Very High Resolution SAR Change Detection with Siamese Networks

Shreya Sharma[1] and Masato Toda[2]

***Abstract* :** This paper proposes a change detection method for very high resolution (VHR) Synthetic Aperture Radar (SAR) images using Siamese network. VHR SAR provides detailed surveillance of a crowded area through detection of very small objects day-and-night and under all weather conditions. However, due to inhomogeneous scattering by an object sub-parts in VHR, conventional methods detect many irrelevant changes. Contrastively, in optical images, Siamese network can suppress irrelevant changes by learning discriminative features automatically. In this paper, we implement Siamese network in VHR SAR images for the application of parking lot monitoring. Experiments on 1 m resolution images show that the proposed method improves f-measure by 15% over conventional methods.
Keywords : very high resolution, change detection, synthetic aperture radar, Siamese network.

1. Introduction

Change detection is a widely researched problem in satellite image processing and is considered an important preliminary analysis before any advanced analysis such as object recognition. Given a pair of images, it aims to infer changes occurred between them over a period of time. Among remote sensing technologies, Synthetic Aperture Radar (SAR) is an ideal technology for change detection because of its ability to capture images even under bad weather and no-sunlight conditions. With the advent of very high resolution SAR sensors such as TerraSAR-X and ASNARO-2, it has become possible to capture changes due to small objects such as cars, humans and containers. Change detection of such small objects is of interest because it helps in effective monitoring of crowded and dynamic areas, and thus provides detailed surveillance.

Traditional methods of change detection employ a pixel-to-pixel based difference between images wherein each pixel of the first image is compared to the corresponding pixel of the second image [1] [2]. These methods, however, do not work well in very high resolution SAR images because the pixel is sensitive to SAR artifacts (shadow, layover and speckle noise) and may show a change even if there is no semantic meaning of that change. This is caused due to inhomogeneous scattering by object sub-parts [3]. To tackle this, feature-to-feature based difference has been proposed where features of the target object are manually modelled using domain knowledge [3]. A filter to extract features is applied directly to the images, and the two results are compared to detect changes due to the object. However, these methods are not robust to changes in object orientation and requires domain knowledge abouts the object.

[1]Member: Data Science Research Labs. NEC Corporation
[2]Non-Member: Data Science Research Labs. NEC Corporation
(Address: 1753, Shimonumabe, Nakahara-ku, Kawasaki, Kanagawa, 211-8666, Japan)
(Contact: Tel. +81-44-431-7658,
E-mail. s-sharma@ap.jp.nec.com)

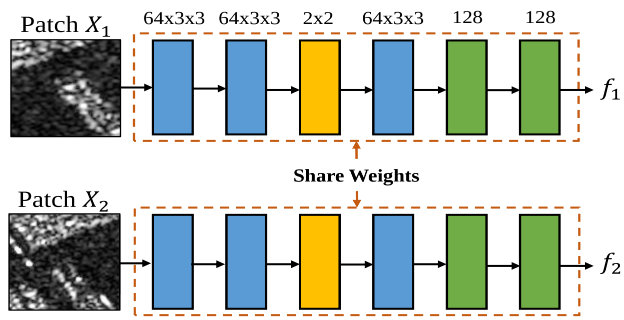

Fig. 1. Architecture of the proposed Siamese network. Color-code used: blue = Conv+ReLU, yellow = Max Pooling, green = Fully Connected Layer.

Over a past few years, neural-network based methods have shown dramatic improvement in a variety of computer vision tasks such as object detection, object recognition, image generation and so on. Neural-network has the capability to automatically extract features of an object robust to changes in orientation and noise. One type of neural-networks, called Siamese network, is well suited for the task of change detection because it inputs a pair of patches, and aims to classify the central pixel of the patch into change or no-change based on neighborhood values. Some of the recent works which have shown the potential of Siamese network for change detection in optical images are [4] and [5]. However, the effectiveness of Siamese network has not been tested in very high resolution SAR images.

