

CLASSIFYING GRAY-SCALE SAR IMAGES: A DEEP LEARNING APPROACH

Haoxiang Wang

Department of Electrical and Computer Engineering, Cornell University,
Ithaca, New York

ABSTRACT

Classifying Gray-scale differencing SAR images into two classes is very difficult due to the changeable impacts caused by the season, the imaging condition and so on. To optimize the state-of-the-art algorithms and to deal with the mentioned difficulty, a novel unsupervised classification algorithm is proposed based on deep learning, where the complex correspondence among the images is built up by Auto-encoder Model. With the proper usage of deep neural network model, we could classify differencing SAR images into two classes more accurately and preciously. Experiments well demonstrate the effectiveness of the proposed approach.

KEYWORDS

Deep Learning, Auto-encoder Model, SAR Images, Classification, Unsupervised Learning

1. INTRODUCTION

Classifying Gray-scale differencing images into two classes is very important for the practical applications [1-2] such as disaster monitoring, target tracking, urban planning, etc. However, change detection of SAR images is more difficult due to the ambiguity in defining “change”. For instance, due to the seasonal changes, the atmosphere condition and/or view-angle variation, the appearances of the same object in the multi-temporal images varies greatly[11]. Compared to the changes of interest such as the addition and removal of the buildings, the changes caused by the above impacts are trivial and undesired. In consequence, it is important to compensate the above impacts before feature comparison.

Many approaches are proposed to deal with the above difficulty. For example, Volpi et al.[3] suggested to use the concatenated features. Falco et al. [4] applied a high-level feature called morphological attribute profiles. Nielsen [5] proposed a Algorithm called IR-MAD to model the linear transformation between different spectral bands, which is implemented by the evaluating the maximum variance between the bases of the subspaces extracted from the multi-temporal images. Li et al suggest to use the metric learning[6] approach to classify. Despite of the novelties of the traditional approaches, most of them are based on the explicit assumption that the transformations between multi-temporal spectral bands are linear transformations. Exactly, for multi-temporal SAR images, the spectral transformation between the unchanged areas is very complex, therefore the traditional linear transformation is limited to capture the complex correspondence.

To our great surprise, humanity could deal with the mentioned impacts intuitively. The real reason is that the connections between brain cells are very deep and the nonlinear spectral transformations between the multi-temporal SAR images could be accurately learned. In recent years, deep learning approach has been proven to be considerable after Hinton's work[6]. Different from the traditional learning approaches, deep learning is promising in simulating human brain network in learning complex correlations between features based on the deep hidden layer, which has got great usage and success in signal and image processing area. Inspired by the powerful utility of deep learning in feature selection and learning, a novel algorithm is proposed for SAR image two-class classification., where the complicated transformations between the features are learned implicitly.

Compared with the traditional methodologies, the novelty of the proposed approach lies in the feature learning fashion: the learning procedure is implicit, it is based on the data in consideration. This paper is organized as follows. In section 2, we elaborate the proposed algorithm step following step. In section 3, sufficient experiments are conducted. Section 4 lists the conclusion.

2. OUR PROPOSED APPROACH

The key point of our proposed approach is to establish the correspondence between the SAR images with deep learning such as Gaussian-Bernoulli, RBM to achieve the difference imaged based on the transformed images and also get the change map with clustering approach. In Fig.1 , we define the proposed approach into three steps: feature learning, feature comparison and clustering.

