Automatic SAR Change Detection Based on Visual Saliency and Multi-Hierarchical Fuzzy Clustering

Yao Peng , *Member, IEEE*, Bin Cui , Hujun Yin , *Senior Member, IEEE*, Yonghong Zhang , and Peijun Du , *Senior Member, IEEE*

Abstract—Change detection based on multi-temporal synthetic aperture radar (SAR) images plays a significant role in environmental and earth observations. With the advancement in deep neural networks, existing work in the literature mainly concentrates on developing self-supervised methods to generate pseudo-labeled samples to guide the subsequent deep learning based detection. However, this way of selecting sample inevitably introduces erroneous labels and imbalance between unchanged and changed classes, thus causing deterioration in change detection performance. To mitigate these issues, we have proposed a SAR change detection network based on visual saliency and multi-hierarchical fuzzy clustering. Specifically, with multi-dimensional difference feature representations, a visual saliency based difference map is constructed for accurate difference feature extraction. By integrating neighborhood information and hierarchical clustering, the multi-hierarchical fuzzy local information C-means clustering algorithm has been developed to identify potential changed regions for sample selection. A class-balanced adaptive focal loss has further been incorporated into the network training to obtain accurate predictions. Extensive experiments and comparisons on five datasets have been performed. The proposed method has achieved averaged accuracy of 99.07% and Kappa coefficient of 79.87%, outperforming other state-of-the-art algorithms both visually and quantitatively.

Index Terms—Change detection, classed-balanced adaptive focal loss, fuzzy clustering, synthetic aperture radar (SAR), visual saliency difference map.

Yao Peng is with the School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: pengyao0130@njupt.edu.cn).

Bin Cui is with the School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China, and also with the School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: bincui@njupt.edu.cn).

Hujun Yin is with the Department of Electrical and Electronic Engineering, University of Manchester, M13 9PL Manchester, U.K. (e-mail: hujun.yin@manchester.ac.uk).

Yonghong Zhang is with the Chinese Academy of Surveying and Mapping, Beijing 100830, China (e-mail: yhzhang@casm.ac.cn).

Peijun Du is with the School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China (e-mail: peijun@nju.edu.cn).

I. INTRODUCTION

CHANGE detection aims at identifying changes across multi-temporal remote sensing images of the same scene [1]. Recent years have witnessed rapid advances in applying multimodal images in change detection, including optical, hyperspectral, and synthetic aperture radar (SAR) images. Of these, SAR images are capable of reflecting backscattering ground surface information regardless of weather and sunlight conditions, making them the only available source in extreme scenarios such as cloud, smoke, haze or other occurring disasters [2], [3]. Additionally, different types of objects in SAR images often exhibit distinct characteristics that can be employed to assist subsequent interpretation. Therefore, research and applications have grown considerably over the last decades, from flood detection [4], [5], postdisaster assessment [6], [7], urban planning [8], [9] to ecological surveillance [10], [11], [12].

A large number of supervised and unsupervised methods have been developed for SAR change detection. Supervised methods often produce superior performance but highly rely on prior knowledge and labelled samples [13], [14]. On the contrary, unsupervised and self-supervising methods can overcome the dependence of manual intervention, effectively remove or reduce the label requirement, and hence have recently become the mainstream approach. Conventional unsupervised SAR change detection methods generally consist of three steps: 1) image preprocessing, 2) difference image generation, and 3) difference image analysis to determine changed pixel (CP) and unchanged pixel (UP) [15], [16], [17]. Most research targets the latter two steps.

