Literature Review on Change Detection Using Remote Sensing Imagery

Dilraj Bhinder[1], Dr. Devaraj Verma C[2]

[1]M. Tech Student, FET, FET, Jain (Deemed to be University)
dilraj1983 [at]gmail.com

[2]Professor, Department of CSE, FET, Jain (Deemed to be University)
devarajverma04 [at]gmail.com

Abstract: ***This paper reviews the subject of Change Detection (CD) using Remote Sensing (RS) or Satellite Imagery. The definition of the problem of CD using RS imagery is explained first followed by the broad framework under which various CD techniques work. This is followed by the survey of different types of CD techniques starting from traditional techniques using only statistics, Machine Learning (ML) techniques followed by the contemporary Deep Learning Techniques with related issues and challenges of all these techniques. Few recent research papers which use DL techniques have also been described in detail with the architectures used at the end of the review. This review will serve as a ready reckoner for researchers wishing to pursue the area of CD using Satellite Imagery.***

Keywords: Change Detection, Remote Sensing, Deep Learning, Neural Networks, Siamese UNet

1. Introduction

Change detection is an integral part of the analysis of satellite imagery, and it has been studied for several decades. It consists of comparing a registered pair of images of the same region and identifying the parts where a change has occurred, e.g. vegetation evolution or urban changes. A label is assigned to each pixel: change or no change. The nature of the changes that are detected may vary with the desired application, such as vegetation changes or urban changes (artificialization). Change detection is a crucial step for analysing temporal Earth observation sequences in order to build evolution maps of land cover, urban expansion, deforestation, etc. The first techniques that were proposed used manually crafted processes to identify changes, while later methods have proposed using machine learning algorithms in different ways [1, 2, 3, 4]. Recent advances in machine learning algorithms for image analysis have not yet taken over the area of change detection due to the lack of large amounts of training data. Thus, some methods have been proposed recently to use transfer learning to circumvent this problem [5]. While transfer learning is a valid option, it may limit the reach of the proposed methods. For example, the vast majority of the large CNNs trained on big datasets use RGB images, while the Sentinel-2 images contain 13 useful bands, most of which would need to be ignored when using such a system. Moreover, the recently proposed deep learning algorithms for change detection have mostly been designed to generate a difference image which is manually thresholded [6]. This avoids end-to-end training, which tends to achieve better results and faster execution. Related to change detection, deep learning techniques have been developed for computer vision applications with the aim of comparing image pairs [7].

<u>Change Detection Framework</u>

Generally speaking, the technological process of change detection is as follows. In order to ensure the consistency of the coordinate system, raw Remote Sensing (RS) images are pre-processed by image registration. According to the description and attributes of the registered RS images, the multitemporal information, namely basic features, is obtained through the relevant image feature extraction algorithms. It contains color, texture, shape, and spatial relationship features of the images. Afterward, the changing features, refined or differentiated from the multi-temporal basic features, are conducted to reveal the location and intensity of extracted change information. The above two steps can be collectively referred to as change information extraction. Finally, feature integration and information synthesis process are conducted to combine global features with the changing judgment criteria, obtaining the final change results, as shown in the blue blocks in Figure 1.

Image registration (i.e., co-registration), which aligns multi-temporal images of the same scene, is essential for RS change detection tasks. As the most common registration strategy for RS images, geographic registration directly maps multi-temporal images via automatic matching of control key points according to geographic coordinates attached to digital raster RS images. However, data requiring registration is available to be captured from different viewpoints, light environments, sensors, and even distinct data models (e.g., digital elevation model). Therefore, interference is often brought to subsequent change detection (e.g., error change boundary caused by registration). Faced with the multi-source and multi-objective scenarios in urban change detection, it is hard to determine which RS image data and which method are more advantageous to balance the effectiveness and the accuracy of change detection results.

<u>Broad Classification of types of Change Detection Approaches</u>

In the literature, change detection algorithms can be summarized into two categories. First, they may follow post classification comparison [8], which classifies the input images separately into many predefined classes and then obtain the change regions by comparing the labels pixel by pixel. Second, the other kind of methods is called post comparison analysis [9]-[14]. These methods follow following procedure: first deriving a similarity-feature map [e.g., a difference image (DI)] from the multitemporal images and then analyzing the feature map to segment changed and unchanged areas. Image arithmetical operations (e.g., differencing [9] and ratioing [10]) and image transformation [14] can be used to obtain the feature map. The algorithms to analyze the feature map include thresholding, clustering [9], and other advanced methods, such as an extreme learning machine (ELM) [11] and an Markov random field [12], [13]. In [11], ELM is applied as the classification model for

