A MANY-TO-MANY FULLY CONVOLUTIONAL RECURRENT NETWORK FOR MULTITEMPORAL CROP RECOGNITION

J. A. Chamorro[1], J. D. Bermudez[1], P. N. Happ[1], R. Q. Feitosa[1,2]

[1] Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil - (jchamorro, bermudez, patrick, raul)@ele.puc-rio.br
[2] State University of Rio de Janeiro, Rio de Janeiro, Brazil

ICWG II/III: Pattern Analysis in Remote Sensing

KEY WORDS: Crop Recognition, Recurrent Networks, Fully Convolutional Networks

ABSTRACT:

Recently, recurrent neural networks have been proposed for crop mapping from multitemporal remote sensing data. Most of these proposals have been designed and tested in temperate regions, where a single harvest per season is the rule. In tropical regions, the favorable climate and local agricultural practices, such as crop rotation, result in more complex spatio-temporal dynamics, where the single harvest per season assumption does not hold. In this context, a demand arises for methods capable of recognizing agricultural crops at multiple dates along the multitemporal sequence. In the present work, we propose to adapt two recurrent neural networks, originally conceived for single harvest per season, for multidate crop recognition. In addition, we propose a novel multidate approach based on bidirectional fully convolutional recurrent neural networks. These three architectures were evaluated on public Sentinel-1 data sets from two tropical regions in Brazil. In our experiments, all methods achieved state-of-the-art accuracies with a clear superiority of the proposed architecture. It outperformed its counterparts in up to 3.8% and 7.4%, in terms of per-month overall accuracy, and it was the best performing method in terms of F1-score for most crops and dates on both regions.

1. INTRODUCTION

The projections of world population for the next decades demand more efficient, comprehensive and precise agriculture. According to the United Nations reports, the world population is expected to reach 8.6 billion by 2030, 9.8 billion by 2050 and 11.2 billion by 2100 (United Nations, 2017). It is therefore necessary to promote policies to increase global agricultural production to ensure food supply with minimal environmental impact. In this context, crop monitoring is very important to develop commercial plans, regulate internal stocks and perform customized management decisions (Leite et al., 2011). Multitemporal remote sensing (RS) imagery has increasingly been applied for this task as a cost-effective way for gathering timely, detailed and reliable information over large areas (Thenkabail, 2015). However, crop recognition from RS data is particularly challenging in tropical regions, because the favorable climate associated with the use of modern technologies makes agriculture highly dynamic (Sanches et al., 2018b).

In recent years, deep learning models have made breakthroughs in several fields such as speech recognition and computer vision (LeCun et al., 2015). In remote sensing, these models have also been successfully tested in diverse applications (Audebert et al., 2017). Such models can be roughly grouped in two main categories: Convolutional Neural Networks (CNN) for understanding spatial context, and Recurrent Neural Networks (RNN), mostly to model data sequences.

In (La Rosa et al., 2018), a type of CNN called Fully Convolutional Network (FCN) was used for crop recognition having as input the stack of a multi-temporal sequence. Although a good performance is reported, the method requires the training of a particular model for each date. Thus, this solution can become computationally expensive depending on the dataset size.

RNNs can be configured to allow sequential inputs and to produce a single outcome that represents the semantic of the whole input sequence. Such "many-to-one" configurations have been used for crop-recognition in temperate regions, where a single crop occurs in each field over the whole season.

In (Ndikumana et al., 2018) two different RNN models, Long short-term memory (LSTM) and Gated Recurrent Unit (GRU), were applied for crop classification upon multi-temporal Sentinel-1 data. In (Bermudez et al., 2018), a CNN was proposed to provide the input to a RNN for the many-to-one crop recognition task.

In (Xingjian et al., 2015), the internal fully connected LSTM layers were replaced by convolutional layers. This type of recurrent convolutional network (ConvLSTM) is able to jointly model the spatial and temporal context from multi-temporal sequences of images. This kind of RNN was used for precipitation forecasting. Later, in (Rußwurm , Körner, 2018), this ConvLSTM network was applied to the multi-temporal land cover classification problem in a many-to-one configuration. Furthermore, (Rußwurm , Körner, 2018) used a bidirectional variant of ConvLSTM to eliminate bias toward the later sequence elements. All aforementioned proposals follow the many-to-one approach.

In areas with complex crop dynamics, such as in tropical regions, multiple crops may come about in a field during the season. Thus, the single crop per season assumption does not hold in those regions. In this case we require a network capable of performing crop recognition for multiple dates.

Our work hypothesis is that many-to-many RNN configurations can be applied for multidate crop recognition, to accurately

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-2-W7-25-2019 | © Authors 2019. CC BY 4.0 License.

identify crop classes in tropical regions at each date represented in a multitemporal sequence. Specifically, we introduce a novel many-to-many configuration of a bidirectional ConvLSTM for multidate crop recognition from multitemporal RS data. The proposed architecture uses a FCN encoder to provide inputs at a lower spatial resolution to a bidirectional LSTM. After processing the input provided by the encoder, the LSTM delivers the output, which is then applied to a decoder that generates the outcome, a pixel-wise label image, at the original spatial resolution.

In addition, we adapted two convolutional many-to-one RNNs, introduced in earlier works (Rußwurm , Körner, 2018), to the many-to-many task and compare them with the proposed architecture. The experiments were carried out upon datasets of two tropical regions charaterized by complex spatio-temporal dynamics and crop rotation practices.

