OPEN ACCESS Freely available online

Journal of Remote Sensing & GIS

Research Article

Partly Uncoupled Siamese Model for Change Detection from Heterogeneous Remote Sensing Imagery

Touati R[1*], Mignotte M[1], Dahmane M[2]

[1]Vision lab of the De´partement d'Informatique et de Recherche Ope´rationnelle (DIRO), Universite´ de Montre´al, Faculte´ des Arts et des Sciences, Montre´al, H3C 3J7, QC, Canada;[2]R&D vision De´partement, Centre de Recherche Informatique de Montre´al (CRIM) Montre´al, QC, Canada

ABSTRACT

This paper addresses the problematic of detecting changes in bitemporal heterogeneous remote sensing image pairs. In different disciplines, multimodality is the key solution for performance enhancement in a collaborative sensing context. Particularly, in remote sensing imagery there is still a research gap to fill with the multiplication of sensors, along with data sharing capabilities, and multitemporal data availability. This study is aiming to explore the multimodality in a multi-temporal set-up for a better understanding of the collaborative sensor wide information completion; we propose a pairwise learning approach consisting on a pseudo-Siamese network architecture based on two partly uncoupled parallel network streams. Each stream represents itself a Convolutional Neural Network (CNN) that encodes the input patches. The overall Change Detector (CD) model includes a fusion stage that concatenates the two encodings in a single multimodal feature representation which is then reduced to a lower dimension using fully connected layers and finally a loss function based on the binary cross entropy is used as a decision layer. The proposed pseudo-Siamese pairwise learning architecture allows to the CD model to capture the spatial and the temporal dependencies between multimodal input image pairs. The model processes the two multimodal input patches at one-time under different spatial resolutions. The evaluation performances on different real multimodal datasets reflecting a mixture of CD conditions with different spatial resolutions, confirm the effectiveness of the proposed CD architecture.

Keywords: Change detection; Deep learning, Heterogeneous Remote Sensing; Multi-source images, Multi-sensor images

INTRODUCTION

In remote sensing imagery, change detection is the process of computing differences in a geographical area by analyzing it at different times. Change detection (CD) problems can be divided into two main types: the monomodal CD problem assumes that the change area occurred between two/multiple images over time under the assumption that these images share the same characteristics i.e. acquired by the same satellite sensor with the same specifications. The multimodal CD problem assumes that the bi-temporal images are acquired by different sensors or with the same sensor but with different specifications. Detecting changes between heterogeneous images is a non-trivial problem as it must take into account multiple sources and characteristics of the acquired data.

This problem is still less explored, although it has recently generated a growing interest in the remote sensing research community, due to the fact that it allows us to exploit, without restriction, the huge amount of heterogeneous data that we can now obtain from the various existing archives including the different types of new and existing earth observation satellites. The technical and practical advantages enable to increase the system performances, and especially to avoid detecting natural changes due to environmental variables such as humidity or

Correspondence to: Touati R, Vision lab of the De´partement d'Informatique et de Recherche Ope´rationnelle (DIRO), Universite´ de Montre´al, Faculte´ des Arts et des Sciences, Montre´al, H3C 3J7, QC, Canada, Email: touatire@iro.umontreal.ca

Received: February 25, 2020; **Acceptance:** March 09, 2020 **Published:** March 16, 2020

Citation: Touati R, Mignotte M, Dahmane M (2020) Partly Uncoupled Siamese Model for Change Detection from Heterogeneous Remote Sensing Imagery. J Remote Sens GIS. 9:271. DOI: 10.35248/2469-4134.20.9.271

phenological state. This challenging task can be viewed as the generalization of the classical monomodal CD problem [1] which is less used asis for solving the same CD problems (e.g environmental monitoring, deforestation, urban planning, land or natural disaster/damage monitoring and management, etc.).

Nowadays, few research works have been addressed in the multimodal CD issue using supervised and unsupervised machine learning methods or digital image processing techniques. The existing research attempts, can be divided into five categories in which we can find parametric models [2], non-parametric methods [1], algorithms based on operators using spatial and temporal similarity measures [3-5], projection based techniques [6-8], and finally machine learning methods [9-11] that will be described in more details since the proposed model belongs to this family.

In parametric methods, a set (or mixture) of multivariate or meta-Gaussian distributions are generally used to model common dependencies between the two imaging modalities, different types of multi-sensor data [2,12-14].

