



**HAL**  
open science

# GSoP Based Siamese Feature Fusion Network for Remote Sensing Image Change Detection

Puhua Chen, Lu Wang

► **To cite this version:**

Puhua Chen, Lu Wang. GSoP Based Siamese Feature Fusion Network for Remote Sensing Image Change Detection. 5th International Conference on Intelligence Science (ICIS), Oct 2022, Xi'an, China. pp.201-213, 10.1007/978-3-031-14903-0\_22 . hal-04666447

**HAL Id: hal-04666447**

**<https://hal.science/hal-04666447v1>**

Submitted on 1 Aug 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# GSoP based Siamese Feature Fusion Network for Remote Sensing Image Change Detection

Puhua Chen<sup>1</sup> and Lu Wang<sup>1</sup>

<sup>1</sup> The Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an 710071, China  
phchen@xidian.edu.cn, 1378039073@qq.com

**Abstract.** Change detection is an important research direction for remote sensing image application. Finding the change location automatically is the goal, which could provide useful information for disaster evaluation, disaster evaluation or urban development planning etc. In this paper, a Siamese feature fusion network is designed for change detection, which applies GSoP, a kind of attentional mechanism, to fuse the information of two feature extraction branches. The feature fusion strategy could build an information bridge which could also ensure the uniqueness of each branch, but also realize the information interaction of them to give more attention to important features during feature learning procedure. In the experimental section, many experiments were designed on many public datasets with some related methods. The results were shown that the proposed Siamese fusion network was efficient for change detection and had obvious advantage than some related methods.

**Keywords:** change detection; Siamese Network; attentional mechanism; remote sensing image.

## 1 Introduction

Change detection is one typical research direction of remote sensing, which have so many application scenes, such as environmental monitoring, marine pollution monitor, urban development, disaster analysis and battle damage assessment etc [1]. Although there are many decades of research on this topic, there still exist some problems appearing real application processing are not solved faultlessly. Firstly, remote sensing images used for change detection almost obtained from different sensors in different times, which could bring in many additional discrepancies to effect the change detection performance. Next, the difference of weather and season also may cause inaccurate change detection results. In addition, recently, many researchers focus into the change detection using different kinds of remote sensing sensors, such as optical sensor and SAR sensor. The difference of imaging mechanism may improve the difficulty of change detection. Therefore, to deal those problem for real applications of remote sensing technique, change detection still is a research topic of great research value.

Difference map generation and difference map segmentation are the key processing of typical change detection methods. Difference map could be obtained by many simple

methods such as ratio method, similarity measurement methods [2] or feature transformation methods [3] etc. Difference map segmentation also could be realized by many methods such as threshold methods, clustering methods and Markov methods [4] etc. With the advent of deep learning, change detection also emerge a deep learning boom. Many change detection methods based on deep learning are proposed, which are clearly analyzed and classified in review papers [5]. The advantage of deep learning on feature learning also brings high change detection results.

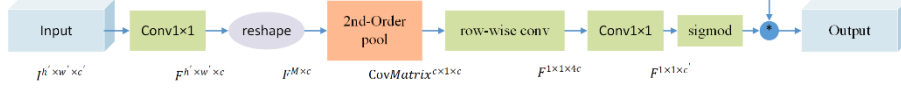
Although many advanced deep learning methods like contrast learning or self-learning, end-to-end pattern also is the most famous learning pattern of deep learning. Under this pattern, based on labeled data, deep learning methods could obtain features with high discrimination ability. It's similar to the requirement of change detection to find better feature space for distinguishing changed pixels with unchanged pixels. For change detection, there are at least two temporal images and one classification output. So, the information of different image must be fused in the network. Siamese network is commonly network for change detection recently. Two temporal image are inputted into feature learning branches with same structure and then combine high level feature together for the following classification layer.

In this paper, we proposed a change detection method based on Siamese network, which denoted as GSoP based Siamese Feature Fusion Network. The purpose of this work was to enhance the information interaction/fusion of two feature learning branches. Global Second-order Pooling (GSoP) is one typical attention module, which were use to built the information bridge of them. Based on GSoP fusion model, important feature channel could obtain much more attention during feature learning process, of which the importance weights were computed from two feature maps together. In experimental section, comparing with Siamese ResNet-32 based method and ResNet-32 based method, the experimental results verified the feasibility of the proposed methods. The efficiency of it was also proved by comparing with related deep learning methods. We also analyzed the effect of fusion location in GSoP module for final performance.

