SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Multi-layer composite autoencoders for semi-supervised change detection in heterogeneous remote sensing images

Jiao Shi[1], Tiancheng Wu[1], Hanwen Yu[2], A. K. Qin[3], Gwanggil Jeon[4] & Yu Lei[1*]

[1] *School of Electronics and Information, Northwestern Polytechnical University, Xi'an* 710129*, China;*
[2] *School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu* 611731*, China;*
[3] *Swinburne University of Technology, Melbourne* 3122*, Australia;*
[4] *Incheon National University, Incheon* 032) 835-0715*, South Korea*

Appendix A Related works

With the development of remote sensing technology, the variety of remote sensing images is also in full bloom subsequently, such as sar images, optical images, multispectral images, hyperspectral images and so on [1–3].

When the sensor sources of the two images are different, they are called heterogeneous images [5]. Different from the traditional change detection of homologous remote sensing images, due to the huge differences between heterogeneous images caused by such as the sensitivity to light, reflection properties of the objects, etc [6], the traditional change detection framework is not suitable for such problems. Therefore, researchers are focusing on new methods for processing heterogeneous remote images.

For the problem of change detection for heterogeneous remote images, it is mean to align the images of two different domains, and then compare the original data in the common domain to highlight the difference [7]. For additional tasks of image alignment, Liu er al. [8] proposed a symmetric convolutional coupling network (SCCN), adopting a probability map to the differences between two images. Liu er al. [9] proposed an unsupervised change detection (USCD), which applied cycle consistency to obtain image mapping relation between heterogeneous images. Luigi er al. [10] proposed code-aligned autoencoders(CAAE), which utilized four learning strategies of reconstruction, cycle consistency, weighted translation and code correlation to enforce alignment of the code spaces. Therefore, more and more methods for processing heterogeneous remote images are under research, which can greatly improve the utilization rate of remote sensing data.

Appendix B Experiments and results

Appendix B.1 Dataset

Four classical datasets of remote sensing images are tested in the experiments, which is shown in Fig. B1- B4. Each pair of them has been radio-metrically corrected and coregistered to make them as more comparable as possible.

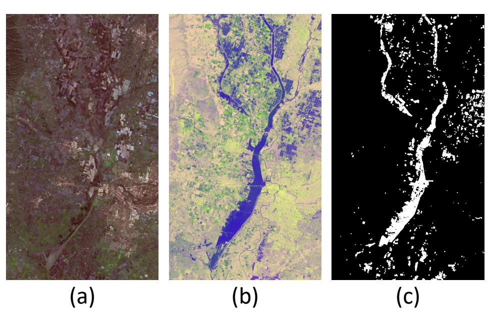

Figure B1 D1—Flood in California: (a)I_1, (b)I_2, (c) Reference map.

1) D1—Flood in California: The first dataset consists of a multispectral image and a sar image, which are size of $875 * 500 * 11$ in Fig. B1(a) and $875 * 500 * 3$ in Fig. B1(b) repectively. The original size is $3500 * 2000$ pixels, and they were resampled to $850 * 500$ pixels for less computation. The ground truth map was marked manually in Fig. B1(c) from [4].

2) D2—Lake Overflow in Italy: The second dataset consists of a near infrared image and an optical image, which are size of $412 * 300 * 1$ in Fig. B2(a) and $412 * 300 * 3$ in Fig. B2(b) repectively. They are both Landsat 5 images with channels being not overlapping. Fig. B2(c) shows the ground truth map.

* Corresponding author (email: leiy@nwpu.edu.cn)

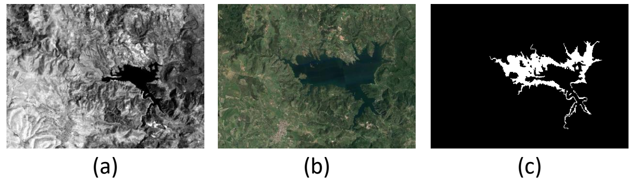

Figure B2 D2—Lake Overflow in Italy: (a)I_1, (b)I_2, (c) Reference map.

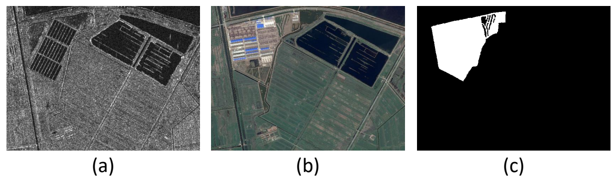

Figure B3 D3—Farmland in Shuguang: (a)I_1, (b)I_2, (c) Reference map.