Difference image generation aims to calculate the differences between pairs of SAR images while suppressing inherent speckle noise, so to highlight potential changed areas. Typical methods include log-ratio operator [18], Gauss-log ratio operator [19], neighborhood-based ratio operator [20], and difference fusion methods [21], [22], [23]. Although still applied, the existing difference image generation algorithms often suffer from poor ability of accurately highlighting potential change areas. The purpose of difference image analysis is to determine meaningful changed areas and can be considered as an image segmentation process. Thresholding and clustering are generally employed. Apart from the classical Ostu algorithm [24], there are several thresholding algorithms based on the combination of specific distributions and fitting criteria [15], [25], [26]. In comparison with the thresholding algorithms, clustering has
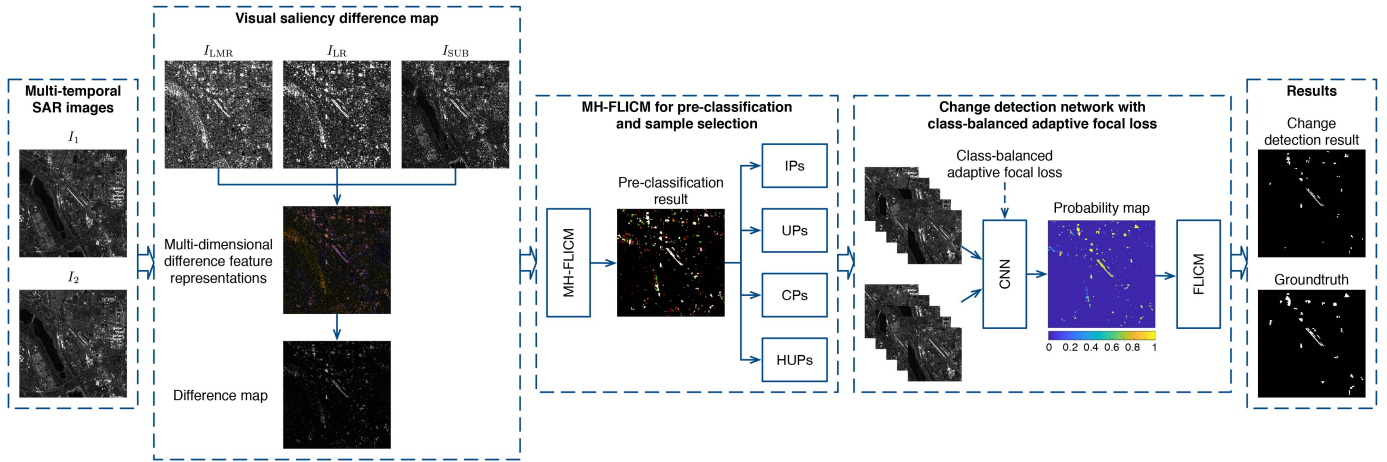
Fig. 1. Framework of the proposed method. Three key steps are: the construction of visual saliency difference map, the MH-FLICM for pre-classification and sample selection, and the change detection network with class-balanced adaptive focal loss.

better generalization ability and is usually more robust to noise. Examples include the fuzzy C-means (FCM) algorithm, fuzzy local information C-means clustering (FLICM) [27], Markov random field fuzzy C-means clustering [28], and reformulated fuzzy local information C-means clustering [29].

Recently, deep neural networks have achieved remarkable successes in remote sensing image analysis, showing superior performances in SAR image change detection. Self-learning strategies are commonly adopted to obtain pseudolabels from unlabeled difference maps. In this direction, many researchers have devoted to exploiting effective classifiers with better performance than the conventional approaches. Gong et al. presented a deep neural network for accomplishing change detection directly from pairs of SAR images [30]. Geng et al. developed a saliency-guided SAR change detection approach based on hierarchical fuzzy C-means (HFCM) and autoencoders [23]. With transfer learning, a multilevel fusion network was introduced for sea ice change detection [31]. In [32], generative adversarial learning was incorporated to classify hard pixels for small area change detection. Saha et al. employed cycle-consistent generative adversarial networks to transcode SAR images into optical images to aid change detection [9]. Despite the remarkable successes achieved, low sample accuracy, and imbalanced class samples remain two common issues in deep learning based SAR change detection.