In the category of non-parametric methods, we can mention the energy based model in the least-squares sense that meet new criteria, and designed for satisfying an over determined set of constraints expressed on every pair of pixels existing in the before and after image change as proposed by Touati and Mignotte [1].

Thirdly, methods mainly using invariant similarity measures (such as correlation, mutual information, etc.) by imaging modality in order to estimate at first, the correspondence between the same existing points in the two images and then to identify and detect, in a second time, the zones of change between the two heterogeneous images.

In the fourth category, the projection or simulation techniques try to transform the two heterogeneous images into a new common feature space, in which the two multimodal images share the same statistical properties, and on which classical monomodal CD models can then be applied.

Finally, in the category of machine learning methods, the authors Merkle et al. [15] used an unsupervised learning algorithm called generative adversarial network consisting of two networks; the first whose first network generates a binary map and the second one tries to discriminate between the output of the generator and the output of a binarization algorithm. Liu et al. [9], the authors proposed to train a couple of convolutional neural networks in order to transform the before and after change images in a feature space allowing to calculate a difference map, and then to apply a thresholding algorithm on the resulting map to generate the final binary detection map. Similar to Liu et al. [9], Zhao et al. [11] proposed to build a symmetric neural network consisting of a restricted Boltzmann machine, whose parameters are then updated based on the clustering result. Another method based on a denoising auto-encoder network uses selected features of the difference image to train the network [10].

In the recent proposed approaches, deep learning has become a methodology of choice as the most advanced form of machine learning for image classification, object detection, segmentation and other applications. In particular, convolutional neural network (CNN) is a descriptor learning frame work with a deep architecture that transforms the input data through many layers to extract high level representations from the inputs. Invariant feature representation learning is types of descriptor learning framework, which can be built on CNN using for instance Siamese network [16]. The Siamese CNN architecture is used for patch comparison and refers to two coupled network streams with the same CNN architecture and the same parameters applied to a pair of input data at the same time. In this point of view, the multimodal CD problem can be considered as a binary classification task in which the Siamese-CNN architecture takes as input the two heterogeneous images.

In this work, we are concerned with a heterogeneity problem. We propose a CD model principally designed to deal with different imaging sources under different spatial resolutions and which is well adapted for representing and detecting temporal changes between two heterogeneous remote sensing images. The CD model learns directly a binary classification function from various types of patch pairs coming from different sources, which are processed through two CNN streams that share the same architecture configuration but with partly uncoupled weights between them, in order to extract descriptors independently for each multimodal input patch. The final stage of the proposed model consists to combine the two output descriptors from each stream in a single multimodal representation, which is then used to learn the binary classification cost function. The built model classifies the new temporal input images by processing the input patch pairs in parallel using the learned convolutional streams and the decision network for binary classification.

The rest of this paper is organized as follows: section II describes the designed CD model and its architecture to identify change or non-change input pairs as similar/dissimilar classes. Section III presents the evaluation strategy used to assess the performance of our CD model and the obtained results compared to the state-of-the-art multimodal techniques. Finally, section IV concludes the paper.

MATERIALS AND METHODS

Proposed change detection model

The two/multiple remote sensing input images that correspond to the same geographic area are acquired and co-registered at different time by two/multiple different sensors. Dealing with the characteristics of the different sources of image represents the main challenging issue.

One interesting solution is to design a multimodal CD model with two branches that take as input a pair of images instead of one input, in which the image before and after are fed to two branches allowing us to capture both the spatial and the temporal inter-dependencies. Formally, the task of multimodal change detection can be viewed as a pairwise identification problem, where a pair of non-change/non-change images (samples) are called similar pair, and a pair of non-change/change represent dissimilar pair which represents the difference (in the land use) caused by the event and not by differences in the data sources.

In this case, the pairwise learning approach is more appropriate to verify whether a pair of temporal images corresponds to the similar pair or to the dissimilar pair, i.e. corresponds to the non-change class and to the change class respectively. This can be achieved by training a network based on the similarity of the images in order to learn the similarity between pair of images. Among metric learning approaches, Siamese network has already been successfully used in several applications [16-18] such as change detection, geo-localization, signature verification, one-shot image recognition, face verification, learning image descriptors, and image ranking to name a few. The Siamese architecture consists of two identical subsystems sharing the same set of parameters and a cost function module to quantify the pairwise relationship. The cost function can be defined via a distance metric or a similarity measure. The goal consists to increase the similarity score or to decrease the distance between similar pairs, and dually, to reduce the similarity score or to increase the distance between two dissimilar image patches.