3) D3—Farmland in Shuguang: The third dataset consists of a sar image and an optical image, which are size of $549 * 411 * 1$ in Fig. B3(a) and $549 * 411 * 3$ in Fig. B3(b) repectively. It was taken by Radarsat-2 to record the changes of a piece of farmland in Shuguang, Dongying, China. Fig. B3(c) is the ground truth image.

4) D4—Flood in Yellow River: The fourth dataset consists of a sar image and an optical image, which are size of $291 * 343 * 1$ in Fig. B4(a) and $291 * 343 * 1$ in Fig. B4(b) repectively. The former one was obtained from Google Earth and the latter one was taken by Landsat-7. The ground truth map is shown in Fig. B4(c).

Appendix B.2 Experimental setup

1) **Processing flow:** The proposed framework uses a small number of labeled samples to complete the initial detection task, and then captures the difference information of optimized features in a large number of unlabeled data. By filtering with high confidence selection, large amount of new pseudo labeled data set can be obtained in multi-layer composite judgment and then be added to the original true-labeled training set to form new training set. Finally, the final change detection result is iteratively optimized to attain.

The iteration can be stopped when the change degree of the pseudo-labeled training sample is less than a fixed value. The fixed value is set by artificial experience according to the experimental results. Specifically, the degree of change refers to the number change of pseudo-labeled samples and the position repetition of the screened pixels. In the experiment, the number of labels screened after one iteration is usually about 90%, and the pixel position repeats greatly with the second iteration, so the network can basically complete the final output after one iteration

In addition, when the pixel position of the vote is consistent with the position of a true label, the label value of the true label will be directly retained, regardless of whether the result of the pseudo label voting is consistent. When there is no true label of one position, the pixel will vote for screening.

2) **Network parameters:** The proposed framework is composited of two autoencoders(AE). Each AE is a fully convolution neural networks, composed of four encoding layers: $Conv(3 * 3 * 20) - ReLU - Coup(1 * 1 * 20) - ReLU - Coup(1 * 1 * 20) - ReLU - Coup(1 * 1 * 1) - Sigmoid$. At the same time, the decoder has the opposite structure. The learning rate is 10e-4, batch size is 1000, epoch set is 100, weight decay is 0.9, and optimizer is based on Adam algorithm. The experiments are performed on Intel(R) Core(TM) i7-4790 CPU @3.60GHz 3.60 GHz with 16 GB of RAM. Tensorflow 1.4 framework with Python 3.6.10 is as the programing language.

Appendix B.3 Evaluation metrics

Several metrics are used to evaluate the performance of the proposed framework, whose calculation methods are shown as follows. The values are obtained by comparison between the change detection result map and the ground truth map, where 0 represents an unchanged pixel and 1 represents a changed pixel.

1) TP: True positives(changed pixels that are correctly detected).
2) TN: True negatives(unchanged pixels that are correctly detected).
3) FP: False positives(unchanged pixels that are wrongly detected as changed one).
4) FN: False negatives(changed pixels that are wrongly detected as unchanged one).
5) OA: (TP+TN)/(TP+TN+FP+FN); the ratio between correctly detected pixels and the total amount of pixels.

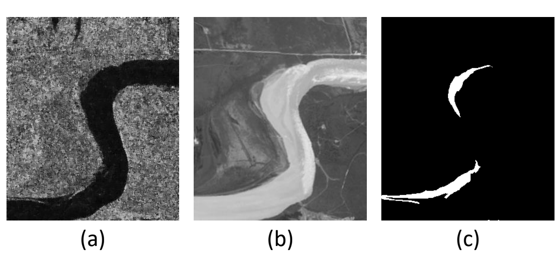

Figure B4 D4—Flood in Yellow River: (a)I_1, (b)I_2, (c) Reference map.

6) PRE: [(TP+FP)(TN+FN)+(FN+TN)(FP+TN)]/(TP+TN+FP+FN)2̂; the probability of random agreement.

7) Kappa: (OA-PRE)/(1-PRE); the consistency of two classifiers.

8) F1: TP/[TP+1/2(FP+FN)]; balanced evaluation of accuracy and recall.

9) AUC: The performance with values from 0 (opposite detection) to 1 (optimal detection).

