A WEAKLY-SUPERVISED CHANGE DETECTION TECHNIQUE FOR SAR IMAGES BASED ON DEEP LEARNING AND SYNTHETIC TRAINING DATA GENERATED BY AN ENSEMBLE OF SELF-ORGANIZING MAPS

VICTOR-EMIL NEAGOE*, ADRIAN-DUMITRU CIOTEC*, LORENZO BRUZZONE**

** Department of Applied Electronics and Information Engineering, Polytechnic University of Bucharest, RO-060042, Bucharest, Romania,*
email: victoremil@gmail.com

*** Department of Information Engineering and Computer Science, University of Trento, I-38123, Trento, Italy,*
email: lorenzo.bruzzone@unitn.it

Abstract

This paper proposes a novel approach to improve accuracy of weakly-supervised change detection in Synthetic Aperture Radar (SAR) imagery. The method is based on the idea to use an initial small size labeled dataset for synthetic training data generation (STDG) using an Ensemble of Self-Organizing Maps (ESOM). The resulting synthetic data set substitutes the initial small set of labeled authentic data and it is used to train a simple deep neural network (DNN) classifier. The proposed DNN architecture consists of 6 layers: an input layer, two Long Short-Term Memory (LSTM) layers, and three Fully Connected (FC) layers. The proposed method is evaluated on a TerraSAR-X image acquired in the Fukushima region, Japan, before and after tsunami. We have used only 200 labeled data (100 for the "change" class and 100 for the "no-change" class). As benchmark method, we have considered a Support Vector Machine (SVM) classifier. The experimental results confirmed the effectiveness of the proposed approach.

1. Introduction

Change Detection (CD) aims to identify land-cover changes occurred between two considered acquisition times in multi-temporal remote sensing images acquired over the same geographical area [1], [2], [3]. It is an important application in both civil and military fields [4]. In the particular case of binary change detection, the goal is to produce binary maps in which changed and unchanged areas can be discriminated [1], [2]. The increased interest in multi-temporal image analysis is due to the increased number of satellites with increased revisit time that allow the acquisition of either long image time series or frequent bi-temporal images. Binary change detection is a key technology for both damage assessment after disasters and monitoring the environment. Examples of natural disasters are related to floods, earthquakes, avalanches, hurricanes, forest fires, and tsunamis [1], [4], [5]. In this paper, we focus the attention on Synthetic Aperture Radar (SAR) images; their active capabilities make the acquisition independent of sunlight, cloud coverage and weather conditions. Thus, SAR imagery is an ideal source for performing change detection [1], [2], [3], [4], [5]. However, due to the presence of speckle noise, SAR image change detection tends to be more difficult than change detection on optical images.

The complexity of collecting ground reference samples, makes unsupervised approaches the most relevant in the field [1], [2]. However, supervised change detection has the advantage of exploiting the semantic of labeled samples for modeling the changed and unchanged pixels. In many cases, ground reference labels are assumed to be available for only a small portion of the scene. This is a realistic assumption for disaster relief or monitoring effort [3]. Since in such cases the number of labeled pixels is much smaller than the total number of image pixels, the framework is called "*weakly-supervised*" [3]. For designing an accurate classifier, the small training set problem should be solved [6]. The training set must contain a sufficiently large number (which depends on the number of parameters to estimate in the learning of the system) of good quality samples. In this paper we address this requirement by proposing a strategy to generate an appropriate number of good quality synthetic samples.

Many tasks can be solved by designing the right set of features to be extracted for each specific task [7]. *Deep learning* solves this central problem seeking to exploit the unknown structure in the input distribution to discover effective representations, with higher level learned features defined in terms of lower level features [3], [7]. A variation of recurrent neural networks with the so-called Long Short-Term Memory units (LSTMs), has been proposed in [8] as a solution to the vanishing gradient problem. LSTMs help preserve the error that can be backpropagated through time and layers. By maintaining a more constant error, they allow recurrent nets to continue to learn over many time steps (over 1000), thereby opening a channel to link causes and effects remotely [8]. LSTM networks are well-suited to classifying, processing and making predictions based on time series data [8].

An interesting traditional model of neural networks is that of the Self-Organizing Map (SOM), also called Kohonen network [9]. In [2] authors present a new neural supervised classification model called concurrent SOMs (CSOM) that has proved to have very good performances for change detection in remote sensing imagery. CSOMs have been also recently proposed for change detection to generate virtual samples, to the purpose to improve the training for Multi-Layer Perceptron [10] and SVM [11] classifiers.