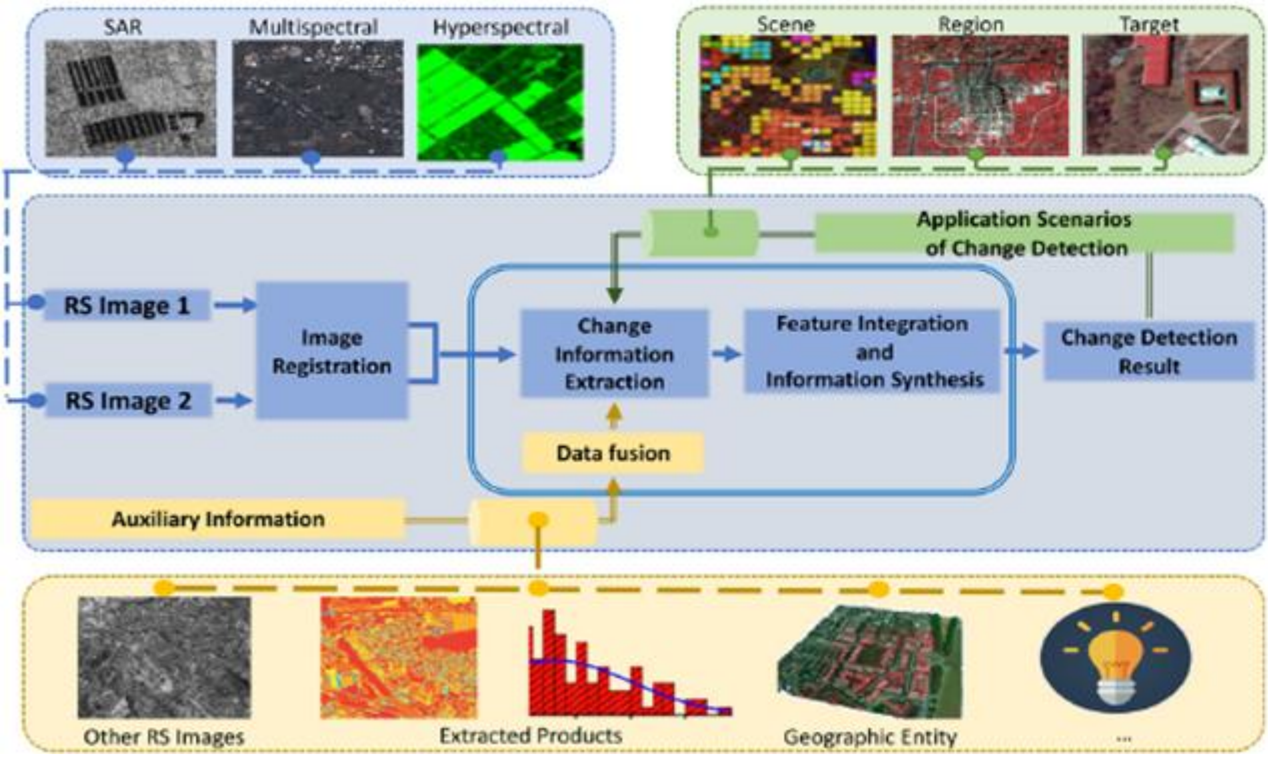


Figure 1: General Framework of Urban Change Detection

change detection, which is trained using some labeled samples obtained by a pre-classification schema. Benedek and Szirányi [12] proposed a conditional multilayer mixed MRF (CXM) model for change detection, which is a three-layered MRF using two weak features: global intensity co-occurrence statistics and block correlation. In [13], a multilayer MRF model (L3MRF) is proposed. The L3MRFmodel uses the similar structure to the CXM [12] but different features: the modified histogram of oriented gradients and gray-level difference features. As described earlier, the features used by conventional change detection algorithms are almost hand-crafted, which are weak in image representation.

Recently, deep neural networks (DNNs) are learned to extract features directly from the input images. The features are more abstract and robust. In literature, some change detection methods [10], [14], [15] using DNN have been proposed and demonstrate good performance. In [3], a DNN based on the restricted Boltzmann machine is learned to classify the DI for synthetic aperture radar (SAR) images. Gao et al. [15] design a pre-classification schema to obtain some labelled samples of high accuracy and then use these samples to train PCANet as the classification model for change detection. In [14], a symmetric convolutional coupling network (SCCN) is applied to detect changes between optical and SAR images. The method extracts features first and then transforms the features into a consistent feature space, where a different map can be obtained. The network is learned by optimizing a coupling function, which is only defined on the point of view of the unchanged pixels. In [14], the method demonstrated superiority over several existing approaches, but the limitation is that the method only considers the unchanged pixels.

The methods described earlier can also be distinguished tobe unsupervised [9]-[11], [14], [15] or supervised [8], [12], [13]. Unsupervised methods do not use the ground truth (GT) data. Thus, they usually rely on some prior assumptions, such as that unchanged regions should demonstrate asmaller pixelwise difference [14]. However, the feature statistics in optical images may be multimodal and strongly overlapping [12]; therefore, it may be challenging for the unsupervised methods in some situation. On the other hand, if training data with GT data are available, they can provide significant additional information for classification.

Methods of Mathematical Analysis

Algebraic Analysis. In early engineering applications, the algebraic difference or ratio between pixel values in the adjacent phases is applied to measure the changes of the grayscale images. In the theoretical analysis, the first attempt can be found in change vector analysis (CVA), which converts the difference of pixel values into the difference of feature vectors. The intensity and direction of change vector (CV) provide reliable facts about the type and status of the change. It has been proved that Markov distance and Manhattan distance are equipped to measure the amplitude of high-dimensional CVs. However, it must be noted that the similar CVs extracted from the pixel-wise algebraic calculation should be clustered in the last step; hence the artificial setting of thresholds is an unavoidable problem in the algebraic method. Confronting the adverse effects, Sun [16] proposed to adjust the weight parameters of CVA according to spectrum standard deviations and the variation amplitude of the features in the adaptive region.