Last, the metrics of FP, FN, OA, Kappa, F1 and AUC will be shown in the exhibition compared with unsupevised methods of SCCN [8], ACE-Net [**?**] and X-Net [**?**], supervised methods of RFR [**?**], classification network based on Resnet [**?**] and classification network based on multilayer perceptron (MLP) [14] and semi-supervised method of Semi-GAN [15]. Besides, in experiments, we purposely compared the classification performance of the pure supervised methods and the semi-supervised methods under the same 60% label quantity. And the experiment of 0.1% label quantity is used to prove that, due to the scarcity of true labels in the remote sensing field, the compared unsupervised models are able to get rid of the label dependency of remote sensing images. And though the proposed model is a semi-supervised network, it only needs 0.1% labeled training samples. So the cost of labeled samples is similar to the unsupervised one. Specially, the experimental result of 0.1% label amount is the result of the minimum label amount that combines the algorithm cost and the experimental performance. The other experimental setup of the comparison algorithms is consistent with Luigi er al. proposed in [**?**]. But the training epoches of ACE-Net and X-Net were set to 60 due to the limitation from the test platform.

Appendix B.4 Experimental results

1) **D1—Flood in California:**

For California dataset, the change detection results are shown in Fig. B5 and test metrics are displayed in Table. B1. From the comparison between unsupervised methods and supervised methods, it is obvious that unsupervised methods such as SCCN, ACE-Net and X-Net do not have the guidance of true-labeled samples, so a large number of red undetected areas appear in the changed areas. Other methods with labels have a large number of green false detected change areas in the upper part of each change map in the Fig. B5. However, by using the information of unlabeled samples, it can be seen in Fig. B5(i) and (j) that the false detected phenomenon has been weakened. On the whole, at the nearly cost of unsupervised models,the proposed framework achieves the performance of 0.914 in OA and 0.505 in Kappa which is shown in Table. B1. And under the unified 60% label usage, the semi-supervised method is obviously superior to the pure supervised method, and the proposed method achieves the best results of OA under the initial network startup with a larger number of labels. For the time cost, the cost of supervised machine learning methods is the least, because they do not include a lot of training time for unlabeled data. In addition to them, the time complexity of the proposed method with 0.1% label amount is the lowest, which proves that our framework is a concise network. However, as the number of labels increases, the time cost of the framework will also increase dramatically. It can be seen that the time cost of proposed method with 60% labels is almost twice that of the one with 0.1% labels, which is more obvious in the large data set D1 data set, for they contain more classification feedback of labeled data.

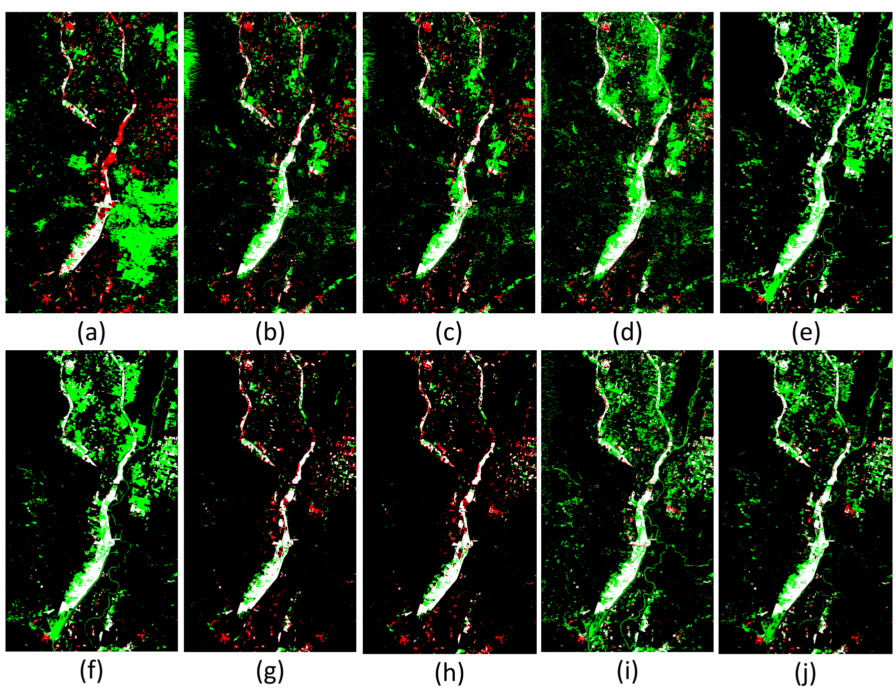

Figure B5 Change detection results of D1—Flood in California achieved by (a) SCCN, (b) ACE-Net, (c) X-Net, (d) RFR, (e) Resnet, (f) MLP, (g) Semi-GAN(60%), (h) proposed(60%), (i) Semi-GAN(0.1%) and (j) proposed(0.1%). (TP: white; TN: black; FP: green; FN: red)