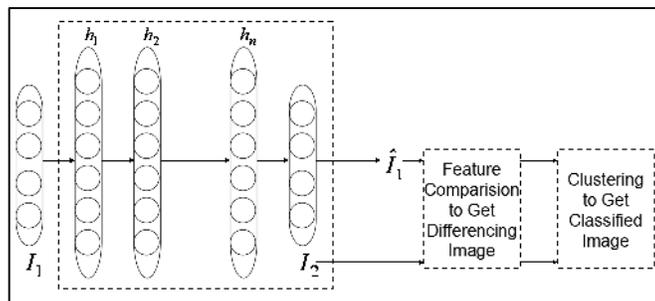


Figure 1. Framework of our proposed approach

2.1. Feature learning based on auto-encoder model

In general, for the co-registered images and , the spectral transformation caused by the outer impacts such as seasonal changes is nonlinear. To solve the problem, we use the auto-encoder model to get the correspondence of features for two important reasons: (1) Auto-encoder model is an undirected graph with symmetrical architecture, so we could take the advantages of graph theory application; (2) Auto-encoder model contains a automatic feature extraction process, which lead to the optimization of the state-of-the-art algorithms. Restricted Boltzmann Machines is the core component of the auto-encoder model. We should introduce the RBM first.

The undirected connection graph of RBM is shown in Fig.2. v_i is the node in visible layer, h_j is the hidden layer node, w_{ij} is the weight coefficient between v_i and h_j . Given the visible variables v_i while $i \in [0, M]$, the hidden variables h_j are independent while $j \in [0, N]$.

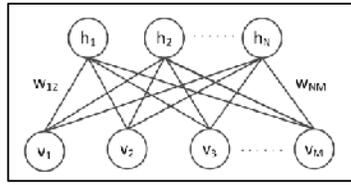


Figure 2. The connection graph of RBM

Let θ denote all the parameters $\{w, b, c\}$ in RBM model. In order to explore the θ , the log-likelihood of $P_{\theta}(v)$ should be maximized. Here, $P_{\theta}(v)$ is determined by the energy of the whole model, which is shown in Eq.(1). \hat{v} and \hat{h} represent every possible value of v and h .

$$P_{\theta}(v) = \sum_h \exp(-E(v, h)) / \sum_{\hat{v}, \hat{h}} \exp(-E(\hat{v}, \hat{h})) \quad (1)$$

The energy function of the whole model in Eq.(1) is determined by every joint configuration and the biases corresponding to every node. Let b_i and c_j respect for the biases of v_i and h_j respectively. We could then get the following formula.

$$E(v, h) = - \sum_{i=1}^M \sum_{j=1}^N v_i w_{ij} h_j - \sum_{i=1}^M v_i b_i - \sum_{j=1}^N h_j c_j \quad (2)$$

The auto-encoder model is consist of RBMs. To efficiently train an auto-encoder model, two phases should be fulfilled: pre-training and fine-tuning[7]. In Fig 3, we illustrate the diagram of a stack of three RBMs. In pre-training phase, the dataset is input to the visible layer of RBM₁. After training the first RBM, the value of its hidden layer is obtained and supplied inturn to the visible layer of RBM₂. Every RBM is trained one by one in the same methodology. At a global scene, the hidden layer of RBM₁ is the visible layer of RBM₂, and so on.

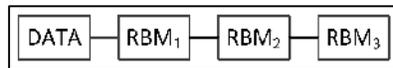


Figure 3. Diagram of pre-training phase

In the fine-tuning phase, firstly, the network in pre-training phase is unrolled and a rudiment of autoencoder is formed. In Fig. 4, the three RBMs in Fig. 3 are expanded to a stack of six RBMs. In this network, the visible layer of RBM₄ is just the hidden layer of RBM₃. The connection weight matrix w_4 of RBM₄ will be equal to the transpose of the connection weight matrix w_3 of RBM₃, and the visible layer biases b_4 and hidden layer biases c_4 of RBM₄ will be equal to c_3 and b_3 , respectively. Then back propagation of the error between the output and the input data is used to tune all of the parameters in the whole model [9].