## 2 Methodology

### 2.1 GSoP Module

In [6], GSoP module was proposed to learn the attention factor for each channel, which could obtain better performance than Global Average Pooling in many visual tasks. The structure of GSoP module is shown in Fig. 1. In this module, the core is 2nd-Order pool. Firstly, the input feature maps denoted as  $I \in R^{h' \times w' \times c'}$  is transformed to  $F \in R^{h' \times w' \times c}$  through convolution with  $1 \times 1$  kernel. In this step, each channel of feature map F is treated as one random variable with  $M = h' \times w'$  random independent samplings. Therefore, feature maps  $F \in R^{h' \times w' \times c}$  can be represented as Eq. 1.



**Fig. 1.** the schematic diagram of GSoP.

$$F = \begin{bmatrix} F_{1,1} & F_{1,2} & \dots & F_{1,c} \\ F_{2,1} & F_{2,2} & \dots & F_{2,c} \\ \vdots & \vdots & \vdots & \vdots \\ F_{M,1} & F_{M,2} & \dots & F_{M,c} \end{bmatrix} = [c_1, c_2, \dots, c_c] \quad (1)$$

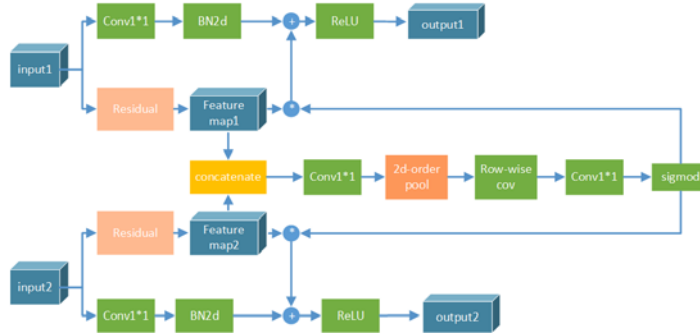
Then, the covariance matrix of  $F$  can be computed through Eq. 2, where  $I$  is a  $M \times M$  matrix, of which whole elements are 1 and  $E$  is a  $M \times M$  diagonal matrix.

$$CovMatrix = F^T \left[ -\frac{1}{M \times M} I + \frac{1}{M} E \right] F \quad (2)$$

The covariance matrix contains the relevant information among  $c$  feature channels, which is better than average value to present the importance of channels. After that, considering the intrinsic structure information, covariance matrix is operated by row-wise convolution which is realized by group convolution with  $c$  group. The combination of row-wise convolution with  $1 \times 1$  convolution is equivalent to depthwise separable convolution. Through the active layer with S function, the weight vector  $c \times 1$  is outputted to multiply with original feature map  $I$ . Through above options, feature learning process could give more attention to more important feature channels.

## 2.2 Information Fusion based on GSoP

Here, three different information fusion models based on GSoP are proposed. The first fusion model is shown in Fig. 3. Feature maps of two branches are spliced on channel dimension and inputted into GSoP block. The final attention weights are computed on these feature maps through those operators in GSoP block, and then they are superposed



**Fig. 2.** The first information fusion model.

into the original feature maps through channel weighting, which reflect the importance of each channel. Based on the weighting processing, feature learning processing could pay much attention on those more important feature channels. In this block, this attention computation can be treated as information fusion of two feature learning branches.

The second fusion model is shown in Fig. 4. In this model, the feature maps of two branches are changed the channel number through  $1 \times 1$  convolution respectively and reshaped as same as GSoP module mentioned above. Then, these reshaped feature maps are concatenated together to computed the important weight and jointed into the original feature maps of two branches.

The third fusion model is shown in Fig. 4. Feature maps of two branches are concatenated after  $1 \times 1$  convolution, reshaping and second-order pooling operators to compute the important weight.