2) **D2—Lake Overflow in Italy:**

For Italy dataset, the change detection results are shown in Fig. B6 and test metrics are displayed in Table. B2. The phenomenon of missing detection and false detection is the most obvious in the Italy dataset of all results. For example, in unsupervised methods, SCCN has a large number of undetected changed areas in Fig. B6(a), while ACE-Net and X-net has fewer undetected changed areas, but then a large number of green false detected areas appear in the lower left corner area, which is shown in Fig. B6(b) and (c). This means that unsupervised methods are difficult to balance changed and unchanged identification information. In

Table B1 Experimental results of D1—Flood in California

D1	Methods	FP	FN	OA	Kappa	F1	AUC	Time cost
Unsuper-vised	SCCN	60163	15201	0.828	0.162	0.237	0.700	925.7
	ACE-Net	34767	9795	0.898	0.385	0.434	0.875	6918.3
	X-Net	31837	10207	0.904	0.395	0.443	0.879	3419.4
Super-vised(60%)	RFR	61399	6573	0.845	0.311	0.374	0.871	299.3
	Resnet	53529	2124	0.873	0.418	0.471	0.916	387.2
	MLP	50796	2491	0.878	0.426	0.478	0.918	179.5
Semi-supervised(60%)	Semi-GAN	7146	9624	0.961	0.652	0.672	0.812	3654.1
	Proposed	5356	10873	0.963	0.645	0.664	0.791	2951.4
Semi-supervised(0.1%)	Semi-GAN	47102	3949	0.883	0.421	0.472	0.869	865.9
	Proposed	33248	4339	0.914	0.505	0.546	0.879	592.6

contrast, the proposed framework makes good use of unsupervised information and the coarse-and-fine features to fuse and filter a large number of noise points with 0.1% label quantity. So it achieves excellent change detection results. At the same time, with 60% label quantity, Semi-GAN and the proposed framework almost perfectly detect the changes in the Italy dataset, leading the Table. B2.

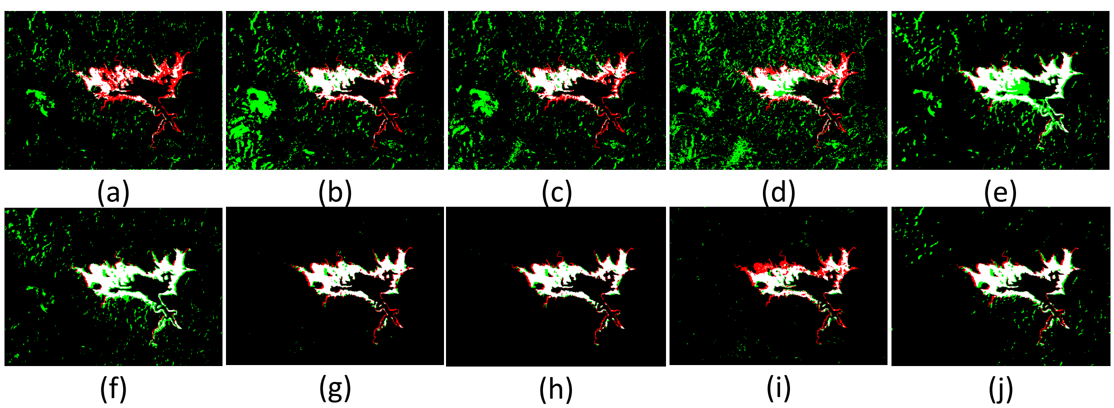

Figure B6 Change detection results of D2—Lake Overflow in Italy achieved by (a) SCCN, (b) ACE-Net, (c) X-Net, (d) RFR, (e) Resnet, (f) MLP, (g) Semi-GAN(60%), (h) proposed(60%), (i) Semi-GAN(0.1%) and (j) proposed(0.1%). (TP: white; TN: black; FP: green; FN: red)

3) **D3—Farmland in Shuguang:**

For Shuguang dataset, the change detection results are shown in Fig. B7 and test metrics are displayed in Table. B3. The detection difficulty of shuguang dataset is that $I1$ and $I2$ have different surface types in the changed area on the upper left corner, which is shown in Fig. B3(a) and (b). This leads to either a horizontal test result close to $I2$ in Fig. B7(a) or an obvious vertical test result close to $I1$ in Fig. B7(b) and (c). In addition, because the imaging mechanisms of the two heterogeneous images in Shuguang dataset are different, the uppermost road area is often over-detected into a changed area, which is particularly obvious in the labeled supervised methods such as Fig. B7(e), (f), (i) and (j). This over-fitting phenomenon occurs because the model excessively learns the characteristics of changed samples. However, when the number of labeled samples is increased, Semi-GAN and the proposed method can achieve a better balance between changed features and unchanged features. In general, the proposed method achieves the best results at 60% label amount, and achieves the best detection performance at 0.1% label amount near the condition of no label.