In our case, Siamese network architecture is able to support as input a pair of images. Since the image pair is multimodal, i.e., composed of two different imaging modalities (acquired from different sources), the Siamese network architecture is less effective when the weights are shared between the parallel network streams (parallel subsystems) [19]. Ultimately, using a cost function based on a distance metric or a similarity measure to distinguish between similar and dissimilar pair images is less suitable for evaluating similar pair images coming from two different sources due to the fact that there was not a strong enough correlation between heterogeneous similar pair images.

Inspired by the Siamese network, we propose and adapt a pseudo Siamese network model that handles multimodal pair of images with multiple heterogeneous sensors and image resolutions.

Pseudo-Siamese are a variant architecture closely linked to the basic Siamese architecture [16]. It is well adapted to our multimodal CD problem since it is a less restricted network in terms of weights which are not shared between the two network branches. This leads to increase the number of parameters to be adjusted during the training phase, giving a more flexible network than the Siamese network proposed Zagoruyko and Komodakis [16]. Let us also add that recently pseudo-Siamese network has been successfully used in different computer vision tasks such as multimodal patch matching and identification applications in remote sensing imagery, between SAR and optical satellite images [19], or for finding the correspondences between image patches from two highly different modalities, visible and near-infrared patches in natural scene images [20].

The training of the pseudo-Siamese network is accomplished using a pairwise learning approach that involves a loss function depending on pairs of input examples. We formalized the pair-wise learning task as a classification of temporal multimodal image pairs into two categories change/non-change. More precisely, our pseudo-Siamese network based CD model performs both a supervised multimodal dimensionality reduction and a binary classification tasks.

Our multimodal CD model architecture is mainly based on pseudo-Siamese network architecture, having two branches that share exactly the same configuration architecture, but with fewer restrictions on the set of weights. Each branch acts as a feature extractor that takes as input one of the two multimodal patches, which can be also a multichannel patches with respect to the number of bands in the input patches. Let us note that the architecture of our proposed CD model is similar to the network proposed by Hughes et al. [19] that uses a pseudo-Siamese network with uncoupled weights between the two CNN branches in order to identify corresponding patch in SAR and optical images. In this architecture the first and second CNN streams receive respectively one SAR and one optical image modality. Generally, in the heterogeneous CD case, the before and after images can be acquired with two different (multiple) modalities, and possibly one of the two CNN branches can receive a mixed modality with different distributions, i.e. optical and SAR or inversely SAR and optical images, etc. Theses multiple pair modalities can be fed to the network streams, and the model learns more generic features. Instead to learn features from unshared convolutional layers, and in order to reinforce the learning ability of our CD model to learn more robust modality specific features, we propose a partly pseudo-Siamese connected network which ensures that the weights between the layers of the two CNN branches are both shared and unshared depending on the abstraction level (depth). In our architecture, the weights of the first (two) convolutional layers are unshared between the two streams, and the layers has the objective to capture generic features of multimodal patches, i.e. to transform the local patch into more high-level features for each modality. Whereas, the rest of the convolutional layer is coupled (connected) and share the same set of weights. The coupled layer try to learn to generate the latent multimodal correlation features of the corresponding feature patch pairs for the binary classification problem, instead to use features only from the last unshared convolution layer.

Our overall CD framework includes a decision network as a top network that forms a descriptor within a lower dimensional space in which features are combined, hence the loss function learns a decision function from the compact feature space.

The input to the CD model is considered to be a pair of image patches, from which descriptors are first computed independently using two parallel streams and then concatenated with a top network module that decide if the two multimodal input patches present a similar or dissimilar pairs.