However, such methods are computationally intensive to cope with the high-resolution images and in addition; the high-dimensional CVs will be generated in the multispectral data, which restricts the effectiveness and popularization of the relevant methods.

Statistical Analysis Likewise, as the mathematical analysis method, the statistical analysis shifts the focus from pixel to region. Depending on the order in which statistical analysis is performed, such methods can be divided into two categories, namely direct calculation and indirect calculation.

Direct calculation: The direct calculation methods make a difference on the individual statistics results of the original multi-temporal images. Without a doubt, even though the independent images are calculated, the correlation between multi-temporal images is still necessary for the direct calculation method. For example, iteratively regularized multivariate change detection (IR-MAD) transformation is of capacity to measure spatial correlation, namely achieving a new invariant mapping on multi-temporal images in unsupervised fashion, and then make individual statistics on this basis. In addition, there are other statistical parameters available to measure the spatial correlation of multi-temporal data, e.g., Moran's index, likelihood ratio parameters, and even trend of spectral histogram. The multi-scale object histogram distance (MOHD) is created to measure the "bin-to-bin" change, contrasting the mean values of the red, green, and blue bands of the pair wise frequency distribution histograms. In order to achieve targeted statistics of the concerned objects, the covariance matrix of MAD, calculated through weighted functions, i.e., the normalized difference water index (ND-WI) for the water body, and the normalized difference built-up index (ND-BI) for urban building, plays an important role.

Indirect calculation: There are two situations feasible for indirect calculation. One is carrying change statistics on the refined features. In fact, some statistical functions are difficult to be applied in the original data domain of the extracted features, taking the probability density function (PDF) as an example. However, in the dual-tree complex wavelet transform (DTCWT) domain, PDF is effective for probability statistics of image features. In addition, with the purpose to optimize change results, performing statistics on the raw difference results of multi-temporal images also plays an important role.

Methods for Feature Space Transformation Ignoring the potential relationship hidden in the training samples, namely mainly paying attention to the mathematical representation of pixels value, makes the mathematical analysis method lack generalization ability. Contrarily, it should be noted that the high computational overhead caused by mining potential relationships between data must be considered. Therefore, at the demand of optimizing data redundancy and reducing the feature dimensions, the feature space transformation methods have been developing, which can derive into three schools for various application purposes.

Naive dimensionality reduction: The method aims at reducing redundancy and improving the recognizability of change by converting the original images into analyzable feature space. As the basic dimensionality reduction operations, principal component analysis (PCA), and mapping of variable base vectors in sparse coding are suitable for urban change detection. It has been proved that the specific filtering operation and wave frequency transformation highlight the high-frequency information and weaken the low-frequency information. For example, Gabor linear filtering, quaternion Fourier transform (QFT), and Hilbert transform are supplementary means for localized analysis of time or spatial frequency. Relatively, the wave frequency transformation method is more flexible. At present, it is advisable to conduct conversion of the high-low frequency on the multi-source multi-temporal images, and then make difference on the results of wavelet inverse transform, or directly obtain DI with wavelet frequency difference (WFD) method.

Noise suppression: Noise interference is an unavoidable problem in image detection, especially for SAR images. Singular value decomposition (SVD) and sparse model can map high-dimensional data space to low-dimensional data space, meanwhile undertaking auxiliary denoising. For example, the adapted sparse constrained clustering (ASCC) method integrates the sparse features into the clustering process, utilizing the coding representation of only meaningful pixels. Or based on relationships between whole and part, processing filtering operation on the boundary pixels to confirm properties of centre pixels is also desirable for noise suppression.

The main highlight of feature space transformation is that it reduces redundant information. As a consequence, it is available to cope with large image data with high data redundancy when the computing resources and time are sufficient. However, once a wrong judgement happened at extracted features, real change information is likely to be eliminated. In addition, setting parameters artificially are non negligible restrictive conditions for feature space transformation.

Feature Classification Methods

There are two methods for feature classification. One is to classify the images of every phase based on the category of the ground objects, and then carry out a comparison on the classification results, that are post-classification comparison (PCC) methods. Although it is friendly to urban change detection tasks with predictable types of objects and multi-classification tasks, an important fact is that the PCC methods excessively depend on the pre-order classification accuracy of every single phase, concluding to accumulated error. Contrarily, other groups have disputed the PCC and put forward to directly classify whether the information is changed or not.

Despite the undisputed success of kernel SVM in feature classification, it is not sensitive to outliers, e.g., the inherent pattern differences existed in the multi-temporal images, such as season changes and solar angle changes. In addition, it is also bedevilled by fitting kernel parameters and operating efficiency to large datasets.