4) **D4—Flood in Yellow River:**

For Yellow River dataset, the change detection results are shown in Fig. B8 and test metrics are displayed in Table. B4. It can be seen from Fig. B8(a), (b) and (c) that the YellowRiver dataset is prone to global noise in the image, because $I1$ image in Fig. B4(a) has many hollow scattered points due to lighting and other reasons, while $I2$ as an optical image in Fig. B4(b) is more smooth. Therefore, there is serious interference when comparing the two. In addition, due to the subtle texture differences on the optical image $I2$, there are wrong-detected place in triangular areas on the left half and river areas on the right half in Fig. B8(e), (i) and (j). So how to convert and eliminate the difference between heterogeneous images is the key to change detection of heterogeneous images. And even if 60% of the labels are used, Semi-GAN is still difficult to reduce the influence of fine horizontal lines in the upper part of the optical image $I2$. However, the proposed framework finally achieves the best change detection results through the fusion of multiple change maps, which is shown in Table. B4, the highest OA of 0.981 and Kappa of 0.691.

Table B2 Experimental results of D2—Lake Overflow in Italy

D2	Methods	FP	FN	OA	Kappa	F1	AUC	Time cost
Unsuper-vised	SCCN	3947	4456	0.932	0.394	0.430	0.844	397.3
	ACE-Net	9543	2294	0.904	0.427	0.474	0.894	6374.3
	X-Net	7861	2826	0.914	0.429	0.473	0.872	3327.8
Super-vised(60%)	RFR	13980	2776	0.864	0.306	0.367	0.866	16.5
	Resnet	8085	689	0.929	0.578	0.613	0.959	112.2
	MLP	5529	652	0.950	0.667	0.693	0.971	54.1
Semi-supervised(60%)	Semi-GAN	796	1641	0.980	0.820	0.831	0.889	447.6
	Proposed	812	1523	0.981	0.829	0.839	0.897	393.1
Semi-supervised(0.1%)	Semi-GAN	871	3210	0.967	0.667	0.684	0.786	247.6
	Proposed	2491	1367	0.969	0.748	0.764	0.900	197.1

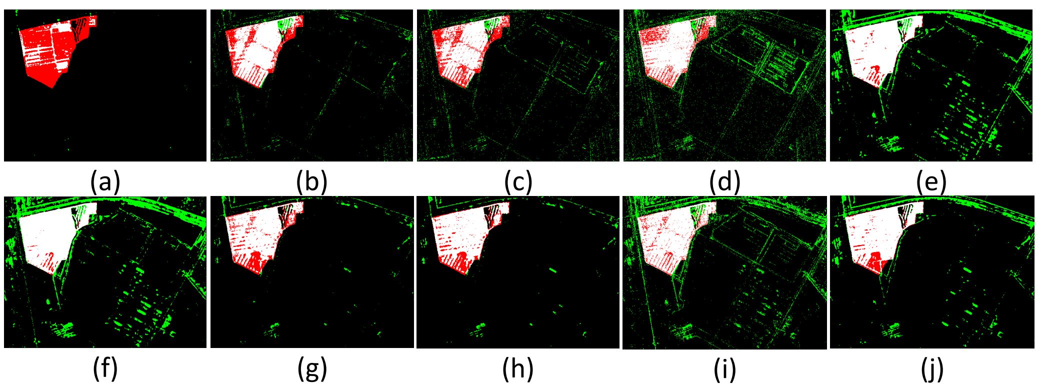

Figure B7 Change detection results of D3—Farmland in Shuguang achieved by (a) SCCN, (b) ACE-Net, (c) X-Net, (d) RFR, (e) Resnet, (f) MLP, (g) Semi-GAN(60%), (h) proposed(60%), (i) Semi-GAN(0.1%) and (j) proposed(0.1%). (TP: white; TN: black; FP: green; FN: red)