In this paper, we address the problem of change detection in very high resolution SAR images using Siamese network. Unlike traditional methods, the Siamese network eliminates the pixel-to-pixel based difference, and extracts robust features directly from the image pair without any additional domain knowledge. As a result, the false changes detected in very high resolution images due to SAR artifacts and variations in object are reduced.

The remainder of the paper is organized as follows. Section 2 describes in detail about the Siamese network with training and inference. Section 3 contains quantitative and qualitative evaluation with previous change detection methods. Finally, Section 4 provides conclusion of the paper.

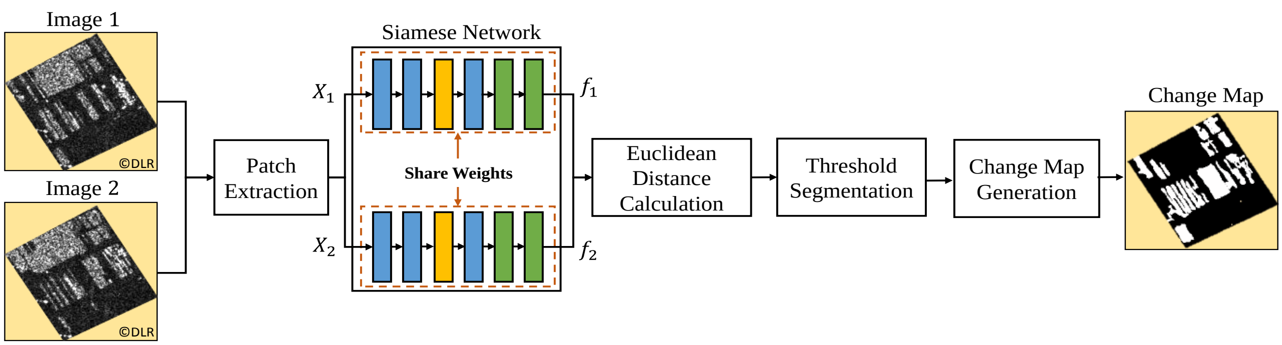

Fig. 2. Schematic of inference with the proposed Siamese network.

2. Methodology

2.1 Siamese Network

Siamese network is a type of neural-network used for comparing image patches where the network has two identical feature extraction branches [6]. Each branch takes as input one of the two image patches, applies a series of non-linearities and encodes the patch into features. In [6], three configurations of the Siamese network have been shown: siamese, pseudo-siamese and 2-channel. Out of these, we have used the siamese configuration in which weights of the two feature extraction branches are shared. As the two images we use for change detection are both SAR and have similar characteristics, sharing weights ensure that both the branches extract features using same approach.

The architecture of our Siamese network, shown in Fig. 1, is inspired by VGGNet [7] since we use a 3x3 filter size in all convolutional layers. This ensures a larger effective receptive field while keeping the number of parameters to be trained less. However, our network is not as deep as the VGGNet because a SAR image comprises much less complex features as compared to an ordinary RGB image. The proposed network contains three convolutional layers (Conv), one max-pooling layer and two fully-connected layers. Each convolutional layer is followed by a batch normalization layer and a dropout layer to prevent overfitting. We use Rectified Linear Units (ReLU) as non-linearity for convolutional layers. Each branch takes as input a patch centered at each pixel of the two SAR images and encodes the patch into a one-dimensional feature vector of size 128.

Two key points in training the Siamese network are patch size and optimization function. The patch size indicates the number of neighborhood pixels taken into account around each pixel of a SAR image. The neighborhood pixels provide a spatial contextual information to extract features of the central pixel. These features are used to make a decision of a change or a no-change class. A small patch size reduces the amount of the contextual information and can be strongly affected by noise in the neighborhood. Whereas a large patch size provides too much contextual information such that the characteristic of the central pixel diminishes which can result in less discriminative features. In this paper, we apply the proposed Siamese network for two image patch sizes - 10×10 and 16×16, results of which are discussed in the Section 3.