In more details, our CNN architecture network is composed of a set of convolutional, ReLU, max-pooling, and fully connected layers. It takes patches as input and apply on them three convolutions and max-pooling, ReLU operations and one concatenation operation of the last layer that is followed by a fully connected layer. Our proposed CNN architecture is inspired by the MatchNet network architecture [21], but with a few layers. The main difference comes from the layer settings. This means that our architecture favors sparse-dense features and disadvantages sparse-sparse features produced by the ReLU. Note also that performing a mean-pooling operation instead of max-polling, does not significantly increase the performance of

the CD model. The structure of the CNN architecture uses small filters of 5 × 5 for all convolutional layers that effectively increases the model performance and reduces the number of filter parameters to be learned. A ReLU function is used after the three convolutional layers, which helps to generate sparse features. The last fully connected layer acts as a linear dimension reduction layer, and project convolutional features in lower dimensions. The ReLU function is removed after this layer to favorize dense representation from activated neurons. The output of the fully connected layer is the feature representation of the input patch. The spatial padding of the convolutional layer input is 2 pixels for the three convolutional layers with 5 × 5 filter size. The convolution stride is set to 1 pixel. Three max-pooling are performed using 3 × 3 spatial pooling kernel with a stride of 2.

In the fusing stage, the two output descriptors of each CNN stream are concatenated using a fusion layer that merges the two input features in one single 128-dimensional feature representation, which is then reduced using 2 fully connected (FC) layers but without ReLU function. The first FC layer contains 16 features and the second has 2 outputs corresponding to the change/non-change binary mapping.

In the proposed approach, the CD model takes a single input which is a pair of patches stacked along the depth dimension that requires being splitted to feed each patch into the corresponding CNN stream. This is ensured using a slice layer that splits the single input into two patches which are in fact the original patches.

As mentioned earlier, the input of the CD model is considered to be a pair of patches, and by definition the Siamese networks use a contrastive loss to learn a new metric to assess the similarity score between these pairs. In the proposed pseudo-Siamese architecture, we adopted the binary cross entropy as a loss function. Figure 1 shows the overall pseudo-Siamese-based change detection framework. Table 1 summarizes the details of our CNN architecture settings.

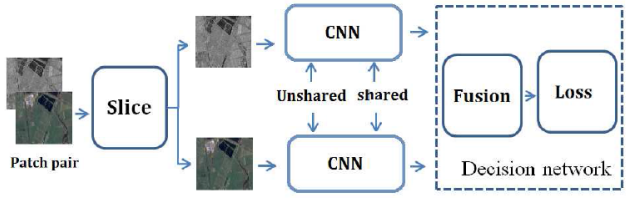

Figure 1: Network architecture of the pseudo-Siamese based change detector model.

Table 1: Details of the model architecture for CNN.

Name	Type	Input size	Filter number	Filter Size conv	Filter Size pool	Stride	Pad	Stride	ReLU
Conv1/Pool1	conv/max pool	32 x 32	32	5 × 5	3 × 3	1	2	2	Yes
Conv2/Pool2	conv/max pool	32 × 16 × 16	32	5 × 5	3 × 3	1	2	2	Yes
Conv3/Pool3	conv/max pool	32 × 8 × 8	64	5 × 5	3 × 3	1	2	2	Yes
FC1	fully-conn	64 × 4 × 4	64	N/A	N/A	N/A	N/A	N/A	No

RESULTS

In order to validate and to show the strength of the proposed model, we conduct the experimentations on five realistic multimodal datasets, reflecting different imaging modalities cases under different change detection conditions with different spatial resolutions, including multi-sensor (heterogeneous optical images) and multi-source (optical and SAR images), showing construction and destruction of buildings in different area. For each multimodal dataset, the ground-truth is provided by a photo-interpreter as a change mask.

In our application, the classification performance of the proposed CD model is assessed using the leave-one-out test procedure. In this well-known evaluation strategy, one entire multimodal dataset is removed from the whole training multimodal images, whereas the training phase is performed on the remaining heterogeneous datasets (Figure 2). The built CD model is then evaluated on the removed dataset to generate binary maps. This process is repeated five times and at each time two multimodal images were retained to form the validation subset.

Heterogeneous dataset description

The first multimodal dataset is a pair of SAR/optical satellite images (Toulouse, France), with a size of 4404 × 2604 pixels, before and after construction. The SAR image was taken by the TerraSAR-X satellite (Feb. 2009) and the optical image by the Pleiades (High-Resolution Optical Imaging Constellation of CNES, Centre National d'Etudes Spatiales) satellite (July 2013). The TSX image was co-registered and re-sampled by Prendes [23] with a pixel resolution of 2 meters to match the optical image.

The second dataset shows two heterogeneous optical images acquired in Toulouse (France) area by different sensor specifications (size 2000 × 2000 pixels with a resolution of 0.5 meter). The before image is acquired by the Pleiades sensor in May 2012 before the beginning of the construction work, and the after image is acquired by WorldView2 satellite from the red, green and blue spectral bands (11 July 2013) after the

construction of a building. The WorldView2 VHR-image was co-registered to match the Pleiades image.

