2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) | 978-1-6654-7069-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/WHISPERS56178.2022.9955130

SELF-SUPERVISED CONFIDENT LEARNING FOR HYPERSPECTRAL IMAGE CHANGE DETECTION

*Haonan Wu, Zhao Chen**

School of Computer Science and Technology, Donghua University, Shanghai 201620, China
(chenzhao@dhu.edu.cn)

ABSTRACT

Hyperspectral images (HSIs) contain rich spectral information required by detection of multiple ground changes. HSI change detection (CD) via deep learning attracts lots of attention in recent years for its advantages in deep feature representation and nonlinear problem modeling. However, unexpected changes can be hardly labelled beforehand, which limits the application of supervised models for HSI CD. In this paper, we propose a novel self-supervised framework based on confident learning (SSCL) for HSI CD. It mainly comprises two branches, binary classification and label correction. For starters, two unsupervised methods are combined to select credible samples as the initial training set. It is used to train a feature extractor in the classification branch while being corrected by the CL and prototype selection in the correction branch. These two branches are updated alternatively as the former share the feature extractor with the latter and the latter provides credible pseudo labels for the former. Experiments on real HSI datasets have shown that our method outperforms many state-of-the-art methods.

***Index Terms*—** Hyperspectral images, change detection, confident learning, credible labels, self-supervised

1. INTRODUCTION

HSI CD [1, 2, 3] aims at detecting change regions between multitemporal HSIs and is widely applied from ecological environment change studies to urban development trace. In recent years, deep learning attracts more and more attention in HSI CD for its powerful nonlinear expressive ability. However, labelled samples are limited in HSI CD for the hard and time-consuming annotation process. An alternative solution uses pseudo labels which can be generated automatically to replace true labels [4]. But the CD results rely highly on the accuracy of pseudo labels in the mentioned solution [5]. To address the above issue, a novel self-supervised framework based on confident learning (SSCL) is proposed in this paper to cope with HSI CD. It mainly consists of binary classification branch and label correction branch, while a sample selection strategy is designed to generate initial training set. The main contributions are threefold. 1) Two unsupervised methods are integrated to generate initial labels with high confidence and the change detector can be thus trained effectively. 2) The strategy of confident learning and feature prototype selection is improved for CD to correct wrong initial labels and further enhance its confidence. 3) A novel self-supervised mechanism is proposed to achieve high-precision CD where change features are used to correct pseudo labels and the feature extractor is optimized using initial and corrected labels simultaneously. Experiments demonstrate that the proposed method obtains better CD results than many state-of-the-art methods.

This work was funded in part by "Chenguang Program" supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission under Grant 18CG38, the National Natural Science Foundation of China under Grant 61702094 and the Science and Technology Commission of Shanghai Municipal under Grant 17YF1427400. *The corresponding author is Zhao Chen.

2. BACKGROUND KNOWLEDGE

2.1. Change vector analysis (CVA) and cosine angle distance (CAD)

CVA [2] and CAD [3] are two widely applied easy but efficient unsupervised methods in HSI CD. Euclidean distance and cosine angle distance between corresponding spectral vectors are used to evaluate the change magnitude of every pixel region respectively. Then the CD results can be obtained using unsupervised clustering methods like K-means.

2.2. Confident Learning

CL can be leveraged to identify the error of pseudo labels by estimating the joint distribution $\mathbf{Q}_{\mathbf{y},\widetilde{\mathbf{y}}} \in \mathbb{R}^{m\times m}$ between the pseudo ground truth $\mathbf{y}$ and real labels $\widetilde{\mathbf{y}}$, where m is the number of categories [6]. Firstly, the confident joint matrix $\mathbf{C}_{\mathbf{y},\widetilde{\mathbf{y}}} \in {}^{m\times m}$ is calculated according to the pseudo labels $\mathbf{y}$ and predicted probabilities $p(\mathbf{y}|\underline{\mathbf{X}},\theta)$ from deep learning models, where $\underline{\mathbf{X}}$ and θ are the corresponding sample tensor and parameter set of the model respectively. Then the joint distribution $\mathbf{Q}_{\mathbf{y},\widetilde{\mathbf{y}}}$ can be obtained by normalizing the confident joint matrix $\mathbf{C}_{\mathbf{y},\widetilde{\mathbf{y}}}$. Finally, errors of pseudo labels can