2.2 Optimization Function

For training the proposed Siamese network, we need an objective function whose optimization can produce good performance in the feature extraction task. The purpose of training the network is to learn a parametric function G_w that non-linearly maps the patches to points in a feature space such that distance between the points in feature space is small if the patches belong to a no-change class and large otherwise. As shown in [8], a contrastive loss function can learn the parametric function G_w.

Let X_1 and X_2 be a pair of patches centered at each pixel extracted from a pair of multi-temporal SAR images. The Siamese network transforms the patches into feature vectors $G_w(X_1)$ and $G_w(X_2)$, denoted as f_1 and f_2 respectively. A Euclidean distance $D_w(f_1, f_2)$ is computed between the feature vectors as a similarity measure using (1).

$$D_w(f_1, f_2) = \|G_w(X_1) - G_w(X_2)\|_2 \tag{1}$$

A small Euclidean distance means high similarity between the features implying a no-change class whereas a large Euclidean distance means low similarity between the features implying a change class. A short notation of $D_w(f_1, f_2)$ is written as D_w. The process of Euclidean distance calculation is repeated for all the training patches. Then, the contrastive loss $L(W)$ is computed using (2).

$$L(W) = \sum_{k=1}^{P} \Big[(1 - y^k) \frac{1}{2} (D_w^k)^2 + y^k \frac{1}{2} \{\max(0, m - D_w^k)^2\} \Big], \tag{2}$$

where P is the number of training patches, y^k is the true class of the central pixel of k^{th} patch obtained from ground truth and m is a margin set to 1. $y = 0$ denotes no-change and $y = 1$ denotes change shown as black and white respectively in the ground truth. Change pixels contribute only if their parameterized distance is within the margin [8].

Table 1. Change detection results on testing site pair 1.

Method	Precision	Recall	f-measure	AUC
PCA-K [1]	0.20	0.85	0.32	0.64
SAE-K [2]	0.27	**0.89**	0.41	0.75
Siamese-10	0.62	0.63	0.62	0.88
Siamese-16	**0.73**	0.60	**0.66**	**0.93**

Table 2. Change detection results on testing site pair 2.

Method	Precision	Recall	f-measure	AUC
PCA-K [1]	0.46	0.82	0.59	0.68
SAE-K [2]	0.62	**0.84**	0.72	0.80
Siamese-10	0.84	0.69	**0.76**	0.89
Siamese-16	**0.93**	0.64	**0.76**	**0.93**

2.3 Inference

After the Siamese network is trained with the contrastive loss function using training patches, inference is carried out on a new pair of multi-temporal SAR images. The schematic of inference with the trained Siamese network is shown in Fig. 2. First, patches are extracted centered at each pixel from the new pair of SAR images. Then, the trained Siamese network takes as input the patches and outputs two feature vectors f_1 and f_2, respectively. Next, a Euclidean distance is computed between the feature vectors as a similarity measure. The computed Euclidean distance is thresholded to assign a change or no-change class. A threshold value of 0.5 is selected based on the best f-measure rate in training phase. This process is repeated for all the pixels and finally a change map is generated where each pixel is assigned either a change (white) or a no-change (black) class.

3. Experiments

3.1 Dataset

The experimental data consists of 10 SAR images acquired by TerraSAR-X with HH polarization. The images are acquired on different dates between June 27, 2017 and January 1, 2018 with an interval of 21 days between two consecutive acquisitions. The azimuth and range resolution of the images is $1\,m \times 1\,m$. We paired the 10 images chronologically to create 9 image-pairs for change detection. We consider change detection for the application of parking lot monitoring where a change is caused by movement of cars. Parking lot monitoring is important to analyze economic activity of an area such as a retail area to predict future economic trends. For evaluation, we selected 5 parking lot sites from different parts of Japan.