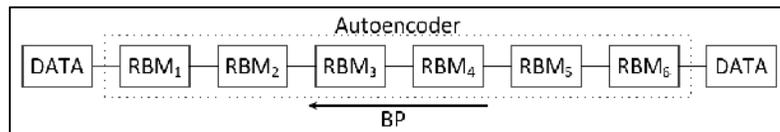


Figure 4. Diagram of fine-tuning phase

In RBM, $P(h|v)$ and $P(v|h)$ subjects to Bernoulli distribution with the prior knowledge that input and output are both binary data. Considering that the input data of the two-class classification in continuous, the Gaussian-Bernoulli RBM(GBRBM) is employed here to learn features. Specifically, $P(h|v)$ is subject to the Bernoulli distribution, while $P(v|h)$ will be employed to constrain the Gaussian distribution[7]. The energy function of GBRBM can be formulated as follows[8].

$$E(v,h) = -\sum_{i=1}^M \sum_{j=1}^N \frac{v_i}{\sigma_i} w_{ij} h_j - \sum_{i=1}^M \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{j=1}^N h_j c_j \quad (3)$$

The feature learning procedure is consist of training step and detection step. Similar to the auto-encoder model, the training step in our algorithm are conducted into the pre-training and fine-tuning process. Gig.5 shows the pre-training phase in our experiment. D_1 and D_2 represent the set of patches extracted from every pixel in I_1 and I_2 , respectively. We use D_1 and D_2 train a GBRBM and RBM contemporary, similar to the pre-training procedure.



Figure 5. The pre-training step in our method

The fine-tuning phase in our method is quite different from the auto-encoder model. Fig 6. Shows the difference. First, we supply D_1 from the left end of the model, and then make use of the difference between the output data D_2 to tune the parameters. Later, we input D_2 from the right end of the model before calculating the difference between the output data and D_1 to tune the parameters. These two steps are implemented alternately, till the whole model is tending to be stable.

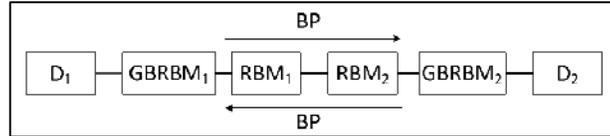


Figure 6. The fine-tuning step in our method

2.2. Compare features to get differencing map

The best use of two-class classification is “Change Detection” which means classify the gray-scale image into to classes, changed or unchanged. We train our model under the usage of detecting changes. In the training step, the model has learned the pervasive ‘trivial changes’ through the whole images. Although D_1 and D_2 contain the patches extracted from the ‘real changes’ areas, these patches do not follow the same change rules. In addition, the changed areas are not the majority of the whole images. So they can be viewed as noises and do not influence the model obviously. In the detection phase, we input D_1 and D_2 from the left and right ends respectively and get two output data R_1 and R_2 , which can be viewed as the simulate data of D_1 and D_2 going though ‘trivial changes’. So the difference between the D_1 and D_2 , as well as R_1 and D_2 could indicate the “real changes” between I_1 and I_2 . In our method, we use wavelet fusion technic[10] to create differencing gray-scale change map.

Considering SAR Images I_1 and I_2 are acquired over the same geographical region at two different times t_1 and t_2 . The mean-ratio image could be defined as:

$$I_m(i, j) = 1 - \min\left(\frac{\mu_1(i, j)}{\mu_2(i, j)}, \frac{\mu_2(i, j)}{\mu_1(i, j)}\right) \quad (4)$$

where $\mu_1(i, j)$ and $\mu_2(i, j)$ represent the local mean value. In the similar way, we define log-ration operator as follows:

$$I_l = |\log(I_2 / I_1)| = |\log I_2 - \log I_1| \quad (5)$$

With the wavelet technics, we take wavelet fusion in to consideration to create a better performance of getting differencing image. The fusion rule is as follows:

$$D_{LL}^F = \alpha D_{LL}^l + \beta D_{LL}^m \quad (6)$$

$$\alpha = \min(|D_{LL}^l|, |\bar{D}_{LL}|) / \max(|D_{LL}^l|, |\bar{D}_{LL}|) \quad (7)$$

$$\beta = (1 - \min(|D_{LL}^l|, |\bar{D}_{LL}|)) / \max(|D_{LL}^l|, |\bar{D}_{LL}|) \quad (8)$$

$$D_e^F(i, j) = \begin{cases} D_e^m(i, j), & \sigma_e^m(i, j) < \sigma_e^l(i, j) \\ D_e^l(i, j), & \sigma_e^m(i, j) \geq \sigma_e^l(i, j) \end{cases} \quad (9)$$