To the best of our knowledge, this is the first work that addresses many-to-many recurrent networks as unique, end-to-end architectures, for pixel-wise crop recognition of entire image sequences. The contributions of this work are threefold:

1. a novel recurrent network architecture that combines bidirectional LSTM and FCN for multidate crop recognition,
2. an extension of convolutional LSTMs originally designed for single crop per season applications to multidate crop recognition,
3. an experimental analysis of the aforementioned network designs on datasets that represent highly dynamic agriculture typical of tropical regions.

The remainder of this paper is organized as follows: Section 2 briefly explains the concepts of RNNs, bidirectional RNNs and ConvLSTMs. In Section 3, the assessed methods for many-to-many multi-temporal crop recognition are presented, including the proposed one. Section 4 describes the study areas and the experimental protocol adopted in this work. Experimental results are presented in Section 5 and conclusions are outlined in Section 6.

2. FUNDAMENTALS

2.1 Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNN) are a type of neural network designed for processing sequential data. These models are regarded as the state-of-the-art for temporal modeling tasks (Ma et al., 2019). RNNs can be seen as neural networks with feedback. Given an input sequence ($\boldsymbol{x} = \boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_t$), the output of such network is given by the equations:

$$\boldsymbol{h}_t = f(\boldsymbol{b} + \boldsymbol{W}\boldsymbol{h}_{t-1} + \boldsymbol{U}\boldsymbol{x}_t) \quad (1)$$

$$\boldsymbol{y}_t = g(\boldsymbol{c} + \boldsymbol{V}\boldsymbol{h}_t) \quad (2)$$

Where $\boldsymbol{h}_t$ is the state at time step t, $\boldsymbol{W}$, $\boldsymbol{U}$ and $\boldsymbol{V}$ are weight matrices, $\boldsymbol{b}$ and $\boldsymbol{c}$ are bias vectors and $\boldsymbol{y}_t$ is the network output

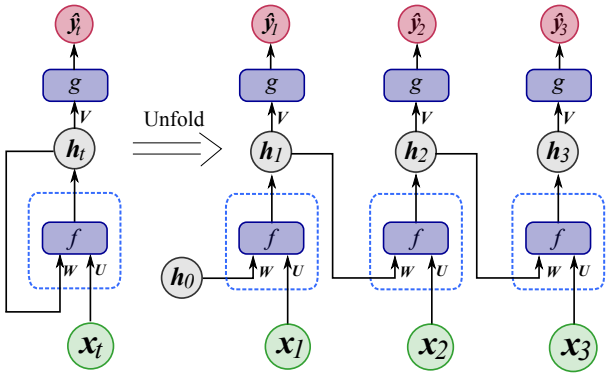

Figure 1. Many-to-many basic RNN.

for time step t. f and g are activation functions, usually *tanh* and *softmax*, respectively.

The training loss of a many-to-many recurrent network considers the entire output sequence. In this case, the total loss is computed by the sum of the losses over all time steps. This configuration is useful for multidate crop recognition because predictions for the entire image sequence can be obtained by a single model. Figure 1 shows on the left the basic RNN architecture and on the right its unrolled representation for three time steps.

To produce the outcome $\boldsymbol{x}_t$ at time t the basic RNN relies on the current input $\boldsymbol{x}_t$ and on a summary of prior time steps coded in the previous state $\boldsymbol{h}_{t-1}$. When available, inputs at posterior instants can be used to improve the prediction at time t. This is achieved by bidirectional RNNs. They consist of two RNNs trained simultaneously. The first RNN is trained in the temporal forward direction, whereas the second one is trained in the backward direction (Schuster , Paliwal, 1997). Correspondent state vectors from both RNNs, $\vec{\boldsymbol{h}}_t$ and $\overleftarrow{\boldsymbol{h}}_t$ are usually concatenated to form the unified state vector $\boldsymbol{h}_t$. This scheme is illustrated in Figure 2 for a sequence of length equal to 3.

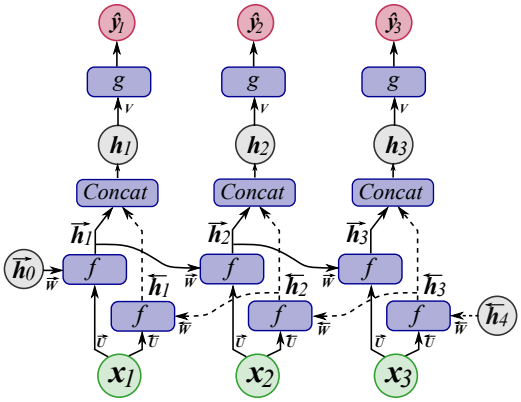

Figure 2. Bidirectional RNN for three time steps (Unfolded representation).

2.2 Convolutional Long Short Term Memory (ConvLSTM)

Traditional RNNs fail when it comes to modeling long-term dependencies and suffer from some stability issues. A special type of RNN, called *Long Short Term Memory (LSTM)*, was conceived to mitigate these problems (Hochreiter ,

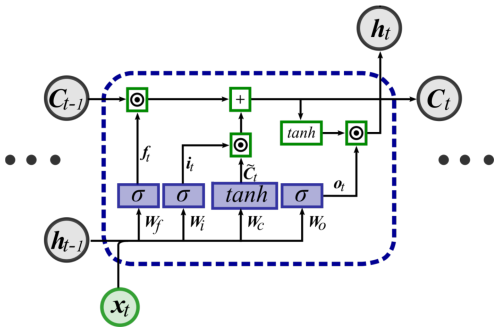

Figure 3. LSTM structure diagram (Bermudez et al., 2017).