In this this paper we propose to use an *Ensemble of Self-Organizing Maps (ESOM)* as a neural *Synthetic Data Generator (SDG)* useful to train a change detection classifier in SAR remote sensing images. The ESOM substitutes the initial small set of authentically labeled samples by a larger set of more informative and filtered synthetic samples. The proposed SAR change detection model uses a simple *Deep Neural Network (DNN)* architecture with six layers based on *Long Short-Term Memory* units (*LSTMs*). The proposed method is implemented and evaluated using a TerraSAR-X bi-temporal image set acquired in the Fukushima region, Japan, before and after tsunami.

2. Proposed STDG-ESOM-DNN Change Detection Algorithm

Let us assume to have a bi-temporal SAR image corresponding to the time instants t_1 and t_2 and a small set of samples labeled as "change" and "no-change". The proposed method has the phases shown below.

2.1. Concatenation of bi-temporal images

Feature concatenation is used to build the feature vector to be given as input to the classifier. Let $a_{ij}^{(1)}$ and $b_{ij}^{(2)}$ be the feature vectors characterizing each spatial position in the M×M bi-temporal image acquired at times t_1 and t_2, respectively. After concatenation, each spatial position will be modeled by a feature vector V_{ij}^T defined as

$$V_{ij}^T = (a_{ij}^{(1)}\ b_{ij}^{(2)})^T,\ (i, j=1...M)$$

2.2. Synthetic Data Generation using an Ensemble of Self-Organizing Maps (STDG-ESOM)

This phase contains the following steps (Fig. 1).

Step 2.2.1. Building two pattern subsets.

One splits the original labeled data set into two pattern subsets, corresponding to classes "change" and "no-change".

Step 2.2.2. Training of each SOM(k) module (k=1,2)

One uses the SOM unsupervised algorithm [9] to train each of the two SOM modules. Namely, each SOM module has been trained only with the samples having the same class ("change" or "no-change") with the neural module label.

Step 2.2.3. Synthetic training data generation.

After training, the weight vectors of the SOM module neurons have become the final synthetic training data for change detection classifier, which substitute the input original labeled samples.

2.3. DNN as change detector

We have trained a Deep Neural Network (DNN) classifier for change detection using the synthetic dataset derived by the STDG-ESOM technique. At this stage we used a simple DNN architecture having the following six layers:

1. Input layer (2 neurons)
2. Long Short-Term Memory (LSTM) layer of N neurons with dropout (DLSTM)
3. LSTM layer of N neurons with dropout (DLSTM)
4. Fully Connected (FC) layer of N neurons with dropout (DFC) and RELU activation function
5. FC layer of N neurons with dropout (DFC) and RELU activation function
6. Output FC layer with 2 neurons ("change" and "no-change") and softmax activation function.

As a benchmark classifier, we have chosen the Support Vector Machine (SVM).

We have evaluated the two considered classifiers (DNN/SVM) performances when training them with *STDG-ESOM data* versus training them with the *original small authentic labeled data set*. Each classifier provides a binary decisions ("change" or "no-change"), to obtain the change map compared with the ground truth map.

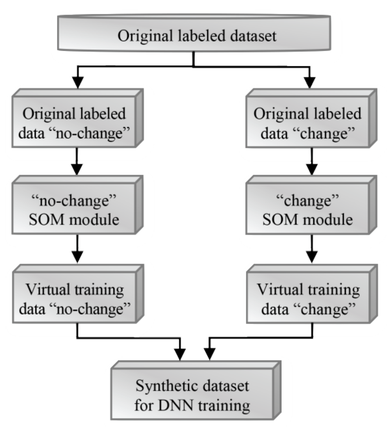


Fig. 1. STDG-ESOM model

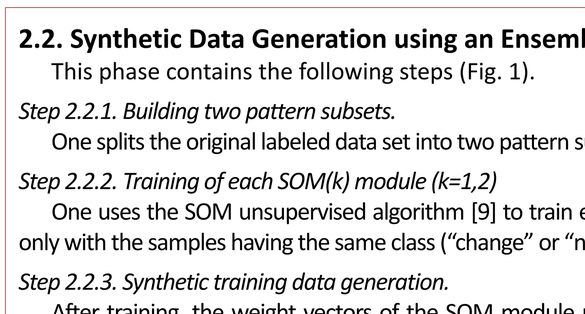
Table 1. Change detection performances of the STDG-ESOM-DNN method (choosing best kappa as target) as a function of both te STDG factor and also of the DNN neuron number (**N=200** original labeled data)