3.2 Training and Testing Data Construction

Out of the 5 parking lot sites, we selected 3 sites for training the network, 1 site for hyper-parameter tuning and 1 site for testing the trained network. We used separate sites for training and testing to ensure that there is no information leakage from the testing data while training the network. We got 27 image-pairs for training, 9 image-pairs for hyper-parameter tuning and 9 image-pairs for testing. To create patches, we cropped the images overlappingly using two different patch sizes (10×10 and 16×16) to examine the effect of patch size on change detection performance. The ground truth (GT) is created through manual interpretation.

3.3 Optimization Settings

We implemented the Siamese network using Keras [9] framework. The training is carried out with Adam optimization using learning rate of 0.0001. The mini batch-size is set to 32 and the number of epochs is set to 20. A dropout ratio of 0.2 is used to prevent overfitting. The training and testing are performed in a single NVIDIA GTX 1080 GPU with 8GB memory.

3.4 Results and Evaluation

We compare the Siamese network with two other conventional methods (PCA-K [1] and SAE-K [2]) on the testing site. The Siamese network with patch size 10×10 is denoted as Siamese-10, while the one with patch size 16×16 is denoted as Siamese-16. The evaluation metrics used for comparison are precision, recall and f-measure with respect to the change class. We also used AUROC (AUC) for comparison since it is a common evaluation metric used for binary classification. We show the results for 3 pairs out of the 9 pairs of testing site for brevity. However, we obtained similar results in the remaining pairs.

Table 1 and table 2 contains the quantitative results for pair 1 and pair 2 respectively where the best results are marked bold. For pair 1, Siamese networks outperforms the PCA-K and SAE-K methods in terms of precision, f-measure and AUC metrics. While the SAE-K method achieves the best recall rate, it performs poorly on the precision metric that results in low f-measure rate. Siamese-16 performed better than Siamese-10, which shows that a large patch size improves features extraction in pair 1. The qualitative results for pair 1 are shown in Fig. 3. From the change maps, we conclude that the Siamese networks significantly reduce the irrelevant changes as compared to other methods. For pair 2 also, Siamese networks performed the best as compared to the conventional methods with the best precision, f-measure and AUC rate. However, unlike pair 1, increasing patch size does not show much improvement. With a large patch size, certain line-shaped changes are missed as observed in Fig. 3 from change map (g) of pair 2. This shows a patch size should be selected based on the characteristic of changes.

Figure 4 shows an illustration of change maps for pair 3 where the true positives, true negatives, false positives and false negatives are color-coded white, black, green and red respectively. From the figure, we observe that the Siamese networks significantly reduce the number of false positives

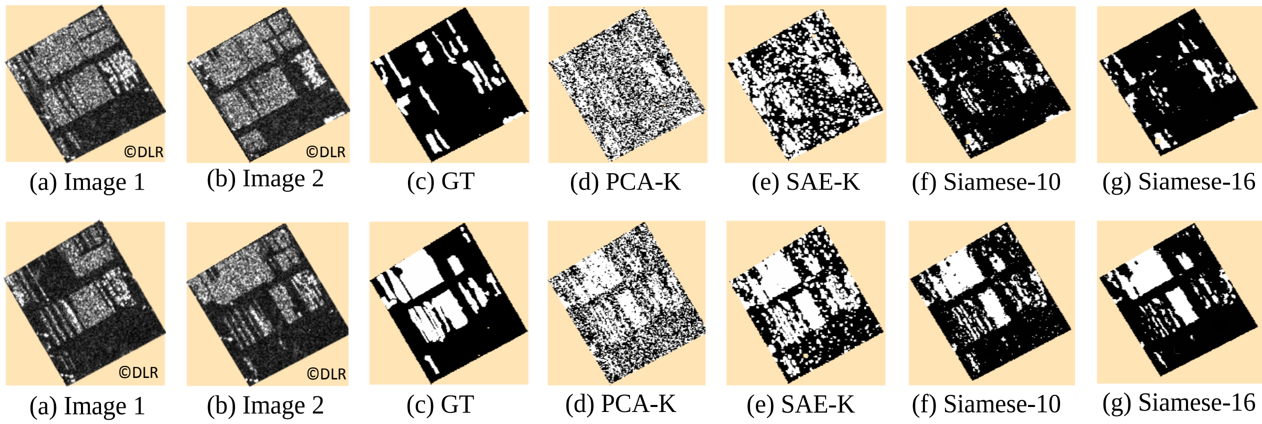

Fig. 3. Comparison of change maps on testing site obtained using the methods. First row: pair 1, second row: pair 2.