2.3. Clustering

General algorithm used to classify the differencing is through thresholding methodology which has gained great success in many other fields. However, a better approach is through fuzzy clustering. Here, we take optimized FLICM[10] algorithm into conduction.

The characteristic of FLICM algorithm is the use of fuzzy local similarity measurement aiming at guarantee noise intensiveness and detail preservation. Particularly, a novel fuzzy factor G_{ki} is introduced to enhance clustering accuracy and efficiency. The mathematical formula is:

$$G_{ki} = \sum_{j \in N_i} \frac{1}{d_{ij} + 1} (1 - u_{kj})^m \|x_j - v_k\|^2 \quad (10)$$

By using defined G_{ki} , the objective function of the FLICM could be defined as:

$$J_m = \sum_{i=1}^N \sum_{k=1}^c [u_{ki}^m \|x_i - v_k\|^2 + G_{ki}] \quad (11)$$

In addition, the calculation of the membership partition matrix and the cluster centers is performed as follows:

$$u_{ki} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_i - v_k\|^2 + G_{ki}}{\|x_i - v_j\|^2 + G_{ji}} \right)^{1/(m-1)}} \quad (12)$$

$$v_k = \frac{\sum_{i=1}^N u_{ki}^m x_i}{\sum_{i=1}^N u_{ki}^m} \quad (13)$$

Here, we describe the algorithm with flowchart below:

<i>Step1</i>	Set the number c , m and stopping condition ε .
<i>Step2</i>	Initialize randomly the fuzzy partition matrix.
<i>Step3</i>	Ste the loop counter $b = 0$.
<i>Step4</i>	Compute the cluster prototypes using (13).
<i>Step5</i>	Calculate the fuzzy partition matrix using (12).
<i>Step6</i>	$\max\{U^{(b)} - U^{(b+1)}\} < \varepsilon$ to stop; otherwise return to step (4).

3. EXPERIMENT

In order to apply our methodology into reality, we firstly simulate our algorithm in Fig 8, Fig 7 is our test dataset, the Bern dataset and the Ottawa dataset.

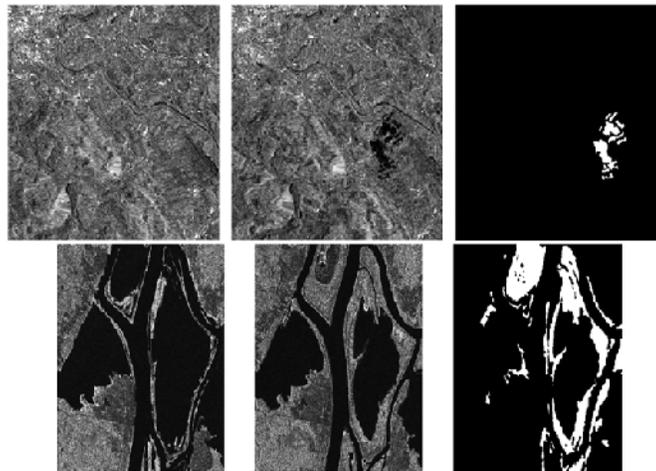


Figure 7. The testing dataset with ground truth

The visualable testing result in Fig 8 compares the proposed deep-learning based classification method with some state-of-the-art approaches to classify gray-scale images into two class which indicates the effectiveness and robustness of our method.