Symmetrical SOM module size	No of ESOM generated training data	STDG Factor γ	No of neurons for each of the {2,3,4,5} DNN layers	Kappa accuracy	Overall Accuracy (OA) [%]	Missed Alarm (MAR) [%]	False Alarm (FAR) [%)]
7 x 7	98	0.49	161	0.813	92.36	4.35	15.61
10 x 10	200	1	110	0.809	92.24	4.18	16.42
14 x 14	392	1.96	148	0.797	91.62	5.54	15.24
17 x 18	612	3.06	91	0.825	92.63	6.49	9.50
22 x 23	1012	5.06	120	0.837	93.05	7.19	6.38
26 x 27	1404	7.02	57	0.856	94.17	2.66	13.50
29 x 29	1682	**8.41**	**57**	**0.889**	**95.51**	**1.46**	11.83
31 x 31	1922	9.61	70	0.832	92.72	8.68	**3.91**

3. Experiments and Results

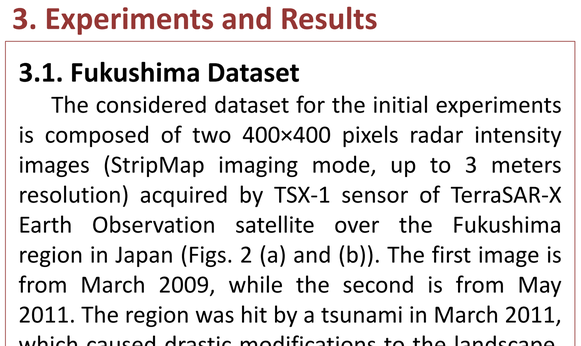
3.1. Fukushima Dataset

The considered dataset for the initial experiments is composed of two 400×400 pixels radar intensity images (StripMap imaging mode, up to 3 meters resolution) acquired by TSX-1 sensor of TerraSAR-X Earth Observation satellite over the Fukushima region in Japan (Figs. 2 (a) and (b)). The first image is from March 2009, while the second is from May 2011. The region was hit by a tsunami in March 2011, which caused drastic modifications to the landscape. The dataset contains 46836 pixels of change (~29.27%) and 113164 unchanged pixels (~70.73%). The reference map has been labeled by experts using a photointerpretation method.



(a) (b)

Fig. 2. Fukushima TerraSAR-X bi-temporal image of 400x400 pixels. (a) March 2009. (b) May 2011.



(a) (b)

Fig. 3 (a) Change map obtained by the STDG-ESOM-DNN with γ=8.41 and N=57 (OA of 95.51% and kappa of 0.889); (b) Reference map (black pixels are changed; white pixels are unchanged).

3.2. Experimental Setup

The training parameters of the DNN have been defined as follows.

a) For SDG-ESOM-DNN:
- (dropout) DLSTM = 0.42; • (dropout) DFC = 0.17;
- number of epochs = 10; • learning rate = 0.00019;
- momentum = 0.94; • batch size = 8.

b) For DNN trained with authentic data (without ESOM):
- (dropout) DLSTM = 0.19; • (dropout) DFC = 0.15;
- number of epochs = 46; • learning rate = 0.00011;
- momentum = 0.85; • batch size = 8.

Regarding VTDG-ESOM, each ESOM module uses a sheet architecture and rectangular topology.

We have evaluated the SAR image change detection performances in terms of Overall Accuracy (OA), Missed Alarm Rate (MAR), False Alarm Rate (FAR) (all in (%)) and also in terms of kappa.

We have also defined $\gamma = STDG\text{-}ESOM\ Factor = \frac{(No\ of\ STDG\text{-}ESOM\ synthetic\ data)}{(No\ of\ authentic\ labeled\ data)}$ for analyzing the effects of the number of extracted synthetic samples with respect to the original ones on performances.