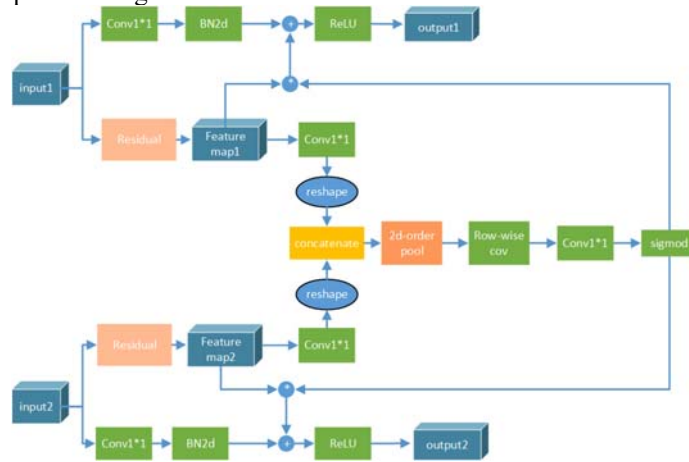


Fig. 3. The second information fusion model.

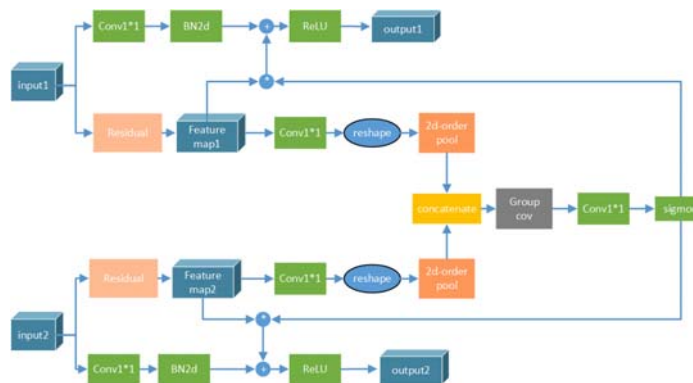


Fig. 4. The third information fusion model.

In above contents, three information fusion model were introduced, which could fuse different information from different images of same scene. The information fusion was

mainly realized by the attention weight computation based on GSoP. The difference of those three models is the concatenation of two branch information is operated in different stages of GSoP. In the experimental section, some experiments were designed to evaluate the performance of them.

### 2.3 Siamese Feature Fusion Network with GSoP for change detection

In this work, ResNet-32[7] was chosen as the basic network framework, which have three Layers one of which contains 5 residual blocks with two convolution operators respectively. Here, aiming to verify the performance of information fusion for change detection, original change detection framework based on ResNet-32 and Siamese ResNet-32 are also discussed, which are shown in Fig. 5 and Fig. 6. In Fig. 5, there is just a data fusion, where two different images are treated as different data channel and learn features synthetically. In Fig. 6, two different images have their own feature learning network and finally learned features are fused together for the next classification.

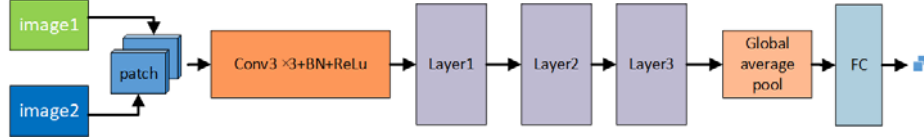


Fig. 5. The schematic diagram of ResNet-32 for change detection.

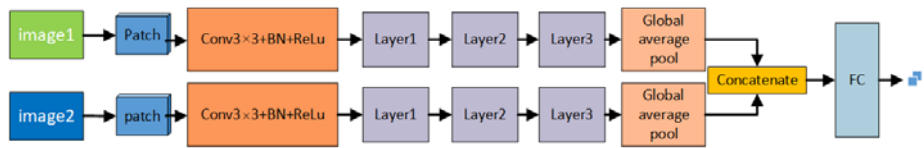


Fig. 6. The schematic diagram of Siamese ResNet-32 for change detection.

Although the method based on Siamese ResNet-32 have two feature learning branches for two different images, there is not any information interaction or information fusion among the feature learning processing. In this paper, a Siamese Feature Fusion Network based on GSoP is proposed for change detection, of which the schematic diagram is shown in Fig. 7. Different from the above methods, the Layers of two branches are not independently, they are connected with each other by using GSoP moduel. Because of this procedure, two advantage could be obtained. First, attention mechanism is involved into feature learning, which could promote the performance. Second, it builds the information bridge of two feature learning branches and could enhance information fusion degree to obtain better results.