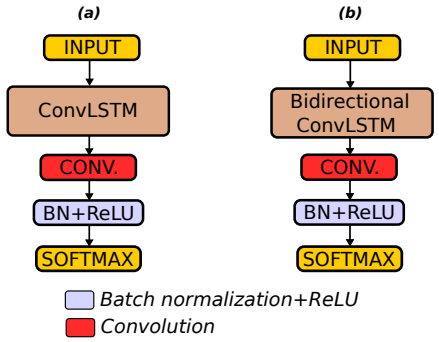

Figure 4. RNN architectures adapted to many-to-many tasks: *(a)* UConvLSTM, *(b)* BConvLSTM.

Schmidhuber, 1997). The main improvement in comparison to traditional RNNs is a memory cell $\boldsymbol{C}_t$ that can be accessed, written and cleared by trainable gates (see Figure 3). Specifically, a LSTM contains an information gate $\boldsymbol{i}_t$ to select which information should be added to the cell, a forget gate $\boldsymbol{f}_t$ to discard useless previous knowledge, and an output gate $\boldsymbol{o}_t$ to decide whether the cell contents should propagate to subsequent steps through the current state $\boldsymbol{h}_t$. In their original form, these gates are implemented as fully connected layers followed by an activation function such as *sigmoid*.

An extension of the LSTM design was proposed for image analysis in (Xingjian et al., 2015). In these networks, called convolutional LSTM (ConvLSTM), the fully connected layers at $\boldsymbol{i}_t$, $\boldsymbol{f}_t$ and $\boldsymbol{o}_t$ are replaced by convolutional layers in order to better capture spatial context. Thus, the input and the output of a ConvLSTM correspond to a sequence of images, as opposed to a sequence of vectors in the basic RNN.

3. RNN ARCHITECTURES FOR MULTIDATE RECOGNITION

In this section we present the recurrent network architectures used for crop mapping from multitemporal RS data. Firstly, we describe the two networks adapted from (Rußwurm , Körner, 2018) for many-to-many tasks that served as baseline in our research. Next, the proposed architecture is presented.

3.1 Unidirectional Convolutional LSTM

The first architecture we consider in this paper is the Unidirectional Convolutional LSTM (UConvLSTM), a unidirectional version of the architecture proposed in (Rußwurm , Körner, 2018), which we adapted to many-to-many tasks. Its architecture is shown in Figure 4a. The input sequence goes first through a ConvLSTM net followed by 1×1 convolutions that produce as many activation maps as the number of classes. Next, batch normalization and ReLU activation functions are applied. In the final layer, a $softmax$ function assigns posterior probabilities to each pixel.

3.2 Bidirectional Convolutional LSTM

The second architecture we tested in this work is the Bidirectional Convolutional LSTM (BConvLSTM), illustrated in Figure 4b. The BConvLSTM also derives from the architecture proposed in (Rußwurm , Körner, 2018) that we adapted for many-to-many tasks. It can be regarded as a bidirectional version of UConvLSTM, whereby the plain ConvLSTM layer is replaced by a bidirectional ConvLSTM layer. The BConvLSTM network comprises two ConvLSTMs: one processes the input data in the forward direction, while the other operates in reversed, backward direction. The outputs of both ConvLSTM are concatenated to form a single output tensor. From this point on, the architecture does not differ from the previous one. 1×1 convolutions are applied to aforementioned tensor producing one activation map per class, followed by batch normalization and by a ReLU activation function. A $softmax$ layer delivers posterior probabilities for each pixel.

3.3 Bidirectional Dense Convolutional LSTM - BDenseConvLSTM

The architecture from (Rußwurm , Körner, 2018), which is our main reference for comparison purposes, applies convolutions at the original image scale only. In contrast, modern FCN architectures tend to follow an encoder-decoder pattern to better capture the spatial information at multiple scales. Such structure comprises a downsampling path, so called encoder, which extracts coarse semantic features, followed by an upsampling path, so called decoder, responsible for recovering the input spatial resolution in the final output.

Our proposal combines elements of the architecture presented in (Rußwurm , Körner, 2018) with the encoder-decoder structure from a FCN, as shown in Figure 6. Different FCN architectures could be considered for this encoder-decoder design. In the present work, we use the dense FCN introduced in (Jégou et al., 2017). This architecture consists of three main block types: a) the Dense blocks (*DB*), consisting of sequences of convolutional layers with multiple bypassing connections, b) the Transition Down (*TD*) blocks, which comprise a convolution followed by a downsampling operation, and c) the Transition Up (*TU*) blocks that perform upsampling operations, typically a transposed convolution. Skip connections are used between downsampling and upsampling stages.

4. EXPERIMENTS

4.1 Study Areas

Two publicly available datasets for multitemporal crop recognition in tropical regions were used for performance assessment. The first region is located in Campo Verde municipality, Brazil, with an extension of 4,782 km^2 (Sanches et al., 2018b). It features a sequence of 14 pre-processed, dual polarized Synthetic Aperture Radar (SAR) images from Sentinel-1. These images were taken between October 2015