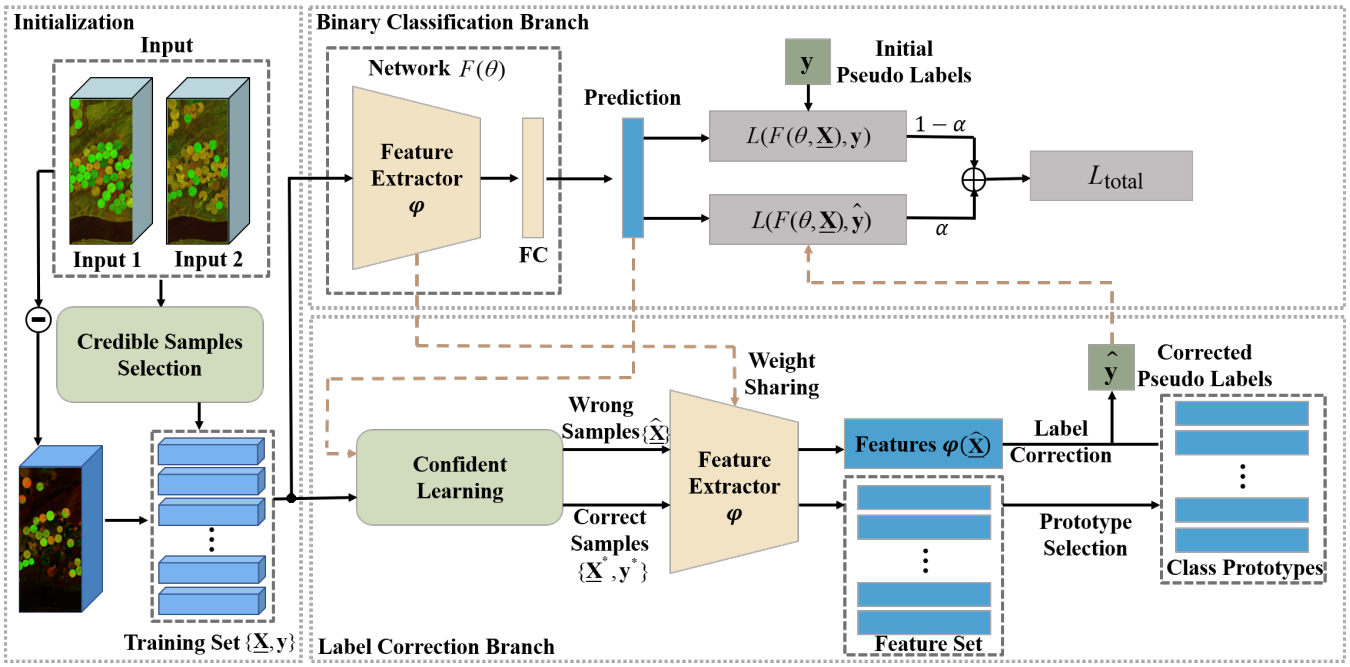

Fig. 1. The proposed framework of SSCL for HSI CD.

be identified from each off-diagonal entry in $\mathbf{Q}_{\mathbf{y},\tilde{\mathbf{y}}}$.

3. METHODOLOGY

The framework of the proposed method is illustrated in Fig. 1. Initial training set is generated in the initialization stage using two unsupervised methods. Then the binary classification branch and label correction branch are updated iteratively to obtain high-precision CD results. The details are as follows.