Decision Tree (DT) obtains conditional probability distribution in feature and class space through the tree structure. It has been proved the ability and efficiency in separating nonlinear features [88]. However, DT is prone to over fitting. In order to adapt to the complex RS data, the ensemble learning algorithm and random subspace theory are combined with DT, bagging the independent weak DT classifiers into a strong one, that is, the random forest (RF). It has been proved that its higher randomization and better variance contribute to stable change detection results. Other than determining whether the change happened, it also can be used to determine the authenticity of the change. For the ubiquitous pseudo changes, the weighted RF can act auxiliary tool for other change detection methods by judging the difference between the generated change and the pseudo change. Moreover, different from RF which predicts in parallel, there is a trend to integrate boosting strategy into DT framework, that is, iterating classification results through the cascade of classifiers. For instance, adaptive boosting (AdaBoost), which emphasizes the fitting samples with predicted errors and conducts iterative training on weak classifier according to the updated sample weights; gradient boosting decision tree (GBDT), which applies residual learning and gradient calculation based on AdaBoost; multi-grained cascade forest, which introduces sliding window and cascade structure, takes full advantage of the spatial expression of images.

Deep Neural Networks Neural network (NN) is a computing model that mimics the structure of biological neuron. The multi-layer hidden layers endow the NN with deep characteristic representation, namely deep neural network (DNN). As an efficient and robust feature extraction method for big data, the DNN eliminates the tedious catastrophe of manually selecting features and the dimensional disaster of high-dimensional data. At present, three theories have been postulated to explain the development of DNN in the change detection task. One is to achieve different network organization by ameliorating the stacking of neurons, which is summarized as the naïve DNN. In addition, apply neuronal structures related to spatiotemporal features, such as convolution cells and recurrent cells. Or, consider the collaborative work of multiple branches NN to generate changing features.

Naïve DNN: As the basic form of DNN, multi-layer perceptron (MLP) reconstructs the change feature space through neurons of hidden layers, and realizes the category expression of basic change information. On this basis, pursuing the nonlinear separability of change in RS images, the nonlinear neurons are utilized in radial basis function (RBF). For acquiring the ability to solve complex problems, the restricted Boltzmann machine (RBM), as a two-layer structure with the visible layer and the hidden layer, reduces the dimension of complex RS data through the neuron association between layers. Deep belief networks (DBN) [17] are stacked by multiply RBMs, its independent hidden layer units are training separately with the joint probability distribution of the data. Experiments show that DBN automatically acquires the abstract change information hidden in the data, which is difficult to interpret, and improves the assimilation effect of the unchanging areas, meanwhile highlighting the change.

DNN for Spatio Temporal Features: In order to cover the shortcomings of naïve DNN, which ignores the two-dimensional spatial information and time-related information of multi-temporal RS images, the application of deep convolutional neural network (DCNN) and sequential neural network are discussed below.

Deep convolutional neural network: Due to the application of the convolution kernel, DCNN achieves the most advanced results on numerous tasks of computer vision and image processing, including multi-temporal RS images. In view of the particularity of the change detection task, three concepts are focused, namely input mode of multi-temporal images, model optimization strategy, and detection solution. Loss Functions: At present, binary cross-entropy and structural similarity index measures (SSIM) [18]are mainstream loss functions to highlight differences and evaluate the similarity of multi-temporal features. As an improvement, for equalizing the proportional relation of unbalanced samples, despite the weighted cross-entropy loss, the random selectivity of training samples is also pivotal. Although CNN is a common supervising strategy, combining CNN with unsupervised theory, such as Kullback-Leibler (KL) divergence loss, also produces good results. In addition, the automation of training and the role of iterative training are advocated. Based on the back propagation and feature differential, the change features are selected from the generated tensors automatically, and the results are iterated by repeatedly comparing with the annotation. Experiences show that it makes the unchanged regions as similar as possible, meanwhile the changed regions as different as possible.

Deep Sequential Neural Nets: At the preliminary stage, as a feature transformation scheme, slow feature analysis (SFA) [19] employs time signals to learn the linear factors of invariant features and extracts the slowly changing features. Depending on the structure of NN, deep sequential neural network takes repeatedly connected neurons with memory function as processing medium, associating the abstract category information of land coverage with the temporal feature space. In fact, REFEREE (learning a transferable change Rule From a recurrent neural network for change detection) [20] is the first attempt to learn “change rule” with RNN. Based on REFEREE, cyclical interference is purposed to be suppressed by a periodic threshold [21]. Nevertheless, to recognize the limitations in only handing pixel-vectors, Wu and S. Prasad [22] proposed to deduce the adjacent property from the convolved window of CNN. It found the first combination of RNN and CNN into an end-to-end network structure. Desired to solve the exponential explosion of RNN, the long short-term memory (LSTM) network becomes a substitute. LSTM collects the short-term memory captured from the recent time steps and preserves the perennial long-term memory. It is available to integrate with CNN [23] into convolution LSTM (ConvLSTM), even with the continuous bag-of-words (CBOW)model [24]. However, it should be figured out that a better detection result is shown in the Faster RCNN-based OD method [25].