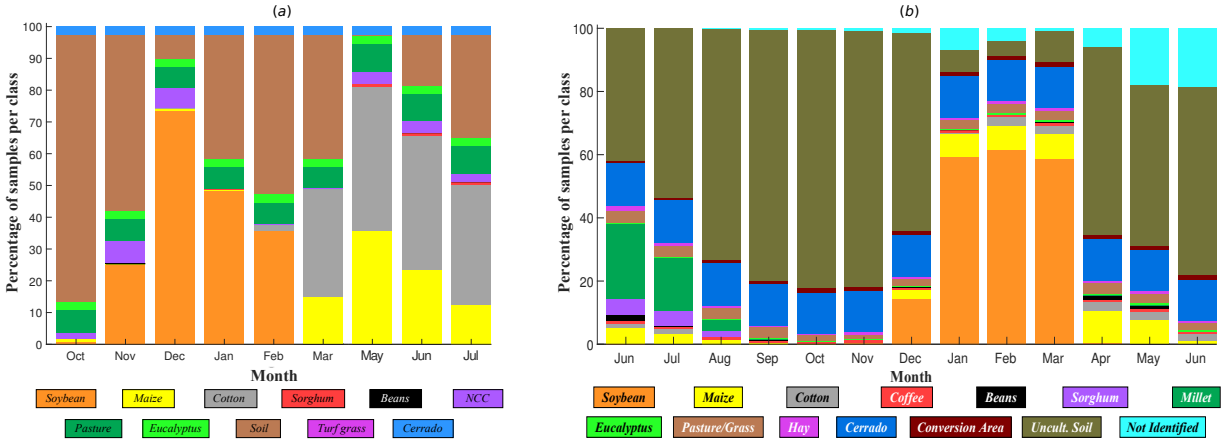

Figure 5. Class distribution in (a) Campo Verde and (b) LEM datasets.

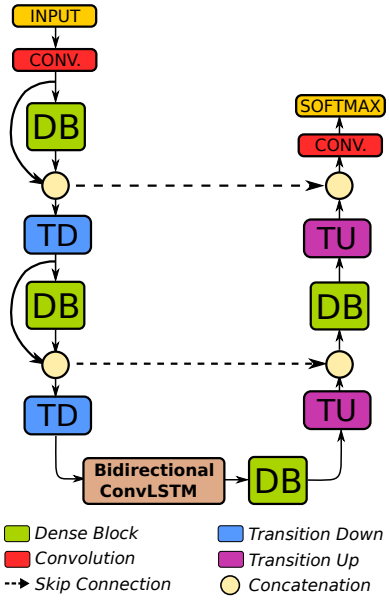

Figure 6. BDenseConvLSTM architecture.

and July 2016, with one or two images per month. The class distribution greatly varies over time (see Figure 5a). *Soybean* is the main crop type from October 2015 to February 2016 and its replaced by *Cotton* and *Maize* in the following months.[1]

The second region is located in Luis Eduardo Magalhães (LEM) municipality, also in Brazil, with an area of 3,940 km^2 (Sanches et al., 2018a). A set of 13 pre-processed Sentinel-1 SAR images acquired between June 2017 and June 2018 was used in our experiments. Similar to Campo Verde, the class distribution in LEM dataset is non uniform along the year, as shown in Figure 5b. The main crop types are *Soybean*, *Maize*, *Cotton* and *Millet*.[2]

4.2 Hyperparameter Setup

We experimented with different hyperparameter values for each method. In this section, we present the configurations that attained the best results.

[1]The Campo Verde database is available in IEEE Daport at https://ieee-dataport.org/documents/campo-verde-database.

[2]The LEM database is freely accessible at http://www.lvc.ele.puc-rio.br/downloads/Databases/LEM/home.html.

Parameter setups for UConvLSTM and BConvLSTM networks are shown in Tables 1 and 2, where T represents the temporal sequence length. 256 convolutional filters were used in the UConvLSTM network for each LSTM internal gate. Likewise, the BConvLSTM model was configured with 256 recurrent filters per gate: 128 for each direction.

Layer	Output Shape	Filters
Input	$T \times 32 \times 32$	2
ConvLSTM	$T \times 32 \times 32$	256
Conv.	$T \times 32 \times 32$	$\#classes$

Table 1. UConvLSTM parameter configuration - T is the sequence length

Layer	Output Shape	Filters
Input	$T \times 32 \times 32$	2
Bidirectional ConvLSTM	$T \times 32 \times 32$	256
Conv.	$T \times 32 \times 32$	$\#classes$

Table 2. BConvLSTM parameter configuration - T is the sequence length.

Following (La Rosa et al., 2018), the BDenseConvLSTM network was built with two convolutional layers per dense block and 20% as dropout factor. Further details from this architecture are presented in Table 3. *Average Pooling* was empirically selected as downsampling operator. Except for the last convolution, we adopted 3×3 filters in all cases.

Layer	Output Shape	Filters
Input	$T \times 32 \times 32$	2
DB	$T \times 32 \times 32$	80
Downsampling	$T \times 16 \times 16$	80
DB	$T \times 16 \times 16$	112
Downsampling	$T \times 8 \times 8$	112
Bidirectional ConvLSTM	$T \times 8 \times 8$	256
DB	$T \times 8 \times 8$	32
Upsampling	$T \times 16 \times 16$	144
DB	$T \times 16 \times 16$	32
Upsampling	$T \times 32 \times 32$	112
Conv.	$T \times 32 \times 32$	$\#classes$

Table 3. BDenseConvLSTM parameter configuration - T is the sequence length.

For all the networks, we applied early stopping with the Adagrad optimizer and a learning rate of 0.01. Mini-batches of size 16 were empirically selected.