Table B3 Experimental results of D3—Farmland in Shuguang

D3	Methods	FP	FN	OA	Kappa	F1	AUC	Time cost
Unsuper-vised	SCCN	195	17452	0.922	0.327	0.351	0.788	684.1
	ACE-Net	3706	5773	0.958	0.753	0.776	0.967	6455.2
	X-Net	6096	7050	0.942	0.666	0.698	0.953	3093.4
Super-vised(60%)	RFR	16234	4727	0.907	0.575	0.626	0.940	49.4
	Resnet	19485	1369	0.908	0.618	0.667	0.949	189.3
	MLP	21577	801	0.901	0.606	0.657	0.945	84.9
Semi-supervised(60%)	Semi-GAN	2499	5211	0.966	0.796	0.815	0.877	565.9
	Proposed	2493	4224	0.970	0.826	0.843	0.900	517.5
Semi-supervised(0.1%)	Semi-GAN	17952	2566	0.908	0.608	0.657	0.898	375.9
	Proposed	9814	2852	0.944	0.723	0.754	0.912	308.4

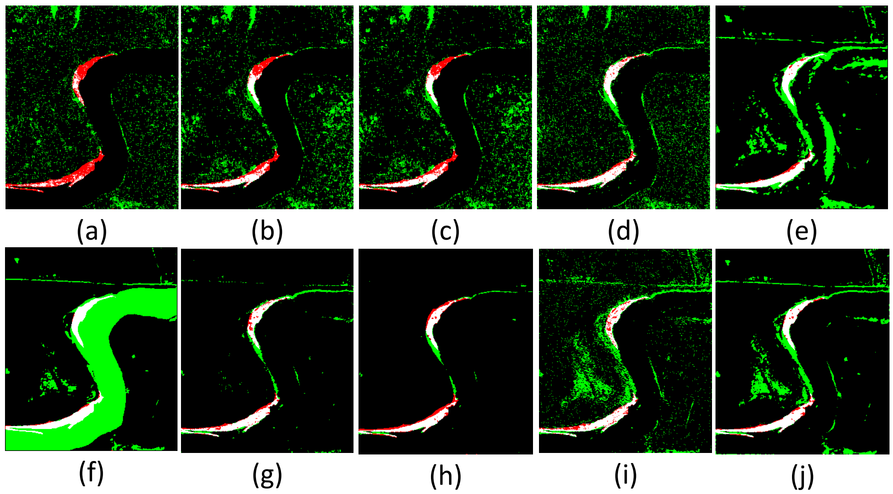

Figure B8 Change detection results of D4—Flood in Yellow River achieved by (a) SCCN, (b) ACE-Net, (c) X-Net, (d) RFR, (e) Resnet, (f) MLP, (g) Semi-GAN(60%), (h) proposed(60%), (i) Semi-GAN(0.1%) and (j) proposed(0.1%). (TP: white; TN: black; FP: green; FN: red)

Table B4 Experimental results of D4—Flood in Yellow River

D4	Methods	FP	FN	OA	Kappa	F1	AUC	Time cost
Unsupervised	SCCN	6958	2363	0.907	0.107	0.147	0.737	194.6
	ACE-Net	8628	1464	0.899	0.214	0.252	0.838	5911.6
	X-Net	5945	1393	0.927	0.294	0.326	0.868	3113.0
Supervised(60%)	RFR	8527	674	0.908	0.318	0.351	0.917	14.9
	Resnet	9080	328	0.906	0.343	0.376	0.947	87.5
	MLP	24269	162	0.755	0.149	0.197	0.867	34.2
Semi-supervised(60%)	Semi-GAN	1758	613	0.976	0.670	0.682	0.894	413.7
	Proposed	806	1027	0.981	0.691	0.700	0.834	350.2
Semi-supervised(0.1%)	Semi-GAN	8647	636	0.906	0.318	0.352	0.854	220.7
	Proposed	5891	465	0.936	0.433	0.459	0.896	181.1

Appendix B.5 Ablation Study

In this subsection, two strategies of the framework will be verified: one is the strategy of coarse-and-fine feature fusion (be expressed as proposed-1 in experimental display). The results of each classification layer will be compared with the result of all layers after fusion to prove the effectiveness of the cooperative judgment of coarse-and-fine features. Another one is the strategy of iterative pseudo label updating. The comparison between the initial network result and the iterative network result will prove the reliability of the pseudo label, which is conducive to improving the learning ability of identification information in network (be expressed as proposed-2 in experimental display).