Building a robust deep model for SAR change detection still poses a great challenge due to the following three issues: 1) inaccurate difference feature representations, 2) insufficient ability of selecting high-quality samples, and 3) imbalanced class samples. To address these issues, we have proposed a robust SAR change detection framework, shown in Fig. 1, based on visual saliency and multi-hierarchical fuzzy clustering. In the proposed method, the visual saliency difference map is first constructed to give prominence to changed areas, and an unsupervised fuzzy clustering algorithm is then developed to generate accurate pseudo-labeled samples. We further incorporate class-balanced adaptive focal loss into the convolutional neural network (CNN) to derive final change detection results. Main contributions of the current work are summarized as follows.

1) In order to enhance the degree of difference information, a method for generating visual saliency difference map has been proposed based on multi-dimensional difference feature representations. Through the difference feature fusion, the proposed method for difference map generation is capable of suppressing false alarms and noise while highlighting real changed regions and providing stronger visual difference contrast.

2) Inspired by hierarchical clustering, we have developed a multi-hierarchical fuzzy local information C-means clustering (MH-FLICM) algorithm for pre-classification and sample selection. The integration of neighborhood information and hierarchical clustering can precisely identify potential CPs, alleviate the imbalance between changed and unchanged classes, as well as select valid sample for subsequent change detection.

3) A change detection network with class-balanced adaptive focal loss has been constructed for generating probability maps to further tackle the sample imbalance problem. FLICM algorithm is then utilized as the probabilistic binary classifier to obtain effective and reliable change detection result.

The remainder of this article is organized as follows. Section II introduces the construction of visual saliency difference map. Section III presents a detailed illustration of the proposed MH-FLICM algorithm. Architecture of the change detection network is described in Section IV. Section V presents experimental results, together with the parameter analysis on several aspects. Section VI concludes this article.

II. CONSTRUCTION OF VISUAL SALIENCY DIFFERENCE MAP

A. Multi-Dimensional Difference Feature Representations

Conventional methods often use one-dimensional difference map to reflect difference information in multi-temporal images between corresponding pixels or neighbors. Although this way of constructing a difference map can be handcrafted with good responses for certain types of changes, it may lose or weaken

general difference information, making it less capable of representing the global differences. To tackle this issue, we construct a visual saliency difference map based on multi-dimensional difference feature representations. Single-dimensional difference maps are concatenated across the dimensions, each reflecting a certain type of difference, and their combination shows superior responses to changes.

In order to moderate the influence of noise in constructing pixel-based difference maps, we calculate the log-mean ratios (LMR) of a neighborhood, which shows tolerance to isolated noise. The larger the difference, the higher the grey level values in difference map. LMR is calculated by

$$I_{\text{LMR}}(p) = \log \left\{ \max \left[\frac{\mu(\Omega_{\Gamma_p})}{\mu(\Omega_{\Upsilon_p})}, \frac{\mu(\Omega_{\Upsilon_p})}{\mu(\Omega_{\Gamma_p})} \right] \right\} \quad (1)$$

where Γ and Υ, respectively, denote the before and after images, Ω_Γ and Ω_Υ denote the neighborhood of pixels in Γ and Υ, respectively. $\mu(\Omega_\Gamma)$ and $\mu(\Omega_\Upsilon)$ are the mean values of the neighborhoods of corresponding pixels, and p represents the pixel position in the two images.

As the most widely applied change detection algorithm, log-ratio (LR) extracts good representations in reflecting differences of corresponding pixels in multi-temporal images. Regions of large grey level values in difference maps correspond to areas with large differences between images taken before and after. LR can be expressed as

$$I_{\text{LR}}(p) = \log \left(\max \left(\frac{\Gamma_p}{\Upsilon_p}, \frac{\Upsilon_p}{\Gamma_p} \right) \right). \quad (2)$$

Although being shown not suitable for overcoming the inherent multiplicative noise in SAR images, the earliest change detection algorithm, the absolute subtraction (SUB), intensely suppresses noise under certain circumstances. For example, low scattering regions erroneously display high differences due to the ratio operation. SUB takes the form

$$I_{\text{SUB}}(p) = \text{abs}\,(\Gamma_p - \Upsilon_p). \quad (3)$$

To unify these three algorithms into similar scale and magnitude, min–max regularisation is processed for I_{LMR}, I_{LR}, and I_{SUB}, respectively. The multi-dimensional difference feature representation, Ψ, is then derived through merging the three difference images across separate channels.