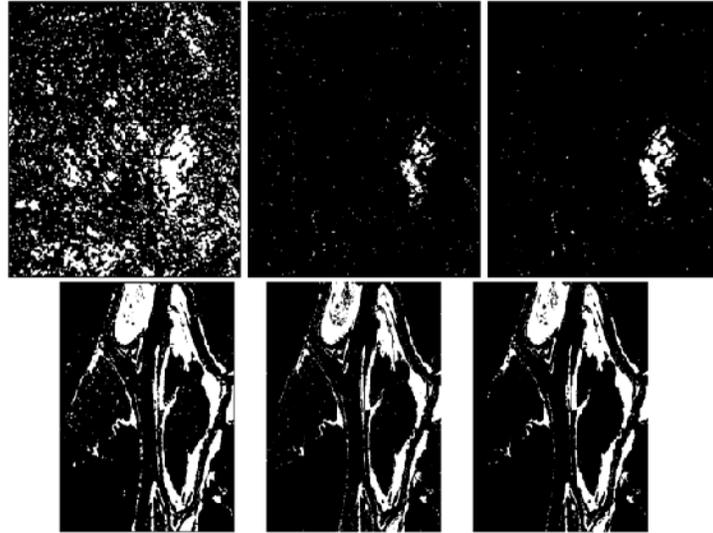


Figure 8. The testing result, first picture is the result of K-mans, the other two are the result of FCM and our proposed method.

4. FINAL CONCLUSION

In this paper, we propose an unsupervised method for SAR images two-class classification with auto-encoder model. This method is based on the assumption that “trivial changes” can be learned while “real change” can not. In the experiments, the effectiveness of our method is proved and we get a better result than two sample methods. In the future work, we focus on localizing the ‘trivial changes’, context information, robust distance measure etc.

ACKNOWLEDGEMENTS

The author would like to thank the helpful discussion with friends Weizhe Hua and Yi Zheng in University of Southern California, USC.

REFERENCES

- [1] D. Lu, P. Mausel, E. Brondizio, E. Moran, “Change detection techniques,” *International Journal of Remote Sensing* 25(12), 2365-2401(2004).
- [2] A. Singh, “Review Article Digital change detection techniques using remotely-sensed data,” *International journal of remote sensing* 10(6), 989-1003(1989).
- [3] M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone, “Supervised change detection in VHR images using contextual information and support vector machines,” *International Journal of Applied Earth Observation and Geoinformation*, 77-85(2013).
- [4] N. Falco, M. D. Mura, F. Bovolo, J. A. Benediktsson, and L. Bruzzone, “Change detection in VHR images based on morphological attribute profiles,” *IEEE Geoscience and Remote Sensing Letters*, 1-5(2012).

- [5] A.A. Nielsen, "The regularized iteratively reweighted mad method for change detection in multi-and hyperspectral data," IEEE Transactions on Image Processing, 463-478(2007).
- [6] Wang, Haoxiang, Ferdinand Shkjezi, and Ela Hoxha. "Distance metric learning for multi-camera people matching." In Advanced Computational Intelligence (ICACI), 2013 Sixth International Conference on, pp. 140-143. IEEE, 2013.
- [7] G. E. Hinton, and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," Science, 504-507(2006).
- [8] N. wang, J. Melchior, and L. Wiskott, "An Analysis of Gaussian-Binary Restricted Boltzmann Machines for Natural Images," European Symposium on Artificial Neural Networks, 287-292(2012).
- [9] G. E. Hinton, S. Osindero, and Y.W. The, "A fast learning algorithm for deep belief nets," Neural computation 18(7), 1527-1554(2006).
- [10] Gong, Maoguo, Zhiqiang Zhou, and Jingjing Ma. "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering." Image Processing, IEEE Transactions on 21, no. 4 (2012): 2141-2151.
- [11] Q. Cai, Y. Yin, H. Man, DSPM: Dynamic Structure Preserving Map for Action Recognition, ICME, 2013.