3.1. Initialization

Let $\underline{\mathbf{X}}_1, \underline{\mathbf{X}}_2 \in \mathbb{R}^{H\times W\times B}$ represent two HSIs acquired at different times, where H, W and B are the height, width and number of bands respectively. $\underline{\mathbf{X}}_D = |\underline{\mathbf{X}}_2 - \underline{\mathbf{X}}_1|$ is the absolute difference tensor. The details of the module credible samples selection in Fig. 1 are described here. To generate initial pseudo labels with high confidence, CVA and CAD are combined. $\mathbf{Y}_1, \mathbf{Y}_2 \in \mathbb{R}^{H\times W}$ represent the change maps generated by them respectively, where pixels with value 1 and value 0 represent the change and unchanged regions. Those pixels which obtain different labels in $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are labelled as uncertain and will not be used in the training phase. Then the initial training set $\{\underline{\mathbf{X}}, \mathbf{y}\}$ is consisted of patches from $\underline{\mathbf{X}}_D$ with the remaining pixels as the center point, where $\underline{\mathbf{X}} \in \mathbb{R}^{s\times s\times B\times N}$ is the tensor representing credible training samples, $\mathbf{y} \in \mathbb{R}^N$ is the initial pseudo labels vector, s and N represent window size of each patch and number of training samples respectively.

3.2. Binary Classification Branch

This branch aims at optimizing parameter θ of the binary classificaiton network F. Because there still exist some errors in the initial pseudo labels, corrected labels from the label correction branch are used as a complementary supervision. The total objective loss function is

$$L_{total} = (1-\alpha)\, L\left(F\left(\theta, \underline{\mathbf{X}}\right), \mathbf{y}\right) + \alpha L\left(F\left(\theta, \underline{\mathbf{X}}\right), \hat{\mathbf{y}}\right), \quad (1)$$

where L is the cross-entropy loss, $\hat{\mathbf{y}} \in \mathbb{R}^N$ is the corrected labels vector and α is the weight factor which controls the importance between the initial and corrected labels.

3.3. Label Correction Branch

This branch aims at correcting the remaining errors in initial pseudo labels to further enhance its confidence. Inspired by [7], single prototype may not be sufficient to represent a category (change or unchanged in our case). Therefore, multiple prototypes can be selected for each class here. Firstly, confident learning is leveraged to identify samples with wrong pseudo labels. Joint distribution $\mathbf{Q}_{\mathbf{y},\tilde{\mathbf{y}}} \in \mathbb{R}^{2\times 2}$ described in section 2.2 can be calculated via initial pseudo labels and the prediction results from training branch. Then the training set can be divided into correct samples $\{\underline{\mathbf{X}}^*, \mathbf{y}^*\}$ and wrong samples $\{\hat{\underline{\mathbf{X}}}\}$ whose labels are aborted. Secondly, the prototypes are selected from change feature set $\varphi\left(\underline{\mathbf{X}}^*\right)$ where $\varphi(\cdot)$ represents the feature extractor which shares the same weights with F except the last fully connected (FC) layer. Prototypes

which can represent one class should be as similar as possible to all samples in this class. Therefore, samples with the first p largest density in each category are selected as the prototypes where p is the number of prototypes. The density of the i-th sample in one class is defined as

$$\rho_i = \sum_{j=1}^{m} \operatorname{sign}(S_{ij} - c), \tag{2}$$

where m is the number of samples in this class, $\mathbf{S} \in \mathbb{R}^{m \times m}$ is the cosine similarity matrix with

$$S_{ij} = \frac{\varphi(\underline{\mathbf{X}}_i^*)^T \varphi(\underline{\mathbf{X}}_j^*)}{\|\varphi(\underline{\mathbf{X}}_i^*)\|_2 \|\varphi(\underline{\mathbf{X}}_j^*)\|_2}, \tag{3}$$

$\underline{\mathbf{X}}_i^*, \underline{\mathbf{X}}_j^* \in \mathbb{R}^{s \times s \times B}$ are the i-th and j-th sample in this class, c is a constant number given by the value of an element ranked top 40% in $\mathbf{S}$ and $\operatorname{sign}(\cdot)$ is the sign function[1]. Finally, corrected labels $\widehat{\mathbf{y}}$ can be obtained by comparing the average cosine similarity between $\varphi(\widehat{\underline{\mathbf{X}}})$ and the selected prototypes in each class. Since the prototypes are selected from correct samples rather than all samples and only wrong samples need to be corrected, the computation cost of selecting prototypes and correcting labels can be greatly reduced compared to [7]. Meanwhile, the selected prototypes are more representative as they are less likely to be mislabelled. Corrected labels $\widehat{\mathbf{y}}$ serve as a complementary supervision in the classification branch.