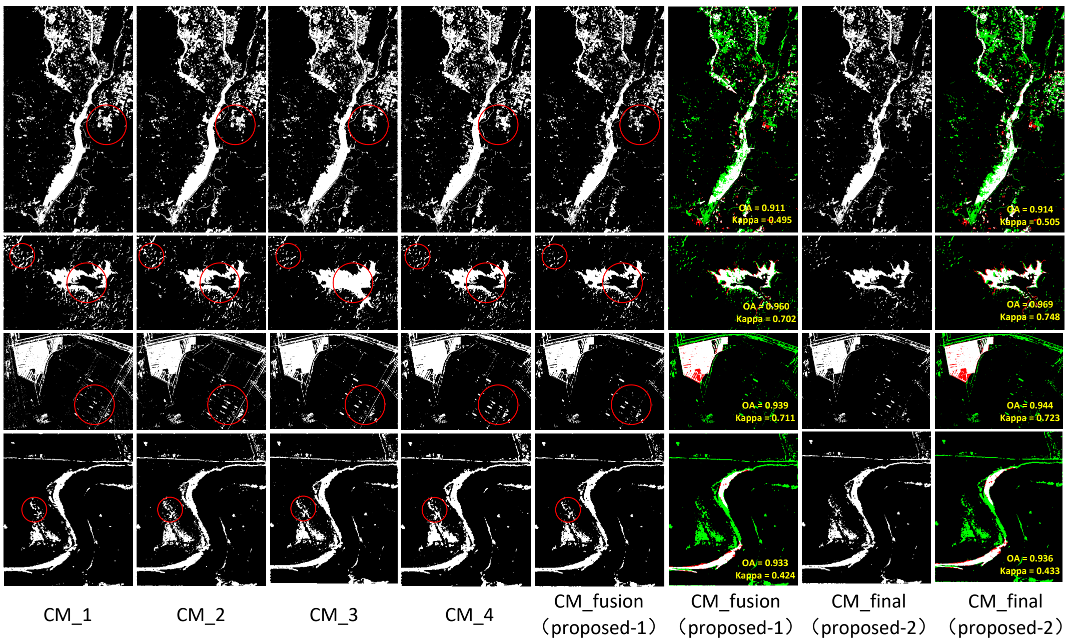

Figure B9 Display of results in ablation experiments. CM_i represents the result of each classification layer. CM_fusion is the result of all layers after fusion (proposed-1). And CM_final is the result after iterative pseudo label updating (proposed-2), which is compared with CM_fusion in the initial network result.

1) **Proposed-1 (coarse-and-fine feature fusion):** It can be seen from the red circle in the left part of Fig. B9 that CM_fusion can filter out some noise interference and improve the performance of change detection after fusion. This is particularly evident in the Italy dataset. For example, in CM_3, there is a serious over-fitting phenomenon, which causes all the hollowed out areas in the middle to be wrongly detected as changed areas. And this interference can be well eliminated by fusing with other change maps. In addition, the noised results in the upper left corner of Italy dataset also achieved a good suppression effect after fusion. The main reason is that the number of true labels is too small, which leads to over-learning of the characteristics of the changed areas in the coarse-and-fine feature extracting. Therefore, the fusion results not only retain the detection of changed areas, but also eliminate part of the error information through different over-fitting detection of multiple change maps.

2) **Proposed-2 (iterative pseudo label updating):** In the right part of Fig. B9, it shows the change detection results of the first iteration and the final iteration, respectively CM_fusion and CM_final. It can be seen that after iteration, the performance of change detection has improved to a certain extent, which is shown in the rise of OA and Kappa index. It is worth mentioning that such iteration can be repeated continuously to optimize the accuracy of the pseudo label pool. However, the experiments prove that after one iteration, the diffusion of the pseudo label pool is basically satisfied, and the iteration will stop. As can be seen from the proposed framework in experiments with 60% true-labeled samples, the semi-supervised method is far superior to the pure supervised method through the utilization of unlabeled samples.

3) **Network convergence analysis:** For network convergence, the experiment is carried on D2 dataset. Since the epochs of network training is 100, the results are displayed in the form of every 5 epochs. It can be seen from Fig. B10 that the loss of the network converges very quickly and finally remains in a stable state. And since the feedback training of the network has not been completed for the first 30 epochs, the output value of the change map is 0, which is not of reference significance. So the accuracy is set to 0. Finally, with the convergence of the network, the accuracy rate also reaches a gradually rising and stable state.