Description of some Older approaches proposed using Machine Learning Algorithms

Gang Liu, Julie Delon, Yann Gousseau, and Florence Tupin, "Unsupervised change detection between multisensory high resolution satellite images," in Signal Processing Conference (EUSIPCO), 2016 24th European. IEEE, 2016, pp. 2435-2439. This approach does not assume any pre-registration between images and also yields satisfactory results when heterogeneous sensors are used. Because the method relies on geometric descriptors and explicitly models local deformations, it is especially adapted to high resolution urban scenes. This operational method does not explicitly rely on building or surface modeling. In short, the proposed method works as follows. Starting with a pair of unregistered images, key points and their corresponding SIFT-like descriptors are first extracted and matched between images. From these matches, a single global transform is inferred between images, as well as several locally evaluated transforms. For each keypoint in one of the image, the computed transforms are tested and the best location in the other image is retained. This location is called the mapping of the key point. Due to the local transforms, this step allows to deal with different incidence angles and disparate heights. Change detection itself is carried out in the next step. Courtesy the a contrario approach [19], keypoints that are significantly different from their mapping are identified as changed. Then, these changed keypoints are grouped, again using an a contrario methodology, resulting in the final detected regions. The method however, will fail in low contrast scenes where less keypoints are detected and the method involves manual setting of various parameters in contrast to end to end learning.

A. Mohammed EI Amin, Q. Liu, and Y. Wang, "Convolutional neural network features based change detection in satellite images, " In this approach it was the first time that a pre-trained CNN was used to extract features for change detection. The methodology involves extraction of features by a pre-trained CNN i.e. AlexNet. The features extracted at different levels are fused because the lower level features carry more spatial information and the higher level features carry better semantic information. After extraction of features from the bitemporal images the pixel wise change is detected based on the Euclidean distance between the extracted features. The results proved that CNN based features were better at detecting changes compared to Image Subtraction, MRF, PCA, EM based approaches.

Arabi Mohammed El Amin, Qingjie Liu, and Yunhong Wang, "Zoom out CNNs features for optical remote sensing change detection," in Image, Vision and Computing (ICIVC), 2017 2nd International Conference on. IEEE, 2017, pp. 812-817. In this paper, a novel approach based on unsupervised optical remote sensing change detection (CD) based on pretrained convolutional neural network (CNN) on ImageNet dataset and superpixel segmentation technique was proposed. The proposed approach can be divided into three steps. First, bitemporal images are stacked, and Principal Component Analysis (PCA) is applied to extract three higher uncorrelated channels, which are later segmented into superpixels. Second, each region is zoomed out into three levels and then fitted separately into a pre-trained CNN. Third, features of different zooming levels that represent the same region (superpixel) are extracted and concatenated. Then the concatenated features are compared to get the final change map. The experimental results demonstrate the efficacy of the proposed approach in comparison to traditional approaches like EM, PCA, MRF and Image Differencing.

T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k-means clustering," IEEE Geosci. Remote Sens. Lett., vol. 6, no. 4, pp. 772-776, Oct. 2009. The author proposed a novel technique for unsupervised change detection in multitemporal satellite images using principal component analysis (PCA) and k-means clustering. The difference image is partitioned into $h \times h$ non overlapping blocks. S, $S \leq h2$, orthonormal eigenvectors are extracted through PCA of $h \times h$ nono verlapping block set to create an eigenvector space. Each pixel in the difference image is represented with an S-dimensional feature vector which is the projection of $h \times h$ difference image data onto the generated eigenvector space. The change detection is achieved by partitioning the feature vector space into two clusters using k-means clustering with $k = 2$ and then assigning each pixel to the one of the two clusters by using the minimum Euclidean distance between the pixel's feature vector and mean feature vector of clusters. It produced results comparable, even better, with the MRF-based approach [26], which requires computationally expensive data modeling and parameter estimation. Simulation results show that the proposed algorithm performs quite well on combating both the zero-mean Gaussian noise and the speckle noise, which is quite attractive for change detection in optical and SAR images.

Contemporary Techniques for Change Detection Using Siamese Convolutional Neural Networks

In the Siamese architecture two bitemporal inputs are provided simultaneously to two CNNs with the same architecture and whose weights are shared. Each branch takes one of the two patches as input. The two branches of the CNNs with their weights shared act as feature extractors. The two branches of the Siamese network share the same weight, which means that they extract features from the two patches using the same approach. As the two input images used for change detection are both optical images, in other word homogeneous, which are captured by the same type of sensor (both optical sensors) and have similar characteristics, it is natural to extract features in the same way.

In the Siamese Network, each branch follows CNN's architectures while there are also some differences from the conventional CNN. In the conventional CNN, there exist three types of layers: convolutional, pooling, and fully connected layers. The convolutional layers can extract the hierarchical features from the input image. The pooling layers' functionalities consist of receptive field enlargement and dimensionality reduction, which means to reduce the size of the output feature maps. The fully connected layers are used as a classifier, which outputs the probabilities predicting the input image to each class. Because our goal is to extract features pixel by pixel, the pooling and fully connected layers are not used in designed network.

Another important issue in the network design is the selection of kernel size of each convolutional layer. The size of kernels keeps gradually increasing with the convolutional layers except the last one. This strategy, which ensures the enlargement of the receptive field, can serve as functional substitutes for the pooling layer. The kernel size of the last convolutional layer, which acts as a selector of feature maps, is 1. In this letter, we select the above-mentioned parameters experimentally.