Table 2. Change detection performances of the STDG-ESOM-DNN method (choosing best kappa as target) as a function of the number of neurons for each of the {2,3,4,5} DNN layers (size of the two ESOM modules: 29x29; STDG factor γ: 8.41**; N=200** original labeled data)

No of neurons for each of the {2,3,4,5} DNN layers	Kappa	Overall Accuracy (OA) (%)	Missing Alarm Rate (MAR) (%)	False Alarm Rate (FAR) (%)
41	0.752	89.57	8.47	15.17
49	0.695	87.90	5.59	27.84
57	**0.889**	**95.51**	**1.46**	11.83
61	0.733	89.26	5.51	23.39
67	0.799	91.98	3.07	19.99
80	0.738	88.12	15.20	3.83
92	0.795	91,30	8,13	10,08
100	0.755	89.61	9.11	13.50
250	0.775	90.85	5.14	18.87

Table 3. Best change detection performances (choosing best kappa as target) as a function of classifier type (**N=200** original labeled data=100 *changed* +100 *unchanged*).

Classifier	Training	No of neurons for each of the {2,3,4,5} DNN layers	STDG Factor γ	Kappa	Overall Accuracy (OA) (%)	Missed Alarm (MAR) (%)	False Alarm (FAR) (%)
DNN	STDG-ESOM	57	8.41	**0.889**	**95.51**	**1.46**	11.83
	authentic data	89	-	0.831	92.62	9.11	3.19
SVM	STDG-ESOM	-	0.49	0.579	80.89	20.39	15.99
	authentic data	-	-	0.542	78.57	24.36	14.35

3.3. Experimental Results

From the above described Fukushima bi-temporal SAR image dataset of 400×400 pixels, we have used only 200 authentic labeled pixels (100 "changed" and 100 "unchanged"), representing 0.125% of the total. These samples have been used either for direct training or for STDG-ESOM. The test set is the whole bi-temporal image. The quantitative experimental results are given in Tables 1-3, while the qualitative results are shown in Fig. 3.

Table 1 shows the influence of SOM module size (STDG factor) on the change detection performances of the STDG-ESOM-DNN method. One can observe that the best performance means to choose the STDG factor value equal to 8.41 (corresponding to SOM module of 29×29 neurons).

In Table 2, one can evaluate the influence of the number of neurons (for each of the layers 2,3,4,5 of the DNN) on the STDG-ESOM-DNN performances. It is easy to see that the optimum number corresponds to N=57 neurons (implying maximum kappa, maximum OA, and minimum MAR).

Table 3 shows a comparison of best performances for each of the considered classifiers (DNN/SVM) using STGN-ESOM training data versus using authentic labeled data. One can see that the advantage of STDG-ESOM-DNN over DNN trained with authentic training data only consists in increasing the OA from 92.62% to 95.51%, the kappa from 0.831 to 0.889, as well as decreasing MAR from 9.11 to 1.46. Applying STDV to the SVM classifier has also a positive effect, leading to an increase of OA equal to 2.32%. The advantage of DNN over SVM is also significant.

The change map obtained by the STDG-ESOM-DNN for the parameters leading to best performances is given in Fig. 3.

4. Concluding Remarks

This paper has presented a neural technique for *weakly-supervised change detection* in SAR images based on the idea to replace the initial small set of labeled data with a set of STDG-ESOM synthetic neural data, generated by an ensemble of self-organizing maps. These synthetic data are used to train a deep learning neural network classifier consisting of six layers: input layer, two LSTM layers, and three FC layers.

In our experiments we used the proposed method with only 200 labeled pixels on a bi-temporal 400×400 pixels SAR image to generate a synthetic data set for training a machine learning classifier. The results have proven the effectiveness of the proposed STDG-ESOM-DNN method in terms of change detection accuracy. The best performance corresponds to a kappa of 0.889, an OA of 95.51%, and a MAR of 1.46%.

As expected, the performances of the proposed STDG-ESOM-DNN model have proved to be influenced both by the size of the ESOM modules, and also by the parameters of the DNN system (number of neurons). We have defined the STDG-ESOM factor **γ** as the ratio between the number of ESOM generated synthetic data and the number of input authentic data. In our experiments the best performance with the STDG-ESOM has been reached for γ=8.41. This corresponds to substitute the small number of 200 original data with an increased number of 1682 synthetic samples, leading to an improved quality of DNN training.

As future development we plan to extend the experimental analysis to a larger data set.