In all the three difference images, the larger the value, the greater the difference, consistent in difference feature responses. Therefore, significant differences can be better observed in changed areas, showing improved robustness than any single-dimensional difference images. On the contrary, insignificant differences, which indicate unchanged areas, will incur weaker difference responses. The multi-dimensional difference map also has a good tolerance in regions with opposite directions of difference indications, thus avoiding erroneous representations generated by a single-dimensional difference map. In summary multi-dimensional difference map has strong robustness and reflection ability, and can enhance the contrast between potential CP and UP for identifying potential changes. Fig. 2(d) shows multi-dimensional difference feature representations for five different datasets, in which regions in white denote significant, potential changed areas, regions in green and purple indicate areas of notably potential changes in certain dimensions of the difference map, and black regions represent negligible changes. Compared to conventional single-dimensional difference map, the multi-dimensional representation holds stronger capability to convey the difference information visually and truthfully.

B. Visual Saliency Difference Map Generation

In the previous section, a multi-dimensional difference map has been constructed to expand the degree of difference between potential CP and UP. However, it is still necessary to determine an optimised selection method to digitally express the "visual difference" so that the difference map has autonomous ability to focus on areas of potential changes. As an important direction in computer vision, visual saliency detection helps to automatically identify the region of interest in an image by simulating the human visual mechanism. For example, when facing a new image scene, human vision automatically pays attention to certain regions of interest and ignores uninterested regions. These detected regions of interest are regarded as visually salient regions [33].

As a classical and effective saliency detection algorithm, the luminance contrast (LC) [34] algorithm aims to obtain the global contrast of single pixels in the entire image. That is, the saliency value of a single pixel is the sum of grey level distances to all other pixels. The calculation focuses on seeking the global contrast but can cause rare pixels to dominate with higher contrast. In multi-dimensional difference map, potential CPs hold a small portion of the image and are considered as rare pixels. This feature can cater for the demand for saliency region extraction [35]. Therefore, we use the LC algorithm to extract visual saliency difference map by highlighting potential change pixels in the multi-dimensional difference map Ψ, so that the main "attention" of the difference map is focused on the potential change pixels. The LC-based saliency can be computed by

$$\Phi_k = \sum_{i=1}^{N} \|\Psi_k - \Psi_i\| \quad (4)$$

where Φ_k is the saliency value of pixel t, $\|\Psi_k - \Psi_i\|$ stands for the Euclidean distance between Ψ_k and Ψ_i, and N denotes the total number of pixels in the image.

Considering that pixels with the same intensity level have the same saliency, we further restructure (4) so that pixels with the same intensity are rearranged together as

$$\Phi_k = \sum_{j=1}^{n_L} f_j \|\Psi_k - L_j\| \quad (5)$$

where L_j represents jth intensity level in Ψ, f_j denotes the frequency of L_j, and n_L is the total number of intensity levels in the image.