In the proposed Siamese fusion network for change detection, GSoP fusion model described in above subsection is applied in each Layer shown in Fig. 8. Just the last residual blocks are replaced by GSoP fusion model of each Layer.

For change detection, Siamese fusion network shown in Fig. 7 are firstly trained on training data of which each data/pixel are labeled as unchanged/0 or changed/1. The

rest pixels are testing data. For testing data, the classification probability could be obtained from trained Siamese fusion network. Commonly, the change probability map is treated as difference map.

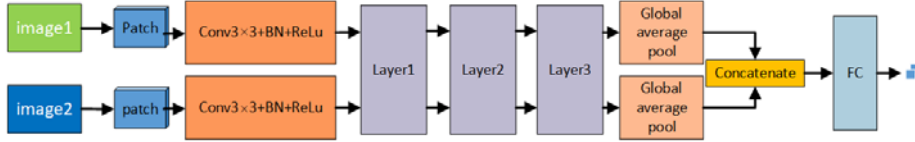


Fig. 7. The schematic diagram of Siamese fusion network based on ResNet-32 for change detection.

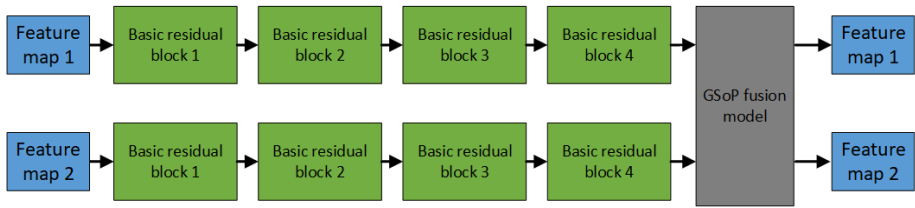


Fig. 8. the schematic diagram of each Layer in Siamese fusion network.

After obtained the different map, a simple threshold segmentation method is used to divide the difference map into two data sets. Pixels with higher difference value than threshold are distinguished as changed pixels and pixels with lower difference value are distinguished as unchanged pixels.

### 3 Experiments and analysis

In this section, aiming to verify the performance of the proposed Siamese fusion network based on GSoP, three public datasets are chosen and DNN based change detection method[8], CNN-LSTM based method[9], ResNet-32 based method and Siamese ResNet-32 based method are also involved in experiments as comparing methods.

#### 3.1 Dataset Introduction and Experiments Setting

There were many public datasets for change detection. In this paper, Tiszadob-3 Dataset, Szada-2 Dataset, QuickBird-1 Dataset and Shuguang Village Dataset were chosen as experimental datasets, which are usually used to verify performance in related works. Tiszadob-3 Dataset and Szada-2 Dataset were selected from SZTAKI Air-Change Benchmark [10], of which the spatial resolution was 1.5m. Tiszadob-3 Dataset displayed the areas changing from one kind of vegetation to another vegetation, which is obvious in two images. However, Szada/2 only labeled areas changing from vegetation to human-made ones, such as roads, buildings and places. QuickBird-1 Dataset was provided by the Data Innovation Competition of Guangdong Government, which was captured in 2015 and 2017. Shuguang Village Dataset is a heterogeneous data which involves two images captured in different time by different imaging sensors.



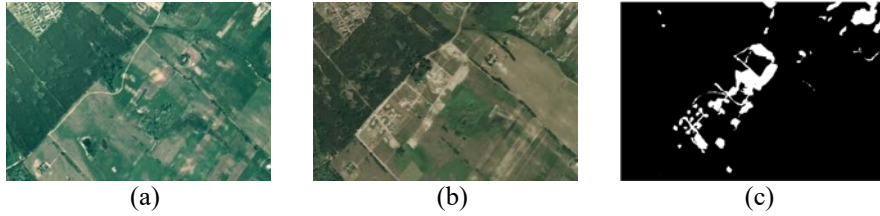
Original images obtained in different times and the ground truthing of those three datasets were shown in Fig. 9, Fig. 10, Fig. 11 and Fig. 12 respectively.