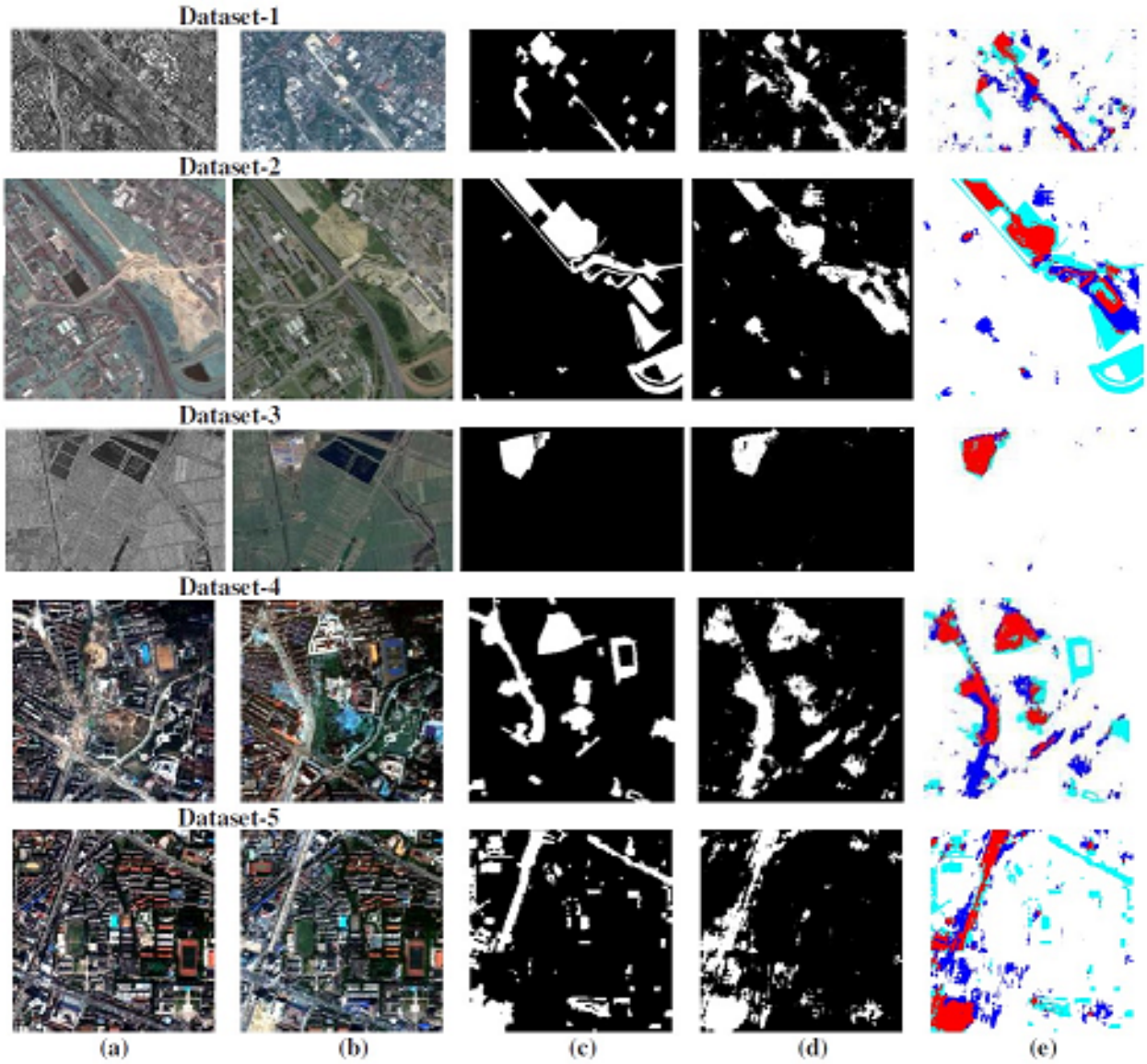

Figure 2: Heterogeneous dataset: (a-c) image t1, t2, ground truth; (d-e) final (changed/unchanged) binary classification and confusion map (white: TN, red: TP, blue: FP, Cyan: FN) obtained by the proposed approach.