4. EXPERIMENT

4.1. Dataset

The CD performance of the proposed method is evaluated on two real-world bitemporal HSI datasets (EO-1 Hyperion images). Each dataset consists of two HSIs and a groud truth map as shown in Fig. 2. The bitemporal HSIs of the Hermiston dataset were acquired on May 1, 2004 and May 8, 2007 respectively [1]. Each HSI comprises 390×200 pixels and 242 spectral bands. And the bitemporal HSIs of the Yancheng dataset were acquired on May 3, 2006 and April 23, 2007 respectively [5]. Each HSI comprises 450×140 pixels and 155 spectral bands.

4.2. Setup

After the initial credible samples have been selected, 80% of them are randomly selected for training and the rest are used for validation. Network is VGG-16 pretrained on the ImageNet[2] concatenated with one flatten layer and three FC layers. Learning rate, batch size and window size s are set as 0.001, 256 and 3 respectively. The total number of epochs is set as 30 for all datasets. In the first 5 epochs, α is set as 0 and only initial labels are used to train F. After that, classification and label correction are updated iteratively in a self-learning manner, whereas α is set as 0.8. The number of prototypes p in each class is set as 4. SSCL is compared with two classical methods CVA [2] and CAD [3] and three state-of-the-art methods, DSFANet [8], HI-DRL [4], and SSTN [9]. All the algorithms are coded by Keras and implemented on 64GB RAM and NVIDIA GeForce RTX 2080 Ti.

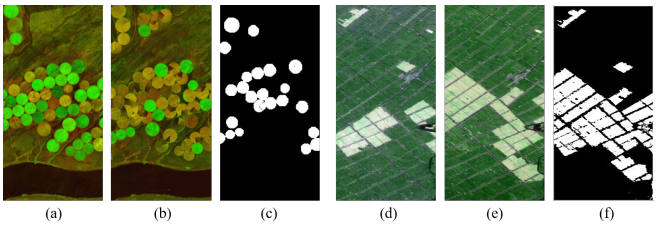

Fig. 2. Illustration of the Hermiston and Yancheng dataset. Hermiston HSI on (a) May 1, 2004, (b) May 8, 2007, and (c) Ground truth change map. Yancheng HSI on (d) May 3, 2006, (e) April 23, 2007, and (f) Ground truth change map.

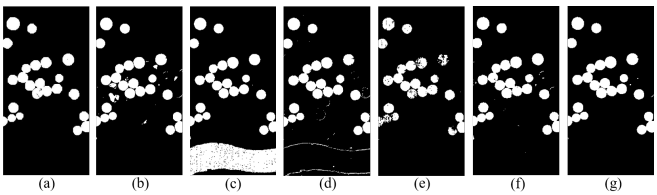

Fig. 3. CD maps of the Hermiston dataset produced by (a) Ground truth, (b) CVA, (c) CAD, (d) DSFANet, (e) HI-DRL, (f) SSTN, (g) The proposed.

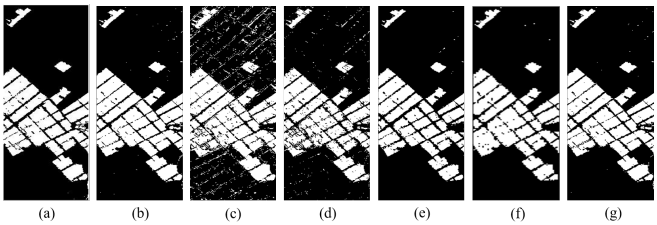

Fig. 4. CD maps of the Yancheng dataset produced by (a) Ground truth, (b) CVA, (c) CAD, (d) DSFANet, (e) HI-DRL, (f) SSTN, (g) The proposed.