In experiments, 400 changed samples and 1600 unchanged samples were chosen randomly for network training. The size of image patch was set to be  $10 \times 10$ . The preliminary learning rate was set to be 0.001 and reduced into 0.1 percentage of original value per 80 iterations. The maximum iteration number was 200. The whole network was built on The PyTorch framework and running in the computer with NVIDIA GeForce RTX2060.

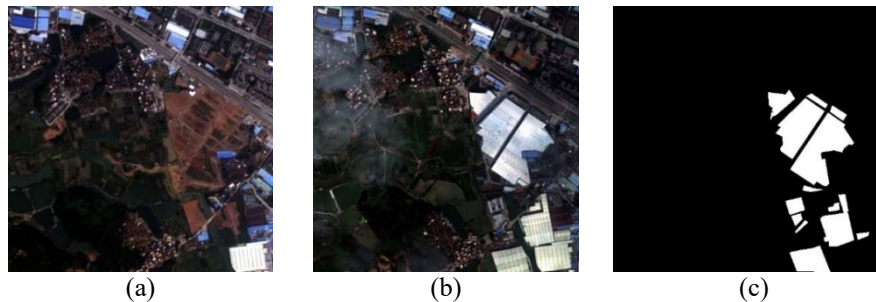
For well evaluating the performance of those methods, five numerical indexes were utilized in experiments. It contains Precision rate (Pre), Recall rate (Rec), Accuracy rate (Acc), Kappa coefficient (Kappa) and F1 coefficient. Those numerical indexes could verify the performance of each method across-the-board.



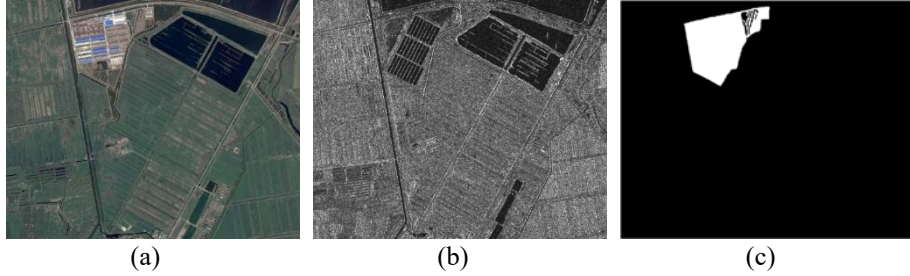
**Fig. 9.** Tiszadob-3 dataset. (a) is the image obtained in time 1; (b) is the image obtained in time 2; (c) is the ground truthing.



**Fig. 10.** Szada-2 dataset. (a) is the image obtained in time 1; (b) is the image obtained in time 2; (c) is the ground truthing.



**Fig. 11.** QuickBird-1 dataset. (a) is the image obtained in time 1; (b) is the image obtained in time 2; (c) is the ground truthing.

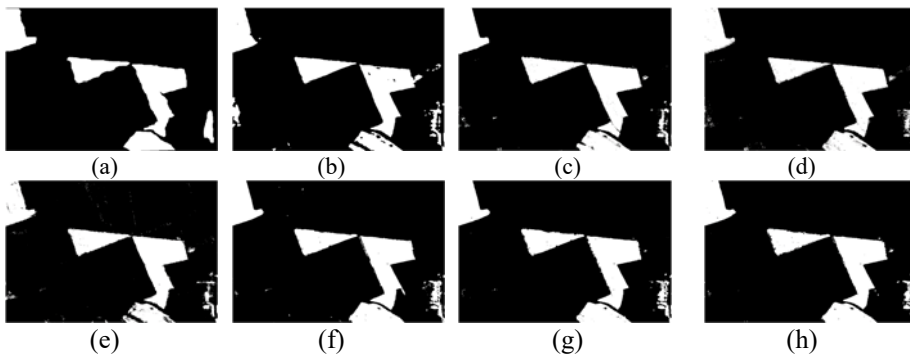


**Fig. 12.** Shuguang Village dataset. (a) is the image obtained in time 1 by optical sensor; (b) is the image obtained in time 2 by SAR sensor; (c) is the ground truthing.