The third multimodal data set consists of one SAR image and one RGB optical image. It shows a piece of the Dongying City in China, before and after a new building construction. The SAR image is acquired by RADARSAT-2 (June 2008) with a spatial resolution of 8 meters. The optical image comes from Google Earth image (Sept. 2012) with a spatial resolution of 4 meters. After co-registration, they are of the same pixel-resolution to give a size of 921 × 593 pixels.

The fourth dataset shows two heterogeneous optical images from another area in the south campus of Hubei province of China, were respectively acquired by the Quick Bird satellite in May 2002 and the IKONOS satellite in July 2009, with a size of 240 × 240 pixels. The images after preprocessing have the same spatial resolution of 3.28 meters.

The fifth dataset shows two heterogeneous optical images covering the campus of Wuhan University in Hubei province of China. They were respectively acquired by the QuickBird satellite in April 2005 and the IKONOS satellite in July 2009, and correspond to 4-bands (red, green, blue, and NIR band) with a size of 400 × 400 pixels. The resolution of these images is

of 2.44 and 3.28 meters. After re-sampling the after image have the same spatial resolution as the before image 2.44 meters.

Table 2: Accuracy rate of change detection on the five heterogeneous datasets obtained by the proposed method and the State-of-the-art multimodal change detectors (first upper part of each table) and monomodal change detectors (second lower Part of each table).

SAR/Optical Dataset [#1]	Accuracy	Optical/Optical Dataset [#2]	Accuracy	SAR/Optical Dataset [#3]	Accuracy
Proposed method	0.870	Proposed method	0.865	Proposed method	0.987
Prendes et al. [22]	0.844	Prendes et al. [22], [23]	0.844	Liu et al. [9]	0.976
Correlation [22]	0.670	Correlation [22], [23]	0.679	PCC [9]	0.821
Mutual Inf. [22]	0.580	Mutual Inf. [22], [23]	0.759		
		Pixel Dif. [23]	0.708		
		Pixel Ratio [23]	0.661		

Quickbird/IKONOS Dataset [#4]	Accuracy	Quickbird/IkONOS Dataset [#5]	Accuracy
Proposed method	0.877	Proposed method	0.837
Tang et al. [24]	0.986	Tang et al. [24]	0.959
Multiscale [24]	0.991	Multiscale [24]	0.966

Table 3: Confusion matrix for each of the five multimodal datasets i.e., [TSX/PLEIADES] (4404×2604 PIXELS), [PLEIADES/WORLDVIEW-2] (2000×2000 PIXELS), [RADARSAT-2/QUICKBIRD] (921×593 PIXELS), [QB02 /IKONOS] (240×240 PIXELS), [QB02 /IKONOS] (400×400 PIXELS).

Multimodal image pairs	TP	TN	FP	FP	FN
TSX/Pleiades	50%	90%	10%	10%	50%
Pleiades/WorldView-2	47%	94%	6%	6%	53%
RADARSAT-2/Quickbird	81%	99%	1%	1%	19%
Quickbird/IKONOS	52%	94%	6%	6%	48%
QuickBird /IKONOS	49%	90%	10%	10%	51%

Training details

In this work, we first convert the multi-bands color image to a grayscale image, and for each multimodal dataset we apply a pre-processing step to extract patches of size 32 × 32 pixels in order to form the training samples. In our application, we also expand the training samples in order to improve the generalization ability of our CD model. For each image patch, we perform standard data augmentation techniques three times by applying some affine transformations between the same multimodal patches. This is simply achieved by performing rotation, translation, and scale image processing techniques. The CD model was trained using the scaled conjugate gradient descent algorithm, with a fixed learning rate of 0.001 and without dropout layer. The momentum and the weight decay were set to 0.9 and 0.004 respectively. The number of epochs was set to 150 epochs. The Training was conducted on GPU clusters with batches of 64 pairs of 32 × 32 patches using balanced classes with leave-one-out evaluation strategy, i.e. the training take around five rounds. Each time a completely different datasets is used for evaluation.

Evaluation results

The evaluation of our CD model is assessed using the accuracy classification rate, which quantifies the percentage of the correct changed and unchanged pixels (Equation 1), in order to compare the obtained results to the state-of-the-art methods:

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

Where TP and TN represent the number of pixels that are correctly classified, FN and FP represent the number of misclassified pixels. In our application, the global accuracy rate represents the average accuracy rate obtained from the five different change detection accuracy rates obtained from the five heterogeneous datasets with the leave-one-out evaluation scenario.