Yang Zhan, Kun Fu, Menglong Yan, Xian Sun, Hongqi Wang, and Xiaosong Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 10, pp. 1845-1849, 2017. In this paper, a novel model for change detection in optical aerial images was proposed, which is based on the supervised deep Siamese convolutional neural network (CNN) [27]. The main contributions are summarized as follows:

a) A Siamese CNN is learned to extract features directly from the images pixel by pixel. The extracted features, which are suitable for change detection, demonstrate a unique property: the feature vectors associated with changed pixel pairs are far away from each other in the feature space, whereas the ones of unchanged pixel pairs are close.
b) In order to distinguish the changed and unchanged pixels more effectively and reduce the influence of imbalance data (i.e., the numbers of changed and unchanged pixels vary greatly) in change detection, therefore, a weighted contrastive loss [28] is used, in which not only the unchanged pixels but also the changed ones are considered as the objective function when training the network. The proposed model is shown in Figure 2.

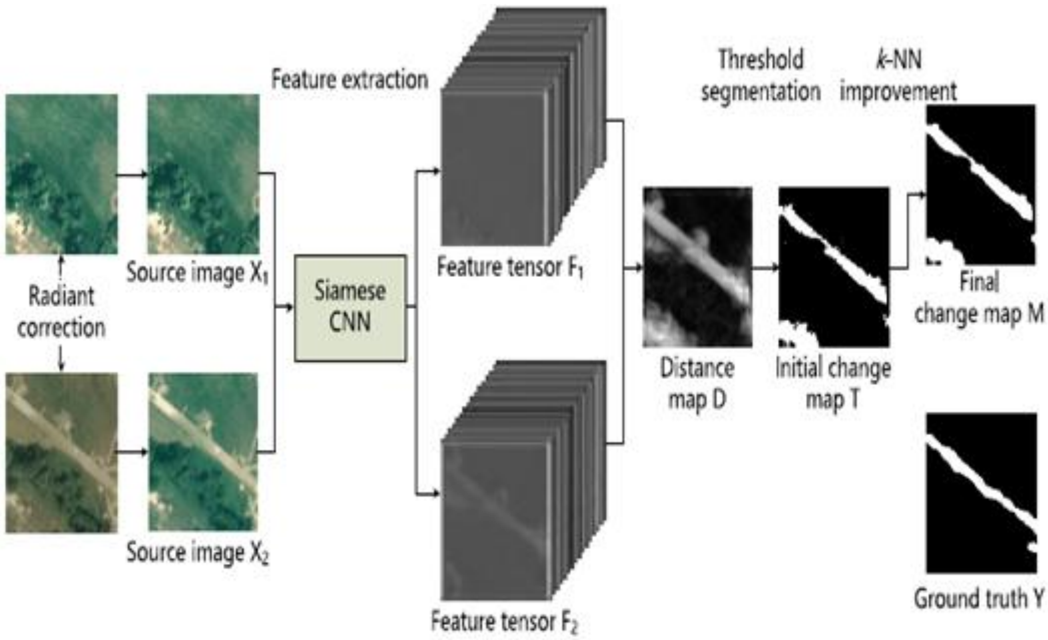


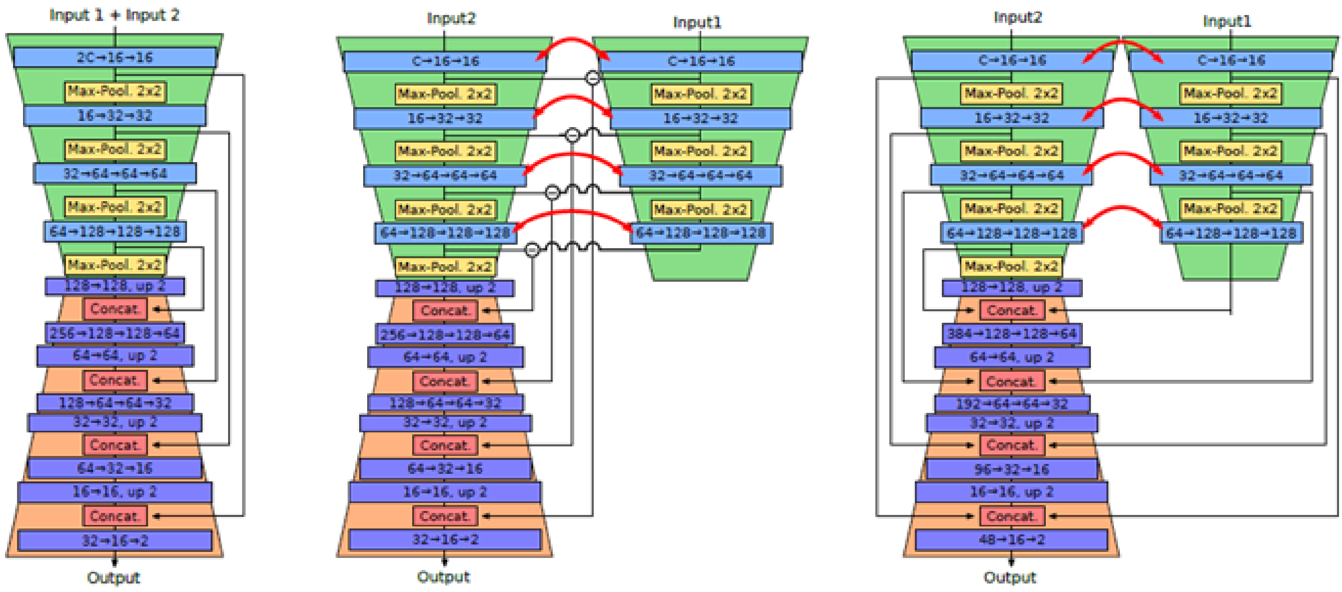
Figure 2: Deep Siamese Convolutional Neural Network