### 3.2 Experimental results and analysis

In this subsection, experimental results on above three datasets were shown and analyzed. For Tiszadob-3 dataset, the change detection maps obtained by different methods were shown in Fig. 13. Model-1, Model-2 and Model-3 were the proposed Siamese fusion network with different fusion models described in subsection 2.2. Most of changes shown in ground truthing were detected by those methods. However, each method had more or less false detection or leak detection. For clearly analyzing the results, Table 1 gave the numerical index results. By analyzing these results, the proposed fusion models could obtain better results because of the advantage shown on Most indicators such as Rec, Acc, Kappa and F1.

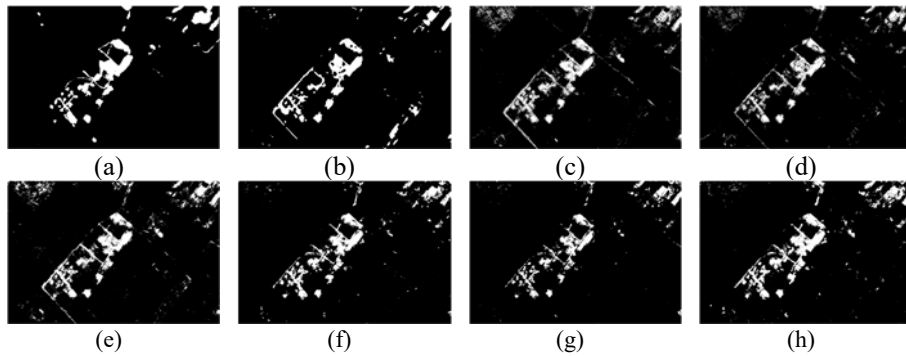
Results on Szada-2 dataset were shown in Fig. 14 and Table 2. Comparing these change detection maps in Fig. 14, less unchanged pixels were detected as changed pixels about these proposed methods (Model-1, Model-2 and Model-3) than other threes comparing methods. In addition, the leak detection of changed pixels also was lower of these proposed methods. The proposed methods with feature fusion strategy could obtain better performance than others, which also was verified by those numerical results shown in Table 2.



**Fig. 13.** experimental results on Tiszadob-3 dataset. (a) ground trothing, (b) DNN, (c) CNN-LSTM, (d) ResNet-32, (e) Siam-ResNet-32, (f) Model-1, (g) Model-2, (h) Model-3.

**Table 1** Numerical index results on Tiszadob-3 dataset.

Methods	Pre	Rec	Acc	Kappa	F1
DNN	0.606	0.624	0.955	0.615	0.591
CNN-LSTM	0.930	0.939	0.981	0.935	0.923
Resnet-32	<b>0.931</b>	0.919	0.978	0.925	0.912
Siam-ResNet-32	0.914	0.914	0.975	0.914	0.899
Model-1	0.913	<b>0.971</b>	0.982	0.941	0.931
Model-2	0.922	0.967	<b>0.983</b>	<b>0.944</b>	<b>0.934</b>
Model-3	0.927	0.955	0.982	0.941	0.930

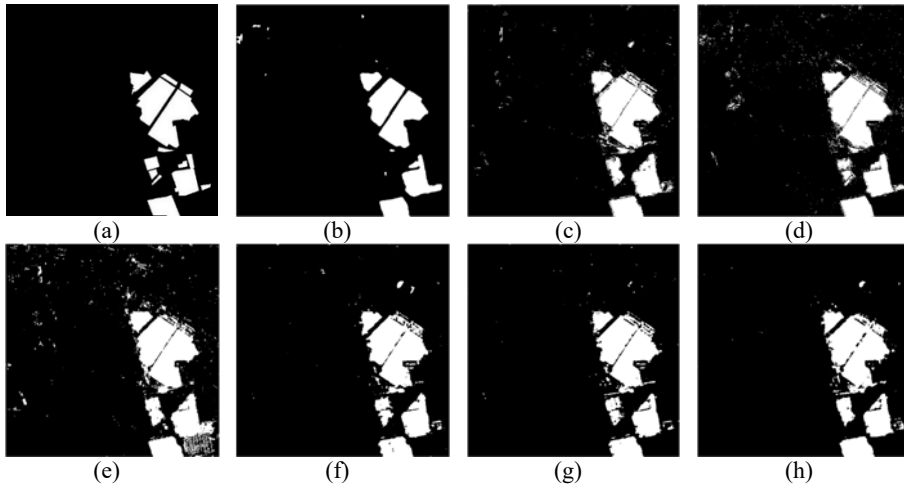
**Fig. 14.** experimental results on Szada-2 dataset. (a) is ground trothing, (b) DNN, (c) CNN-LSTM, (d) ResNet-32, (e) Siam-ResNet-32, (f) Model-1, (g) Model-2, (h) Model-3.**Table 2** Numerical index results on Szada-2 dataset.