In the first experiment, we compare the performance of our CD model with unsupervised image processing CD models. To that end, we summarize respectively in Tables 2 and 3 the accuracy rate and the confusion matrix obtained using the leave-one-out evaluation strategy. As depicted in Tables 2 and 3 the proposed CD model outperforms some of the state-of-the-art methods, it is able to process new probe image pairs under different change detection conditions and without over fitting any of the two classes.

In a second experiment, we compared the performance of our CD model with the supervised pseudo-Siamese, Siamese, and the unsupervised Siamese network models described by Hughes et al. [19] and Rahman et al. [25]. Let us recall that the architecture of our CD model was adapted from the pseudo-Siamese network model of Hughes, et al. [19], where the main difference come from the number of parameters and degree of freedom to map the two modalities onto the new subspace. More precisely, our CD model has fewer parameters and degree of freedom to be tuned during the training phase, i.e architecture with partly unshared-shared weights between the two parallel CNN streams, in which our model uses features from the last shared convolution layer to train the binary decision network, contrary to the network architecture proposed by Hughes et al. [19] Which is unshared parallel network streams, uses features solely from unshared convolution layer to learn the decision network? Let us also add that the supervised and unsupervised Siamese network models [25] are both a fully shared parallel network streams which uses selected features from multiple levels, but with different network decision. Let us also recall that the network CD models were validated under the leave-one-out evaluation strategy in our application. The global average classification rate was 86.8% with the pseudo-Siamese network model of Hughes LH, et al. [19], and 85.6% and 81.7% when we experienced the supervised and unsupervised Siamese network CD models of Rahman et al. [25], for detecting similar and dissimilar heterogeneous remote sensing image patch pairs (change vs. non-change) reflecting different multimodal satellite image sensors and different spatial image resolutions. Based on this comparison, we can draw some observations about the behavior of our CD model, in contrast to the pseudo-Siamese network model [19], and the supervised and unsupervised Siamese networks [25]. First, we can observe that the proposed supervised CD model produces higher average classification rate of 88.7% compared to the obtained rates with the pseudo-Siamese model [19] and the supervised and unsupervised Siamese networks [25]. Second, in all the experiments, we have observed that the CD results obtained by the unsupervised Siamese network [25] depend on the selection of the decision threshold, which varies according to the heterogeneous image pair (CD condition request) to be classified.

Table 4: Average change detection accuracy on the five heterogeneous datasets obtained by the proposed method and the State-of-the-art Siamese cd network models.

CD network model	Average accuracy
Proposed method	0.887
Supervised pseudo-Siamese [19]	0.868
Supervised Siamese [25]	0.856
Unsupervised Siamese [25]	0.817

Table 4 shows the average classification rate obtained with the pseudo-Siamese network model [19], supervised and unsupervised Siamese network models [25], under the leave-one-out validation strategy.

DISCUSSION

The multimodal CD described in this paper turns out to be interesting for multi-resolution change detection. Indeed, the CD model is learning the modality specific features. Globally, the model learn to fuse features of the two multimodal patches which helps to factorize the differences (e.g. land cover changes) and the imaging modalities, but also makes use of standard max-pooling layers to deal with the multi-resolution nature of the data. The model can be also less accurate than some specific CD models that are more specific and only dedicated to a restricted number of imaging modalities.

CONCLUSION

In this paper, we presented a parallel framework based on partly uncoupled learning architecture for change detection from bi-temporal multimodal remote sensing images. The model that combines a pseudo-Siamese CNN feature descriptor, a fusion layer and a cost classification module, is able to properly capture the spatial and the temporal dependencies between the multimodal input image pairs thanks to its ability to process input data pairs in parallel. The experiments using the leave-one-out test strategy demonstrate that the proposed CD model presents an effective way to process new-unseen heterogeneous input image pairs with different spatial resolutions and under different heterogeneous CD conditions such as multi-source and multi-sensor image pairs.

ACKNOWLEDGEMENT

We would like to acknowledge all researchers that made at our disposal the change detection dataset in order to validate the proposed change detection model.