Caye Daudt, R.; Le Saux, B.; Boulch, A. Fully convolutional siamese networks for change detection. Proc.-Int. Conf. Image Process. ICIP 2018, 4063-4067. The work presented in this paper aimed to propose FCNN architectures able to learn to perform change detection solely from change detection datasets without any sort of pretraining or transfer learning from other datasets. These architectures are able to be trained end-to-end, unlike the majority of recent works on change detection. These fully convolutional architectures are an evolution of the work presented in [29], where a patch based approach was used. Moving the patch-based architectures to a fully convolutional scheme improves accuracy and speed of inference without affecting significantly the training times. These fully convolutional networks are also able to process inputs of any sizes given enough memory is available.

Two CNN architectures were compared in [29]: Early Fusion (EF) and Siamese (Siam). The EF architecture concatenated the two patches before passing them through the network, treating them as different color channels. The Siamese architecture processed both images separately at first by identical branches of the network with shared structure and parameters, merging the two branches only after the convolutional layers of the network. To extend these ideas the concept of skip connections that were used to build the U-Net, which aimed to perform semantic segmentation of images [30] was used. In summary, skip connections are links between layers at the same subsampling scale before and after the encoding part of an encoder-decoder architecture. The motivation for this is to complement the more abstract and less localized information of the encoded information with the spatial details that are present in the earlier layers of the network to produce accurate class prediction with precise boundaries in the output image.

The first proposed architecture is directly based on the UNet model, and was named Fully Convolutional Early Fusion (FC-EF). The U-Net model was adapted into the FC-EF taking into account the amount of available training data. The FCEF (Fig. 3 (a)) contains therefore only four max pooling and four upsampling layers, instead of the five present in the UNet Model. The layers in FC-EF are also shallower than their U-Net equivalents. As in the patch based EF model, the input of this network is the concatenation the two images in the pair that is to be compared.

The two other proposed architectures are Siamese extensions of the FC-EF model. To do so, the encoding layers of the network are separated into two streams of equal structure with shared weights as in a traditional Siamese network. Each image is given to one of these equal streams. The difference between the two architectures is only in how the skip connections are done. The first and more intuitive way of doing that is concatenating the two skip connections during the decoding steps, each one coming from one encoding stream. This approach was named Fully Convolutional Siamese-Concatenation (FC-Siam-conc, Fig. 3 (b)). Since in CD we are trying to detect differences between the two images, this heuristic was used to combine the skip connections in a different way. Instead of concatenating both connections from the encoding streams, the absolute value of their difference is concatenated instead. This approach was named Fully Convolutional Siamese-Difference (FC-Siam-diff, Fig. 3 (c)).



(a) FC-EF. (c) FC-Siam-diff. (b)FC-Siam-conc.

Figure 3: Schematics of the three proposed architectures for change detection. Block color legend: blue is convolution, yellow is max pooling, red is concatenation, purple is transpose convolution. Red arrows illustrate shared weights

Dual Learning-Based Siamese Framework for Change Detection Using Bi-Temporal VHR Optical Remote Sensing Images Bo Fang, Li Pan * and Rong Kou. Most of the methods generally focus on the research of simultaneously modeling and discriminating the changed and unchanged features. In practice, for bi-temporal VHR optical remote sensing images, the temporal spectral variability tends to exist in all bands throughout the entire paired images, making it difficult to distinguish non-changes and changes with a single model. In this paper, motivated by this observation, a novel hybrid end-to-end framework named dual learning-based Siamese framework (DLSF) for change Detection was proposed. The framework comprises two parallel streams which are dual learning-based domain transfer and Siamese-based change decision. The former stream is aimed at reducing the domain differences of two paired images and retaining the intrinsic information by translating them into each other's domain. While the latter stream is aimed at learning a decision strategy to decide the changes in two domains, respectively. By training the proposed framework with certain change map references, this method learns a cross-domain translation in order to suppress the differences of unchanged regions and highlight the differences of changed regions in two domains, respectively, then focus on the detection of changed regions.

2. Conclusion

Practically, since there is no one-size-fits-all solution, it is possible to complement the existing methods. For example, the outputs of the feature transformation methods are available to act the input of NN, and the results of the mathematical analysis can be applied for later clustering and classification. Experiments show that combining the shallow mathematical feature extraction method (e.g., PCA, IR-MAD) with the deep image semantic feature extraction method (e.g., SVM, DT, NN) can improve the performance of the results to some extent.