Methods	Pre	Rec	Acc	Kappa	F1
DNN	0.606	0.624	0.955	0.615	0.591
CNN-LSTM	0.532	0.690	0.947	0.601	0.573
ResNet-32	0.639	0.623	0.958	0.631	0.609
Siam-ResNet-32	0.550	0.693	0.950	0.613	0.587
Model-1	<b>0.726</b>	0.731	<b>0.969</b>	<b>0.728</b>	<b>0.712</b>
Model-2	0.682	<b>0.758</b>	0.966	0.718	0.700
Model-3	0.708	0.707	0.966	0.707	0.690

Results on QuickBird-1 dataset were given in Fig. 15 and Table 3. In this dataset, the ground trothing showed the changes from soil to building, which was obvious in original images. In those change detection maps of DNN, LSTM Resnet32 and Siam-Resnet, there were many false detection pixels in the left part. But, in the changed area, more changed pixels were not accurately detected. In numerical index results, just Siam-Resnet obtained the better results than proposed methods only on Rec. On other numerical indexes, proposed methods had obvious advantage performance than others.

Shuguang Village Dataset contains two heterogeneous images obtained by different sensors. The difference of imaging mechanism enhanced the difference of images, which would make change detection more difficult than homologous images. The

change detection results on Shuguang Village Dataset were given in Fig. 16 and Table 4. Because the results of DNN was too bad than other methods, we ignored it on this dataset. Observed from change detection maps in Fig. 16, the detection results of CNN-LSTM, ResNet-32 and Siam-ResNet-32 were better in the outside of the changed area. However, in the inside of the change area, there were some unchanged pixels, which were not detected by them. However, those proposed methods could still detect them. Observed form those numerical results shown in Table 4, those proposed methods had obvious advantages on Pre, Rec, Kappa and F1.



**Fig. 15.** experimental results on QuickBird-1 Dataset. (a) is ground trothing, (b) DNN, (c) CNN-LSTM, (d) ResNet-32, (e) Siam-ResNet-32, (f) Model-1, (g) Model-2, (h) Model-3.

**Table 3** Numerical index results on QuickBird-1 dataset.

Methods	Pre	Rec	Acc	Kappa	F1
<b>DNN</b>	0.606	0.624	0.955	0.615	0.591
<b>CNN-LSTM</b>	0.835	0.906	0.972	0.869	0.853
<b>ResNet-32</b>	0.819	0.894	0.968	0.855	0.837
<b>Siam-ResNet-32</b>	0.745	<b>0.917</b>	0.959	0.822	0.799
<b>Model-1</b>	0.869	0.887	0.974	0.878	0.863
<b>Model-2</b>	<b>0.877</b>	0.883	<b>0.975</b>	0.880	0.866
<b>Model-3</b>	0.865	0.901	<b>0.975</b>	<b>0.883</b>	<b>0.869</b>

From these results shown in Table 1, Table 2, Table 3 and Table 4, we found that the proposed three different fusion models had obtained better results than those comparing methods. Especially, the comparison with ResNet-32 and Siam-ResNet-32 had shown that the information fusion between two feature learning branches could promote the performance of change detection. However, we also found that the performance differ-

ence of Model-1, Model-2 and Model-3 was not very remarkable. Therefore, we considered that the information fusion based on GSoP was efficient but the stages where fusion operated were not important for the final performance.