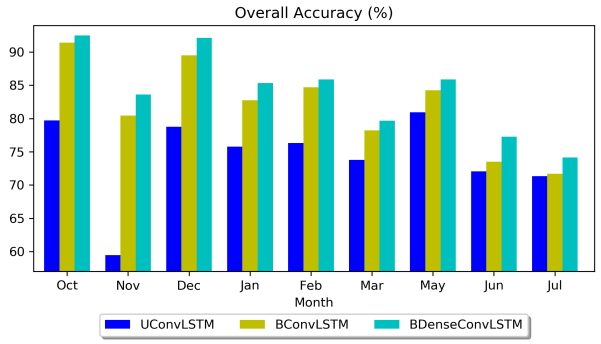

Figure 7. Overall Accuracy for Campo Verde study area, computed in each date.

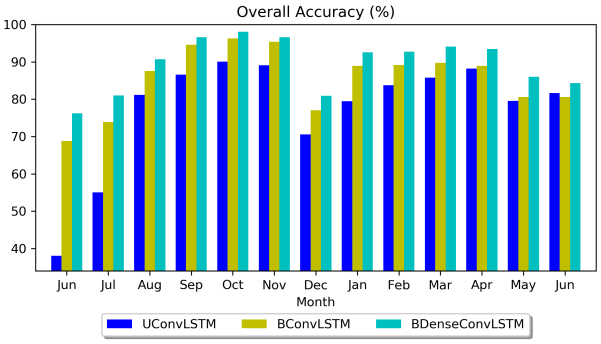

Figure 8. Overall Accuracy for LEM study area, computed in each date.

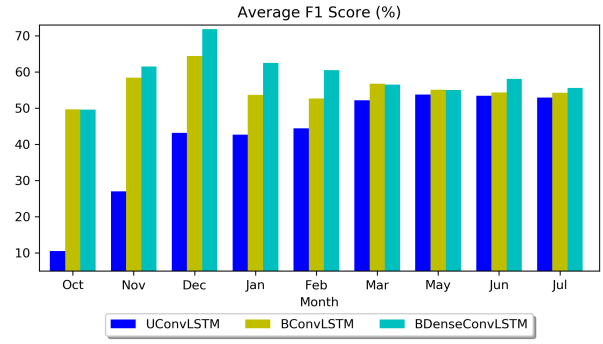

Figure 9. Average F1-Score for Campo Verde study area, computed in each date.

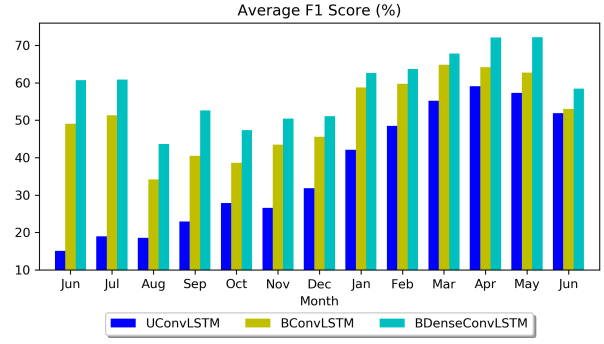

Figure 10. Average F1-Score for LEM study area, computed in each date.

4.3 Experimental Protocol

Parcels present in the dataset were randomly separated in training and testing sets, whereby the training set contained about 50% of all pixels for Campo Verde and 75% for LEM. Each parcel was split into non-overlapping square patches that were processed separately by the networks. After all patches have been processed, the patch-wise classification results were arranged in a mosaic forming the final outcome. The input patch size was set to $14 \times 32 \times 32$ pixels for Campo Verde and to $11 \times 32 \times 32$ for LEM.

In this work, data augmentation strategies such as rotation, horizontal and vertical flip were used, since they were empirically found to improve overall and per-class performance metrics. Experiments were carried out using Keras framework with Tensorflow backend, on a Nvidia GTX Titan GPU.

5. RESULTS

The experimental results are reported in Figures 7 to 10. The figures show one result per month for each architecture. Thus, when the dataset contains more than one image per month, the reported result refers to the latest image.

Figures 7 and 8 present the performance achieved by each architecture in terms of Overall Accuracy (OA) for Campo Verde and LEM, respectively. Each bar group contains the performance of all tested network designs for a month. Our proposed method consistently achieved the highest scores for both datasets, outperforming the second best approach, BConvLSTM, in up to 3.8% for Campo Verde and 7.4% for LEM. The UConvLSTM network presented the lowest OA values (with just one exception) in comparison two the other networks, specially at the earliest dates. This occurred because UConvLSTM uses only data from past dates for predictions, ignoring data of posterior dates.

Figure 9 summarizes the networks' performance in terms of average F1 score for Campo Verde. UConvLSTM presented a poorer performance at the earlier dates, and came closer to the other network designs at the later dates.

BDenseConvLSTM achieved the best F1 scores in most months, with exception of October, March and May, when it performed similarly to the BConvLSTM network. Table 4 sheds light over this results. It contains the F1 scores of most relevant crop types across the entire Campo Verde sequence. The best performance values for each crop and month are highlighted in bold. Clearly, the BDenseConvLSTM network was the best performing architecture in most cases. Exceptions occurred mostly in months when the target crop was weakly represented. This can be inferred by comparing the results of Table 4 with the crop distribution in Figure 5.

Figure 10 shows the average F1-score for the LEM dataset. As in the experiments on Campo Verde, UConvLSTM performed poorer than the other models during the earlier dates and improved for the later dates. BDenseConvLSTM was the best performing network over all LEM sequence also in terms of F1, being 7.5% higher than BConvLSTM in average.

Table 5 shows the class specific F1 scores for the most relevant crops in the LEM dataset. The best performance values for each crop and date are also highlighted in bold. The superiority of BDenseConvLSTM in terms of F1 scores was even more evident here than in the experiments on Campo Verde. The

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-2-W7-25-2019 | © Authors 2019. CC BY 4.0 License.