Appendix B.6 Limitation Analysis

In general, the proposed semi-supervised framework can make good use of the effective information of unlabeled samples to complete the screening of pseudo label samples, and then expand the pseudo label pool to optimize the process of network learning. However, the number of true-labeled samples required for the initial iteration will affect the overall performance of the framework to some extent. When the error change detection results are filtered into the pseudo label pool, it will lead to the classic error accumulation effect in the semi-supervised research field. Therefore, how to screen the available identifying information more reliably will become the key to our work in the future.

References

1 Zhou L F, Yu H W, Lan Y. Deep Convolutional Neural Network-Based Robust Phase Gradient Estimation for Two-Dimensional Phase Unwrapping Using SAR Interferograms. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58:4653-4665

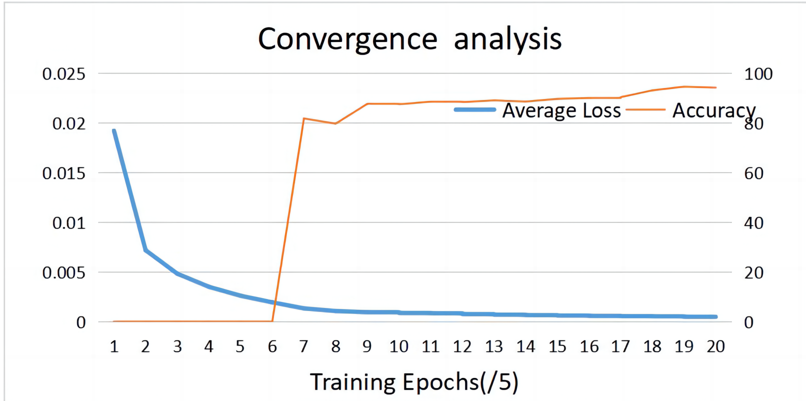

Figure B10 The network convergence of proposed method on D2 data set

2 Zhou L F, Yu H W, Lan Y, et al. Deep Learning-Based Branch-Cut Method for InSAR Two-Dimensional Phase Unwrapping. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60:1-15

3 Zhou L F, Yu H W, Lan Y, et al. CANet: An Unsupervised Deep Convolutional Neural Network for Efficient Cluster-Analysis-Based Multibaseline InSAR Phase Unwrapping. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60:1-15

4 Luppino L T, Bianchi F M, Moser G, et al. Unsupervised Image Regression for Heterogeneous Change Detection. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57:9960-9975

5 Niu X D, Gong M G, Zhan T, et al. A Conditional Adversarial Network for Change Detection in Heterogeneous Images. IEEE Geoscience and Remote Sensing Letters, 2019, 16:45-49

6 Wu Y, Li J, Yuan Y, et al. Commonality Autoencoder: Learning Common Features for Change Detection From Heterogeneous Images. IEEE Transactions on Neural Networks and Learning Systems, 2021, 99:1-14

7 Su L Z, Gong M G, Zhang P Z, et al. Deep learning and mapping based ternary change detection for information unbalanced images. Pattern Recognition, 2017, 66:213-228

8 Liu J, Gong M G, Qin K, et al. A Deep Convolutional Coupling Network for Change Detection Based on Heterogeneous Optical and Radar Images. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29:545-559

9 Liu Z G, Zhang Z W, Pan Q, et al. Unsupervised Change Detection From Heterogeneous Data Based on Image Translation. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60:1-13

10 Luppino L T, Hansen, M A, Kampffmeyer M, et al. Code-Aligned Autoencoders for Unsupervised Change Detection in Multimodal Remote Sensing Images. IEEE Transactions on Neural Networks and Learning Systems, 2022, 1-13

11 Luppino L T, Kampffmeyer M, Bianchi F M, et al. Deep Image Translation With an Affinity-Based Change Prior for Unsupervised Multimodal Change Detection. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60:1-22

12 Tuia D, Verrelst J, Alonso L, et al. Multioutput Support Vector Regression for Remote Sensing Biophysical Parameter Estimation. IEEE Geoscience and Remote Sensing Letters, 2011, 8:804-808

13 He K M, Zhang X Y, Ren S Q, et, al. Deep Residual Learning for Image Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 770-778

14 Shi J, Liu X D, Lei Y. SAR Images Change Detection Based on Self-Adaptive Network Architecture. IEEE Geoscience and Remote Sensing Letters, 2021, 18:1204-1208

15 Jiang F L, Gong M G, Zhan T, et al. A Semisupervised GAN-Based Multiple Change Detection Framework in Multi-Spectral Images. IEEE Geoscience and Remote Sensing Letters, 2020, 17:1223-1227