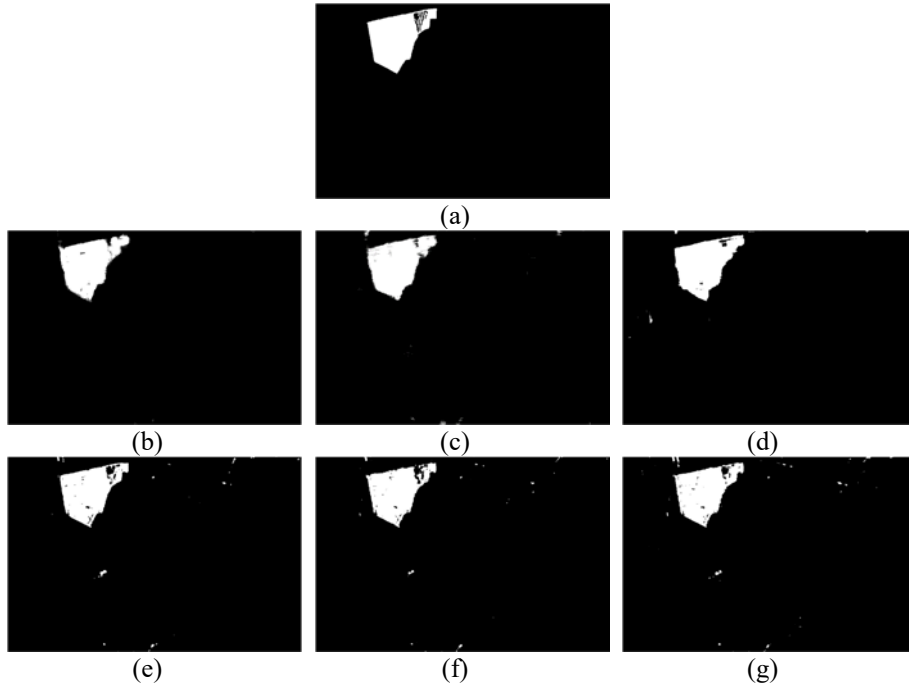


Fig. 16. experimental results on Shuguang Village Dataset. (a) is ground trothing, (b) CNN-LSTM, (c) ResNet-32, (d) Siam-ResNet-32, (e) Model-1, (f) Model-2, (g) Model-3.

Table 4 Numerical index results on Shuguang Village Dataset.

Methods	Pre	Rec	Acc	Kappa	F1
CNN-LSTM	0.893	0.899	0.992	0.896	0.892
ResNet-32	0.884	0.822	0.988	0.852	0.846
Siam-ResNet-32	<b>0.954</b>	0.869	0.993	0.909	0.906
Model-1	0.941	<b>0.957</b>	<b>0.996</b>	<b>0.949</b>	<b>0.947</b>
Model-2	0.953	0.936	<b>0.996</b>	0.944	0.942
Model-3	0.930	0.958	0.995	0.943	0.941

## 4 Conclusion

In this work, considering the importance of information fusion of feature extraction for change detection, a Siamese fusion network for change detection was proposed, which utilized GSoP to construct the information fusion model. Experimental results verified

the efficient of information fusion compared with related methods. However, the performance relied on labeled samples which involved high labor cost. In the future, self-supervised learning combined with fewer human guidance for changed detection would be valuable to be researched deeply.

## References

1. Ban, Y., and O. Yousif: Change Detection Techniques: A Review. Springer International Publishing (2016).
2. An L, Li M, Zhang P, et al.: Multicontextual Mutual Information Data for SAR Image Change Detection. *IEEE Geoscience and Remote Sensing Letters*, 12(9):1863-1867(2015).
3. Seto K C, Woodcock C E, Song C, et al. Monitoring Land-Use Change in The Pearl River Delta Using Landsat TM. *International Journal of Remote Sensing*, 23(10),1985-2004 (2020).
4. Yusif O, Ban Y. Improving SAR-based Urban Change Detection by Combining MAP-MRF Classifier and Nonlocal Means Similarity Weights. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(10), 4288-4300 (2014).
5. Khelifi, Lazhar, and Max Mignotte: Deep Learning for Change Detection in Remote Sensing Images: Comprehensive Review And Meta-Analysis. *IEEE Access*, 8, 126385-126400 (2020).
6. Gao Z, Xie J, Wang Q, et al.: Global Second-Order Pooling Convolutional Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2019).
7. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016).
8. Gong M, Zhao J, Liu J, et al.: Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 27(1),125-138(2017).
9. 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 Transactions on Geoscience and Remote Sensing*, 57(2), 924-935(2018).
10. SZTAKI AirChange Benchmark: [http://web.eec.sztaki.hu/remotesensing/airchange\\_benchmark.html](http://web.eec.sztaki.hu/remotesensing/airchange_benchmark.html).