	Crop Type	Month (%)								
		Oct	Nov	Dec	Jan	Feb	Mar	May	Jun	Jul
UConvLSTM	Soybean	0.0	57.6	90.4	82.0	76.5	**39.2**	-	-	-
	Maize	0.0	0.0	10.9	15.2	26.2	54.4	81.6	64.2	**45.9**
	Cotton	-	-	47.0	52.3	27.7	77.1	89.1	87.8	85.2
	Sorghum	-	-	0.2	0.5	2.8	8.4	50.3	50.9	53.1
	Beans	-	15.4	39.1	-	-	-	36.2	-	-
	Eucalyptus	4.9	39.5	61.8	70.2	75.3	81.2	83.8	84.7	86.0
BConvLSTM	Soybean	27.0	74.5	96.6	85.8	84.5	37.3	-	-	-
	Maize	44.3	73.5	57.5	0.6	3.0	70.1	87.3	66.1	42.1
	Cotton	-	-	73.2	71.4	43.7	80.2	91.8	89.1	86.1
	Sorghum	-	-	14.7	13.4	12.3	11.8	50.5	49.8	50.4
	Beans	-	28.3	29.8	-	-	-	33.9	-	-
	Eucalyptus	95.3	94.4	93.4	93.1	**93.2**	89.1	85.8	85.6	86.3
BDenseConvLSTM	Soybean	**34.7**	**78.3**	**98.3**	**88.2**	**86.2**	38.1	-	-	-
	Maize	**67.1**	**91.3**	**84.8**	**68.2**	**69.6**	**72.8**	**89.7**	**72.3**	43.1
	Cotton	-	-	**76.1**	**72.6**	**46.0**	**81.9**	**92.2**	**90.0**	**87.0**
	Sorghum	-	-	**41.4**	**41.7**	**34.0**	**16.7**	**51.6**	**53.8**	**51.6**
	Beans	-	**46.8**	**59.6**	-	-	-	**40.9**	-	-
	Eucalyptus	**95.6**	**95.3**	**95.1**	**94.5**	92.7	**93.7**	**93.3**	**93.1**	**92.5**

Table 4. Average F1 score for the most relevant crop types in Campo Verde study area, computed at each date from October 2015 to July 2016.

	Crop Type	Month(%)												
		Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
UConvLSTM	Soybean	1.2	0.5	-	-	-	28.1	39.7	88.0	91.4	91.5	58.6	74.2	77.2
	Maize	0.0	1.3	14.9	13.3	-	15.1	43.1	63.5	64.3	64.9	73.7	62.0	35.8
	Cotton	17.1	42.0	0.3	-	-	-	-	29.6	69.4	80.8	95.9	98.0	97.8
	Coffee	8.4	11.8	15.2	17.4	23.9	29.6	35.4	39.6	42.4	45.4	48.6	50.6	51.5
	Beans	3.7	2.4	-	-	-	-	-	-	-	63.1	48.4	43.0	-
	Sorghum	12.0	14.9	33.7	24.4	-	-	-	-	-	-	-	-	-
	Millet	44.7	46.2	14.7	0.0	-	-	-	-	0.0	0.0	11.7	16.1	0.0
	Eucalyptus	2.6	7.7	13.5	17.5	19.9	22.9	26.1	27.8	28.3	28.6	28.7	28.7	28.2
BConvLSTM	Soybean	32.6	15.1	-	-	-	35.6	41.0	94.1	95.0	94.4	60.6	83.4	84.5
	Maize	75.2	77.3	**20.1**	**67.1**	-	**83.8**	76.7	80.3	75.0	72.2	75.3	61.6	25.9
	Cotton	76.6	75.7	0.9	-	-	-	-	77.8	99.3	98.9	97.7	98.7	97.5
	Coffee	75.9	75.6	73.8	71.2	70.2	68.4	66.0	64.0	62.8	64.3	65.5	68.3	70.1
	Beans	30.3	**78.0**	-	-	-	-	-	-	-	**85.7**	69.6	59.3	-
	Sorghum	19.4	19.2	29.4	42.0	-	-	-	-	-	-	-	-	-
	Millet	59.2	47.8	53.8	0.0	-	-	-	-	0.0	0.0	27.1	20.9	0.0
	Eucalyptus	24.2	25.8	27.8	29.0	31.7	33.7	34.5	33.8	32.5	31.6	31.0	30.7	29.7
BDenseConvLSTM	Soybean	**79.5**	**74.1**	-	-	-	**62.3**	**46.0**	**96.1**	**96.4**	**96.6**	**65.4**	**88.9**	**88.1**
	Maize	**84.3**	**81.5**	17.2	59.5	-	67.9	**85.5**	**90.6**	**87.0**	**86.8**	**86.9**	**74.6**	**41.6**
	Cotton	**87.6**	**88.9**	**7.3**	-	-	-	-	**80.5**	**99.7**	**99.6**	**99.4**	**99.8**	**99.8**
	Coffee	**87.4**	**87.8**	**87.1**	**87.2**	**87.3**	**85.9**	**87.4**	**85.8**	**88.7**	**89.5**	**89.3**	**89.8**	**89.6**
	Beans	**31.9**	47.7	-	-	-	-	-	-	-	79.8	**77.8**	**77.7**	-
	Sorghum	**43.1**	**47.4**	**66.0**	**72.2**	-	-	-	-	-	-	-	-	-
	Millet	**68.6**	**59.0**	**57.5**	0.0	-	-	-	-	**2.3**	**9.6**	**55.7**	**49.1**	**0.1**
	Eucalyptus	**60.4**	**62.1**	**61.2**	**63.0**	**66.3**	**65.6**	**66.1**	**64.7**	**63.2**	**63.8**	**65.5**	**66.1**	**62.1**

Table 5. F1 score for most relevant crop types in LEM study area, computed at each date. The sequence starts in June 2017 and ends in June 2018.

few exceptions when BDenseConvLSTM did not achieve the highest score refer mostly to weakly represented classes.