Through the construction of highly responsive multi-dimensional feature representation and the application of global visual saliency extraction algorithm, areas with significant
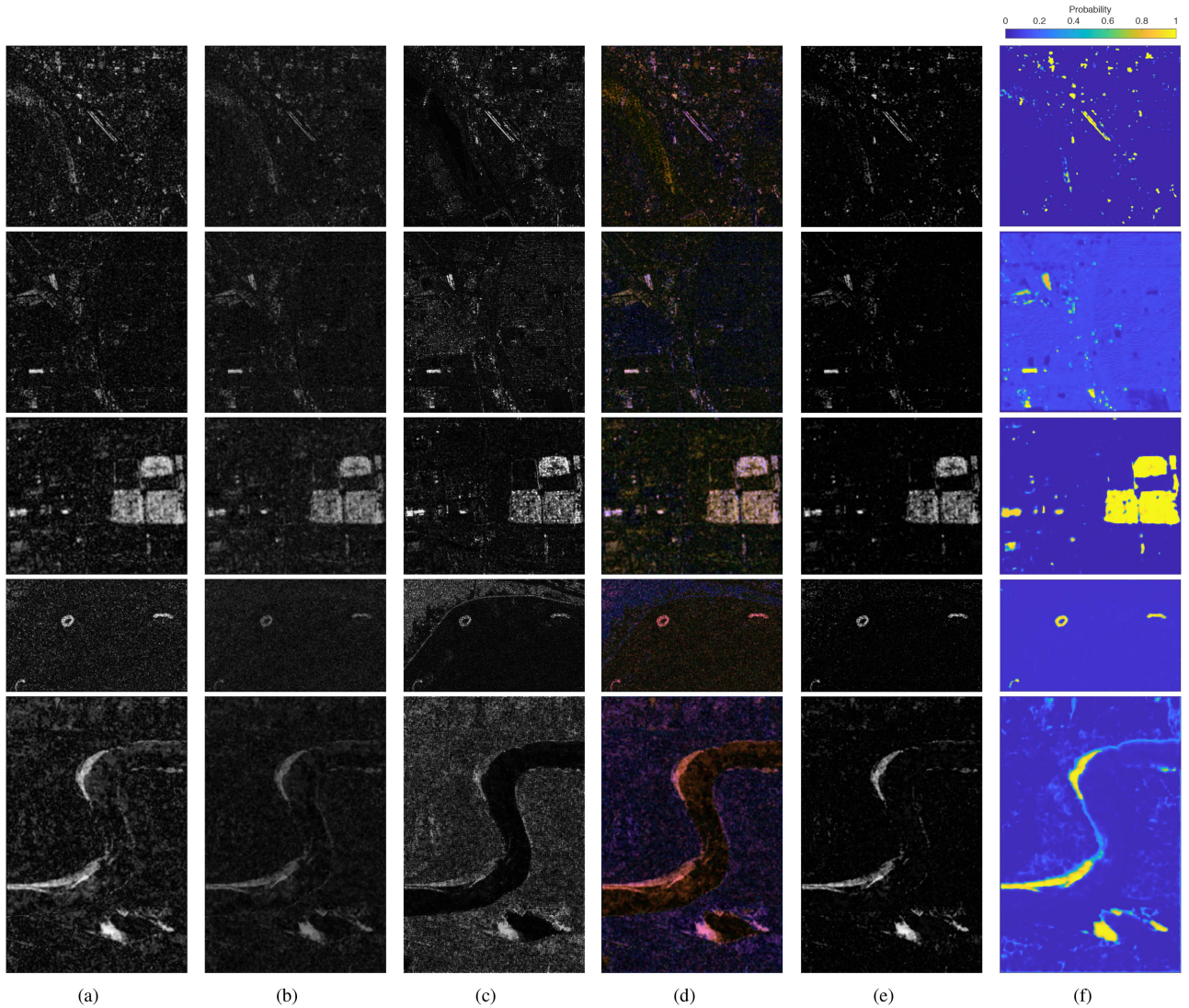
Fig. 2. Step-wise results of the proposed method on Beijing-I, Beijing-II, Beijing-III, Coast and Yellow River datasets: (a) LMR difference image, (b) LR difference image, (c) SUB difference image, (d) multi-dimensional difference feature representations, (e) visual saliency difference maps, and (f) predicted probability maps.

changes in the difference map are targeted and highlighted. This strategy suppresses insignificant areas and expands the degree of difference between different types of pixels, providing a basis for selecting as many as possible potential change pixels with high-confidence in the subsequent process. Fig. 2(e) shows a constructed difference map based on the extraction of multi-dimensional difference feature saliency. It can be seen that potential changed areas possess stronger visual contrast and show stronger resistance to noise.