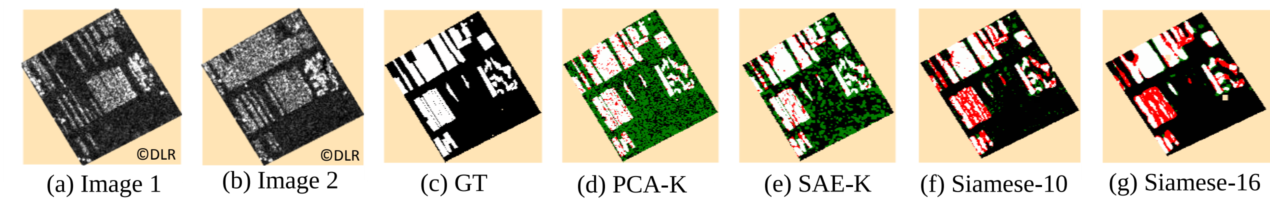

Fig. 4. Color-coded change maps on testing site pair 3. Color code used: white means true positive, black means true negative, green means false positive, and red means false negative.

which indicates that the networks successfully learn features of only relevant changes. Although the number of false negatives increases but overall the change maps generated by Siamese networks are better interpretable. A large patch size can increase false negatives so a trade-off between accuracy and interpretability is needed.

4. Conclusion

In this paper, we presented a change detection method for very high resolution Synthetic Aperture Radar images using Siamese network. Until now, Siamese networks were designed only for optical change detection. We have shown the effectiveness of these networks for VHR SAR using a self-designed architecture trained with a contrastive loss function. Experiments have been conducted for parking lot monitoring with 1 m resolution images. The experiments conclude that the Siamese network outperforms the conventional methods both quantitatively and qualitatively.

References

1) T. Celik: Unsupervised change detection in satellite images using principal component analysis and k-means clustering, IEEE Geoscience and Remote Sensing Letters, vol. 6, no. 4, pp. 772-776, 2009.

2) M. Gong., H. Yang, and P. Zhang: Feature learning and change feature classification based on deep learning for ternary change detection in SAR images, ISPRS Journal of Photogr. and Remote Sensing, no.129, pp.212-225, 2017.

3) F. Bovolo, C. Marin, and L. Bruzzone: A hierarchical approach to change detection in very high resolution SAR images for surveillance applications, IEEE Trans. on Geos. and Remote Sensing, vol. 51, no. 4, pp.2042-2054, 2013.

4) Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu: Change detection based on deep siamese convolutional network for optical aerial images, IEEE Geoscience and Remote Sensing Letters, 14(10), pp.1845-1849, 2017.

5) R. C. Daudt, B. Le Saux, A. Boulch, and Y. Goussea: Urban change detection for multispectral earth observation using convolutional neural networks, IEEE International Geos. and Remote Sensing Symp., pp. 2115-2118, 2018.

6) S. Zagoruyko, and N. Komodakis: Learning to compare image patches via convolutional neural networks, In Proceedings of the IEEE conf. on comp. vision and pattern recognition, pp. 4353-4361, 2015.

7) K. Simonyan, and A, Zisserman: Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.

8) S. Chopra, R. Hadsell, and Y. LeCun: Learning a similarity metric discriminatively, with application to face verification, In CVPR (1), pp. 539-546, 2005.

9) F. Chollet, Keras, 2015.