As stated before, classes with few samples are vulnerable to obtain very different F1 score among different networks. This can be seen in the Campo Verde for *Maize* in January, with an F1 score of 0.6% for BConvLSTM and 68.2% for BDenseConvLSTM.

Figure 11 shows snips of the reference and the classification maps obtained by the methods on a particular location of the test area in Campo Verde at three dates. Consistent with results reported in Figure 9, UConvLSTM performed worse than the other models in the first month (Oct). It predicted *Soil* for all the area. Its performance improved for the later dates coming closer to the other methods. This indicates the importance of exploiting data from posterior dates. Figure 11 also reveals that BDenseConvLSTM results tended to be less affected by the salt and pepper effect.

Figure 12 shows snips of reference and classification maps for the LEM dataset. Similar to what was shown in Figure 10, BDenseConvLSTM clearly outperformed the other methods. In addition, the salt and pepper effect is once again more evident in UConvLSTM and BConvLSTM. In contrast, BDenseConvLSTM obtains smoother predictions in almost all parcels. Recall that BDenseConvLSTM reduces the resolution of the input data, before feeding the recurrent network. Thus, in this network, the LSTM layer operates upon comparatively lower resolution patches. The result produced by the LSTM layer is then upsampled by a decoder that restores the original spatial resolution. We verified experimentally that this last step prevents small regions in the outcome. This explains partially the comparatively smoother outcome of BDenseConvLSTM.

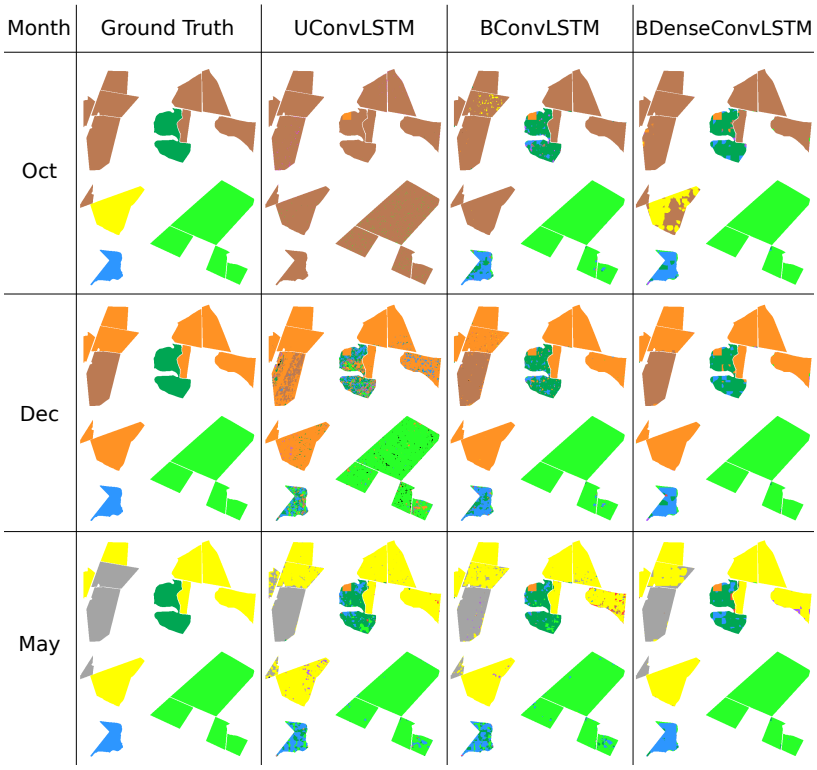

Figure 11. Sample structured output for Campo Verde study area.