References

[1] Masroor Hussain, Dongmei Chen, Angela Cheng, Hui Wei, and David Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 80, pp. 91-106, 2013.

[2] Ashbindu Singh, "Review article digital change detection techniques using remotely-sensed data," International Journal of Remote Sensing, vol. 10, no. 6, pp. 989-1003, 1989.

[3] Bertrand Le Saux and Hicham Randrianarivo, "Urban change detection in sar images by interactive learning," in Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International. IEEE, 2013, pp. 3990-3993.

[4] Gang Liu, Julie Delon, Yann Gousseau, and Florence Tupin, "Unsupervised change detection between multisensory high resolution satellite images," in Signal Processing Conference (EUSIPCO), 2016 24th European. IEEE, 2016, pp. 2435-2439.

[5] Arabi Mohammed El Amin, Qingjie Liu, and Yunhong Wang, "Zoom out cnns features for optical remote sensing change detection," in Image, Vision and Computing (ICIVC), 2017 2nd International Conference on. IEEE, 2017, pp. 812-817.

[6] Yang Zhan, Kun Fu, Menglong Yan, Xian Sun, Hongqi Wang, and Xiaosong Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 10, pp. 1845-1849, 2017.

[7] Sergey Zagoruyko and Nikos Komodakis, "Learning to compare image patches via convolutional neural networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4353-4361.

[8] P. Zhong and R. Wang, "A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3978-3988, Dec. 2007.

[9] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and *k*-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772-776, Oct. 2009.

[10] J. Zhao, M. Gong, J. Liu, and L. Jiao, "Deep learning to classify difference image for image change detection," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2014, pp. 411-417.

[11] F. Gao, J. Dong, B. Li, Q. Xu, and C. Xie, "Change detection from synthetic aperture radar images based on neighborhood-based ratio and extreme learning machine," *J. Appl. Remote Sens.*, vol. 10, no. 4, p. 046019, 2016.

[12] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multilayer conditional mixed Markov model," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 10, pp. 3416-3430, Oct. 2009.

[13] P. Singh, Z. Kato, and J. Zerubia, "A multilayer Markovian model for change detection in aerial image pairs with large time differences," in *Proc. IEEE 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 924-929.

[14] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.

[15] F. Gao, J. Dong, B. Li, and Q. Xu, "Automatic change detection in synthetic aperture radar images based on PCANet," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1792-1796, Dec. 2016.

[16] Sun, H.; Zhou, W.; Zhang, Y.; Cai, C.; Chen, Q. Integrating spectral and textural attributes to measure magnitude in object-based change vector analysis. Int. J. Remote Sens. 2019, 40, 5749-5767.

[17] Cao, G.; Wang, B.; Xavier, H.-C.; Yang, D.; Southworth, J. A new difference image creation method based on deep neural networks for change detection in remote-sensing images. Int. J. Remote Sens. 2017, 38, 7161-7175.

[18] Zhuang, H.; Deng, K.; Fan, H.; Ma, S. A novel approach based on structural information for change detection in SAR images. Int. J. Remote Sens. 2018, 39, 2341-2365.

[19] Wu, C.; Du, B.; Zhang, L. Slow Feature Analysis for Change Detection in Multispectral Imagery. IEEE Trans. Geosci. Remote Sens. 2014, 52, 2858-2874.

[20] Kong, Y. L.; Huang, Q.; Wang, C.; Chen, J.; Chen, J.; He, D. Long short-term memory neural networks foronline disturbance detection in satellite image time series. Remote Sens. 2018

[21] Wu, H.; Prasad, S. Convolutional recurrent neural networks for hyperspectral data classification. Remote Sens.2017, 9, 298.

[22] Mou, L.; Bruzzone, L.; Zhu, X. X. Learning Spectral-Spatial-Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery. IEEE Trans. Geosci. Remote Sens. 2019, 57, 924-935.

[23] Jia, X.; Khandelwal, A.; Nayak, G.; Gerber, J.; Carlson, K.; West, P.; Kumar, V. Incremental Dual-Memory LSTMin Land Cover Prediction; ACM Press: New York, NY, USA, 2017; pp. 867-876.

[24] Gong, M.; Niu, X.; Zhang, P.; Li, Z. Generative Adversarial Networks for Change Detection in Multispectral Imagery. IEEE Geosci. Remote Sens. Lett. 2017, 14, 2310-2314.

[25] Wang, Q.; Zhang, X.; Chen, G.; Dai, F.; Gong, Y.; Zhu, K. Change detection based on Faster R-CNN forhigh-resolution remote sensing images. Remote Sens. Lett. 2018, 9, 923-932.

[26] L. Bruzzone and D. Prieto, "Automatic analysis of the difference image for unsupervised change detection," IEEE Trans. Geosci. Remote Sens., vol. 38, no. 3, pp. 1171-1182, May 2000.

[27] S. Zagoruyko and N. Komodakis, "Learning to compare image patchesvia convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 4353-4361.

[28] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2. Sep. 2006, pp. 1735-1742.

[29] Rodrigo Caye Daudt, Bertrand Le Saux, Alexandre Boulch, and Yann Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2018

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234-241.