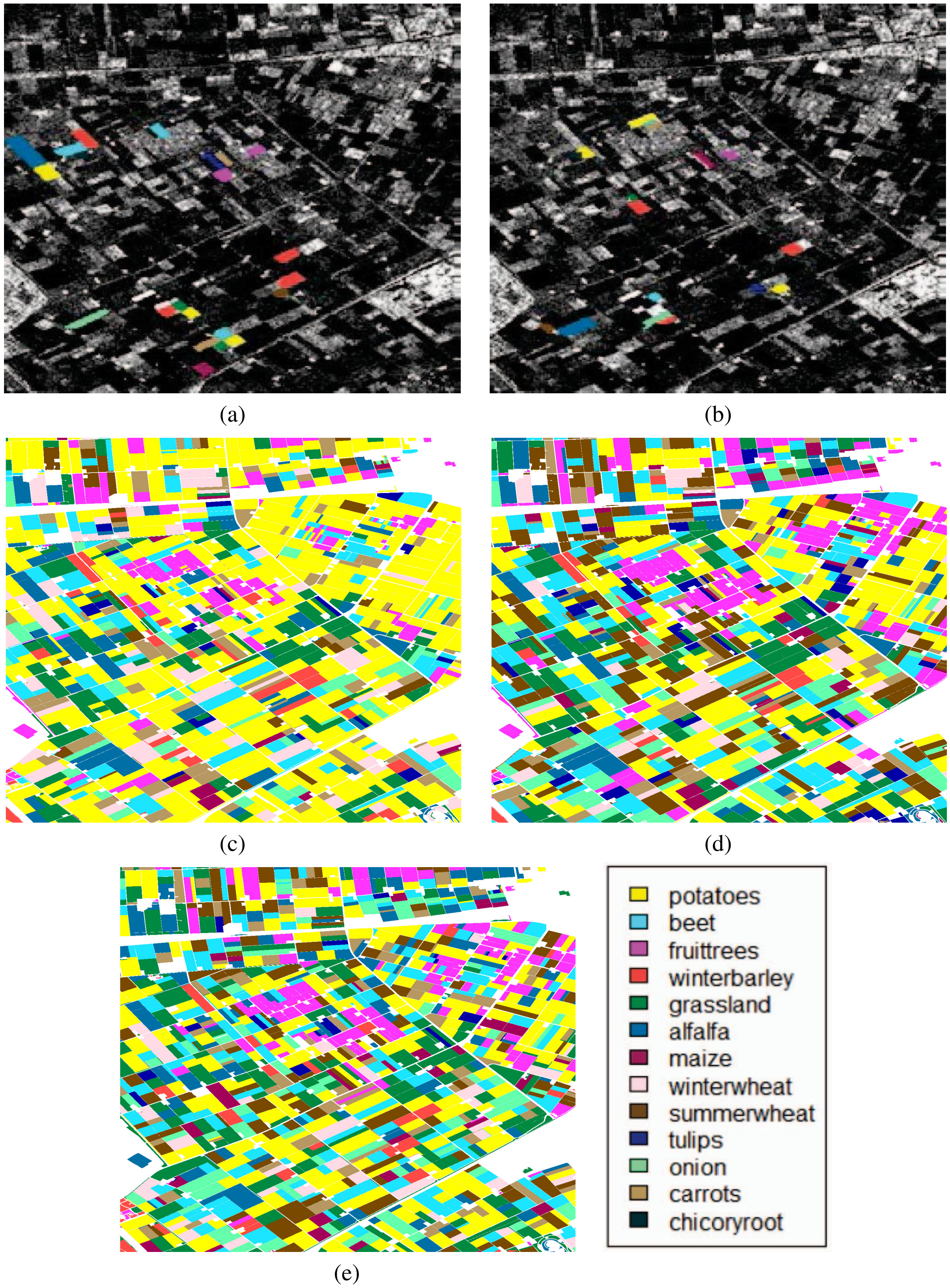

Fig. 2. (a) Ground truth polygons used in training. (b) Ground truth polygons used in validation. Classification results for (c) SVM (d) CNN (e) The proposed FCN-DK3.

(SVM) and patch based CNN. For the application of SVM, we applied a refined Lee filter before classification, and Grey level co-occurence matrix (GLCM) to extract features. The filtered images and GLCM features were used for classification. The patch based CNN is applied without applying any speckle filtering.

Parameter	Value
Kernel size	5
Patch size	15
Number of filters	32
Epoch	1000

Table 3. CNN network training parameters.

3.3. Classification results

To evaluate the results from the proposed FCN we split the reference data into a training set and a validation set. The three layer FCN is trained using the the polygons selected from the ground truth data and the accuracy of the classification scheme is evaluated from a separated set of polygons that were not used in the training procedure. To evaluate the performance of the proposed FCN we qualitatively and quantitaively compared the results from a 5 layer patch based CNN [3] (Table 3) and support vector machine (SVM) applied on a the same set of SAR images. The proposed FCN provided a less noisy classification result when compared with the CNN and SVM classification so we applied majority voting based on polygons obtained from the dutch spatial data infrastructure (SDI). It can be observed from Figure 2 that the different crop classes are better represented in the proposed FCN than the patch-based CNN or SVM. As shown in Table 4. the proposed FCN-DK3 achieves superior classification accuracy when compared with the patch based CNN and SVM classifiers.

Classifier	Overall accuracy	Mean producers accuracy
SVM	49.03%	25.6%
CNN	51.37%	30.77%
FCN-DK3	54.16%	54.37%

Table 4. Classification accuracies.

4. CONCLUSIONS

We introduced a FCN for crop classification of SAR images. The proposed method improves the classification accuracy when compared to patch based CNN. The application of dilated kernels improved the learning capability of the network by improving the spatial support of the network.

5. ACKNOWLEDGEMENTS

We thank NEO BV for supplying the ground truth data.

6. REFERENCES

[1] C. Pelletier, S. Valero, J. Inglada, N. Champion, and G. Dedieu, “Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas,” *Remote Sensing of Environment*, vol. 187, pp. 156–168, 2016.

[2] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[3] J. R. Bergado, C. Persello, and C. Gevaert, “A deep learning approach to the classification of sub-decimetre resolution aerial images,” in *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*. IEEE, 2016, pp. 1516–1519.

[4] Z. Huang, Z. Pan, and B. Lei, “Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data,” *Remote Sensing*, vol. 9, no. 9, pp. 907, 2017.

[5] J. Zhao, W. Guo, S. Cui, Ze. Zhang, and W. Yu, “Convolutional neural network for SAR image classification at patch level,” in *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*. IEEE, 2016, pp. 945–948.

[6] Carolyn Danilla, Claudio Persello, Valentyn Tolpekin, and John Ray Bergado, “Classification of multitemporal SAR images using convolutional neural networks and markov random fields,” in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2017, pp. 2231–2234.

[7] C. Persello and A. Stein, “Deep fully convolutional networks for the detection of informal settlements in VHR images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2325–2329, 2017.