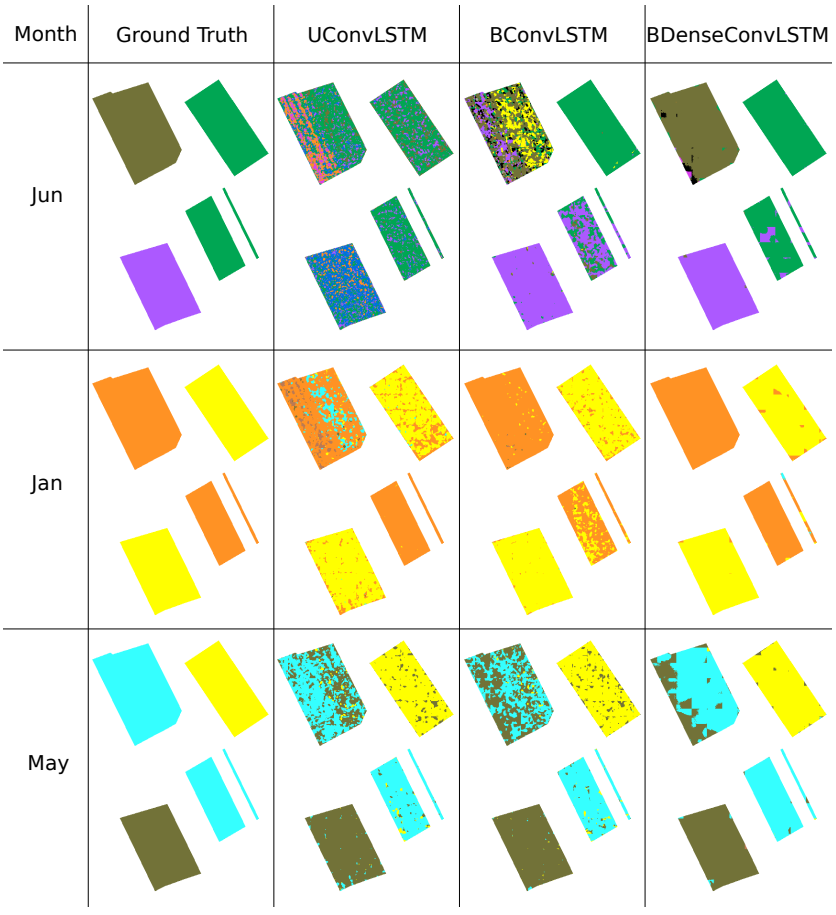

Figure 12. Sample structured output for LEM.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-2-W7-25-2019 | © Authors 2019. CC BY 4.0 License.

6. CONCLUSIONS

This paper introduced an extension of the traditional ConvLSTM networks in multitemporal crop recognition, by performing classification of an entire sequence of multitemporal images. In contrast, previous approaches produced classification for a simple element of the sequence. Furthermore, a novel fully convolutional bidirectional recurrent network, called BDenseConvLSTM, was proposed. The network was validated by comparing its performance against the conventional ConvLSTM network and its bidirectional variant BConvLSTM.

In all cases, the bidirectional networks outperformed the unidirectional approach for the first elements of the temporal sequence. This indicates that the bidirectional variation for recurrent networks is essential in many-to-many configurations.

The UConvLSTM and BConvLSTM networks produced a salt and pepper effect at their outputs. In contrast, BDenseConvLSTM, which includes an additional spatial encoding stage, reduced this effect and obtained smoother predictions. This indicates the importance of multi-scale spatial information in the design of convolutional recurrent architectures.

Finally, BDenseConvLSTM obtained the highest scores across two different datasets, with a more significant performance difference in LEM dataset. Thus, this network is recommended for many-to-many multi-temporal crop recognition applications.

Future works will focus in comparing these approaches with novel recurrent fully convolutional architectures using state-of-the-art semantic segmentation techniques such as *Atrous Pyramid Spatial Pooling*, and the design of more suitable data augmentation techniques for SAR images.

REFERENCES

Audebert, N., Boulch, A., Randrianarivo, H., Le Saux, B., Ferecatu, M., Lefèvre, S., Marlet, R., 2017. Deep learning for urban remote sensing. *2017 Joint Urban Remote Sensing Event (JURSE)*, IEEE, 1–4.

Bermudez, J.D., Feitosa, R.Q., Achanccaray, P., Happ, P.N., Sanches, I.D., Cué, L.E., 2017. Evaluation of recurrent neural networks for crop recognition from multitemporal remote sensing images. *Anais do XXVII Congresso Brasileiro de Cartografia*.

Bermudez, J.D., Feitosa, R.Q., Happ, P.N., 2018. An hybrid recurrent convolutional neural network for crop type recognition based on multitemporal sar image sequences. *IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 3824–3827.

Hochreiter, Sepp, Schmidhuber, Jürgen, 1997. Long short-term memory. *Neural computation*, 9, 1735–1780.

Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y., 2017. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, IEEE, 1175–1183.

La Rosa, L.E., Happ, P.N., Feitosa, R.Q., 2018. Dense fully convolutional networks for crop recognition from multitemporal sar image sequences. *IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 7460–7463.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature*, 521, 436.

Leite, P.B., Feitosa, R.Q., Formaggio, A.R., Da Costa, G.A., Pakzad, K., Sanches, I.D., 2011. Hidden Markov Models for crop recognition in remote sensing image sequences. *Pattern Recognition Letters*, 32, 19–26.

Ma, C.Y., Chen, M.H., Kira, Z., AlRegib, G., 2019. TS-LSTM and temporal-inception: Exploiting spatiotemporal dynamics for activity recognition. *Signal Processing: Image Communication*, 71, 76–87.

Ndikumana, E., Ho Tong Minh, D., Baghdadi, N., Courault, D., Hossard, L., 2018. Deep Recurrent Neural Network for Agricultural Classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, 10. http://www.mdpi.com/2072-4292/10/8/1217.

Rußwurm, M., Körner, M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, 7, 129.

Sanches, I.D., Feitosa, R.Q., Achanccaray, P., Montibeller, B., Luiz, A.J.B., Soares, M.D., Prudente, V.H.R., Vieira, D.C., Maurano, L.E.P., 2018a. Lem Benchmark Database for Tropical Agricultural Remote Sensing Application. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 621, 387–392.

Sanches, I.D., Feitosa, R.Q., Achanccaray, P., Soares, M.D., Luiz, A.J., Schultz, B., Maurano, L.E., 2018b. Campo Verde Database: Seeking to Improve Agricultural Remote Sensing of Tropical Areas. *IEEE Geoscience and Remote Sensing Letters*, 15, 369–373.

Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45, 2673–2681.

Thenkabail, P.S., 2015. *Land resources monitoring, modeling, and mapping with remote sensing*. CRC Press.

United Nations, 2017. *World Population Prospects: The 2017 Revision, Key Findings and Advance Tables.* Working Paper No. ESA/P/WP/248.

Xingjian, S., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C., 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 802–810.