4.3. Results

The efficacy of the proposed SSCL can be validated by the numerical data in Table 1. It shows that the proposed method outperforms other classical and state-of-the-art methods not only in overall accuracy (OA) but also in Kappa coefficient

[1] $\operatorname{sign}(x) = 1$ if $x > 0$; $\operatorname{sign}(x) = 0$ if $x = 0$; otherwise $\operatorname{sign}(x) = -1$.

[2] https://github.com/fchollet/deep-learning-models/releases/download/v0.1/vgg16_weights_tf_dim_ordering_tf_kernels_notop.h5

Table 1. Evaluation of HSI change detection results.

Dataset	Hermiston		Yancheng	
	OA(%)	Kappa	OA(%)	Kappa
CVA	97.81	0.9036	97.44	0.9379
CAD	84.34	0.5112	91.35	0.7996
DSFANet	97.81	0.9018	95.71	0.8942
HI-DRL	97.72	0.8910	97.09	0.9270
SSTN	98.63	0.9380	97.64	0.9420
Ours	**98.79**	**0.9459**	**98.17**	**0.9553**

Table 2. Ablation study of weight factor α on Hermiston dataset.

α	0	0.2	0.4	0.6	0.8	1
OA(%)	98.65	98.67	98.69	98.69	**98.79**	98.66
Kappa	0.9389	0.9399	0.9414	0.9408	**0.9459**	0.9400

(Kappa). It demonstrates that our network is supervised by pseudo labels with high enough confidence thanks to credible samples selection and label correction. The visual CD maps of Hermiston and Yancheng datasets shown in Fig. 3 and Fig. 4 also demonstrate the efficacy of the proposed method.

To validate the ability of correcting pseudo label errors in label correction branch, the ablation study is conducted and the experiment results on Hermiston dataset are shown in Table 2. The worst CD result is obtained when label correction branch is aborted ($\alpha = 0$ in our case) which validates the necessity of label correction. However, it shows that using corrected pseudo labels alone cannot obtain the best CD result, which is obtained when the initial and corrected pseudo labels are leveraged at the same time.

5. CONCLUSION

This paper proposes a novel self-supervised framework based on confident learning (SSCL) to cope with HSI CD. Initial pseudo labels with high confidence can be generated by combining two unsupervised methods which are further corrected using confident learning and prototype selection strategy. Then the classification network is supervised by initial and corrected labels simultaneously so that more effective deep change feature can be extracted. Experiment results demonstrate that the proposed method outperform several state-of-the-art methods for HSI CD.

6. REFERENCES

[1] Lorenzo Bruzzone, Sicong Liu, Francesca Bovolo, and Peijun Du, "Change detection in multitemporal hyperspectral images," in *Multitemporal Remote Sensing*, pp. 63–88. Springer, 2016.

[2] Francesca Bovolo and Lorenzo Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 1, pp. 218–236, 2006.

[3] Hui Zhang, Maoguo Gong, Puzhao Zhang, Linzhi Su, and Jiao Shi, "Feature-level change detection using deep representation and feature change analysis for multispectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 11, pp. 1666–1670, 2016.

[4] Puzhao Zhang, Maoguo Gong, Hui Zhang, Jia Liu, and Yifang Ban, "Unsupervised difference representation learning for detecting multiple types of changes in multitemporal remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2277–2289, 2018.

[5] Qiuxia Li, Hang Gong, Haishan Dai, Chunlai Li, Zhiping He, Wenjing Wang, Yusen Feng, Feng Han, Abudusalamu Tuniyazi, Haoyang Li, et al., "Unsupervised hyperspectral image change detection via deep learning self-generated credible labels," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 9012–9024, 2021.

[6] Curtis Northcutt, Lu Jiang, and Isaac Chuang, "Confident learning: Estimating uncertainty in dataset labels," *Journal of Artificial Intelligence Research*, vol. 70, pp. 1373–1411, 2021.

[7] Jiangfan Han, Ping Luo, and Xiaogang Wang, "Deep self-learning from noisy labels," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5138–5147.

[8] Bo Du, Lixiang Ru, Chen Wu, and Liangpei Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9976–9992, 2019.

[9] Feng Zhou and Zhao Chen, "Hyperspectral image change detection by self-supervised tensor network," in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020, pp. 2527–2530.