REFERENCES

1. Touati R, Mignotte M. An energy-based model encoding nonlocal pairwise pixel interactions for multisensor change detection. IEEE Transactions on Geoscience and Remote Sensing. 2017;56(2): 1046-58.
2. Prendes J, Chabert M, Pascal F, Giros A, Tourneret JY. A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors. IEEE Transactions on Image Processing. 2014;24(3):799-812.
3. Alberga V. Similarity measures of remotely sensed multi-sensor images for change detection applications. Remote Sensing. 2009;1(3):122-43.
4. Touati R, Mignotte M, Dahmane M. A new change detector in heterogeneous remote sensing imagery. In: Seventh International Conference on Image Processing Theory, Tools and Applications. 2017;1-6.
5. Brunner D, Lemoine G, Bruzzone L. Earthquake damage assessment of buildings using VHR optical and SAR imagery. IEEE Transactions on Geoscience and Remote Sensing. 2010;48(5):2403-20.
6. Touati R, Mignotte M, Dahmane M. Change detection in heterogeneous remote sensing images based on an imaging modality-invariant mds representation. In: 25th IEEE International Conference on Image Processing. 2018;3998-4002.

7. Liu ZG, Mercier G, Dezert J, Pan Q. Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning. IEEE Geoscience and Remote Sensing Letters. 2013;11(1):168-72.
8. Liu Z, Li G, Mercier G, He Y, Pan Q. Change detection in heterogenous remote sensing images via homogeneous pixel transformation. IEEE Transactions on Image Processing. 2017;27(4):1822-34.
9. Liu J, Gong M, Qin K, Zhang P. A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. IEEE transactions on neural networks and learning systems. 2016;29(3):545-59.
10. Zhang P, Gong M, Su L, Liu J, Li Z. Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing. 2016;116:24-41.
11. Zhao W, Wang Z, Gong M, Liu J. Discriminative feature learning for unsupervised change detection in heterogeneous images based on a coupled neural network. IEEE Transactions on Geoscience and Remote Sensing. 2017;55(12):7066-80.
12. Mercier G, Moser G, Serpico SB. Conditional copulas for change detection in heterogeneous remote sensing images. IEEE Transactions on Geoscience and Remote Sensing. 2008;46(5): 1428-41.
13. Chatelain F, Tourneret JY, Inglada J. Change detection in multisensor SAR images using bivariate gamma distributions. IEEE Transactions on image processing. 2008;17(3):249-58.
14. Prendes J, Chabert M, Pascal F, Giros A, Tourneret JY. Change detection for optical and radar images using a Bayesian nonparametric model coupled with a Markov random field. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing. 2015;1513-1517.
15. Merkle N, Fischer P, Auer S, Müller R. On the possibility of conditional adversarial networks for multi-sensor image matching. IEEE International Geoscience and Remote Sensing Symposium. 2017;2633-2636.
16. Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition 2015;4353-4361.
17. Hedjam R, Abdesselam A, Melgani F. Change Detection from Unlabeled Remote Sensing Images Using SIAMESE ANN. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. 2019;1530-1533.
18. Merkle N, Luo W, Auer S, Müller R, Urtasun R. Exploiting deep matching and SAR data for the geo-localization accuracy improvement of optical satellite images. Remote Sensing. 2017;9(6):586.
19. Hughes LH, Schmitt M, Mou L, Wang Y, Zhu XX. Identifying corresponding patches in SAR and optical images with a pseudo-Siamese CNN. IEEE Geoscience and Remote Sensing Letters. 2018;15(5):784-788.
20. En S, Lechervy A, Jurie F. TS-Net: Combining modality specific and common features for multimodal patch matching. In: 25th IEEE International Conference on Image Processing. 2018;3024-3028.
21. Han X, Leung T, Jia Y, Sukthankar R, Berg AC. Matchnet: Unifying feature and metric learning for patch-based matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015;3279-3286.
22. Prendes J, Chabert M, Pascal F, Giros A, Tourneret JY. Performance assessment of a recent change detection method for homogeneous and heterogeneous images. Revue Française de Photogrammétrie et de Télédétection. 2015;209:23-9.
23. New statistical modeling of multi-sensor images with application to change detection. In: J. Prendes. Ph.D. dissertation, Toulouse, 2015.
24. Tang Y, Zhang L. Urban change analysis with multi-sensor multispectral imagery. Remote Sensing. 2017;9(3):252.
25. Rahman F, Vasu B, Van Cor J, Kerekes J, Savakis A. Siamese Network with Multi-Level Features for Patch-based Change Detection in Satellite Imagery. In: IEEE Global Conference on Signal and Information Processing. 2018;958-962.