III. MULTI-HIERARCHICAL FUZZY LOCAL INFORMATION C-MEANS CLUSTERING

Coherent speckle noise inherently exists in SAR images and can greatly affect SAR image processing and interpretation. Incorporating neighborhood information is the most efficient and effective way to reduce such noise in SAR images. In this section, we develop an unsupervised multi-hierarchical clustering algorithm, namely MH-FLICM, which takes into account

the neighbourhood information in hierarchically segmenting the difference map. The proposed MH-FLICM algorithm identifies potential change areas as accurately as possible, and at same time optimizes the distributions of pixel classes in the pre-classification result for selecting pixels from valid samples.

As an improvement of the classic FCM algorithm, the FLICM segmentation algorithm incorporates local neighborhood and grey level information to achieve stronger resistance to speckle noise [27]. Specifically, the objective function of the FLICM algorithm is defined as

$$J_w = \sum_{i=1}^{N} \sum_{m=1}^{M} \left[u_{mi}^w \|x_i - v_m\|^2 + G_{mi} \right] \quad (6)$$

where x_i is the gray level value of pixel i, $w = 2$ is the weighting parameter of each fuzzy membership, N denotes the total number of pixels in the image, M is the desired number of classes in the final classification, u_{mi} is the degree of membership of ith pixel in mth cluster, v_m denotes the center of m-th cluster

prototype, $\| \cdot \|$ represents the Euclidean distance, and the fuzzy factor G_{mi} is defined as the weighted sum of distances of the cluster centre from its spatial neighbors by

$$G_{mi} = \sum_{j \in \Omega_i, i \neq j} \frac{1}{d_{ij} + 1} (1 - u_{mj})^w \|x_j - v_m\|^2 \quad (7)$$

where Ω_i is the neighborhood of pixel i, d_{ij} denotes the spatial distance between pixels i and j, and u_{mj} and v_m can be computed by

$$u_{mj} = \frac{1}{\sum_{j=1}^{M} \left(\frac{\|x_i - v_m\|^2 + G_{mi}}{\|x_i - v_j\|^2 + G_{ji}} \right)^{\frac{1}{w-1}}} \quad (8)$$

$$v_m = \frac{\sum_{i=1}^{N} u_{mi}^w x_i}{\sum_{i=1}^{N} u_{mi}^w}. \quad (9)$$

When consecutive update of the membership matrix becomes smaller than the preset threshold, the iteration stops. Therefore, pixels will be attributed to the categories with the highest degrees of membership, and the initial segmentation is completed.

Different from FCM, FLICM algorithm generates fewer isolated pixels and more attention is paid to changed areas in the difference map, producing smoother effect in the spatial domain. The incorporation of neighborhood information often results in blurring of the boundary between changed and unchanged areas, in which pixels may correspond to intermediate category and are most likely to be misclassified in conventional difference image analysis. Therefore, the MH-FLICM algorithm is proposed to accurately detect the boundary regions (pixels belong to the intermediate category), as well as precisely identify changed and unchanged areas of high confidence. The category of pixels in boundary regions will later be determined by the network well trained using those selected changed and unchanged samples. Details of the MH-FLICM algorithm are described as follows.

1) Use FLICM to divide the difference map into two categories: changed C_C and unchanged C_U. N_C represents the total number of pixels in class C_C, and also assumed to be the upper limit of number of CPs in MH-FLICM algorithm. The upper limit of ratio of the number of intermediate pixels (IPs) to the number of CPs is set to be $T = 2$.

2) Use multiclass FLICM to segment the difference map into seven classes, arranged as $C_1, C_2, \ldots, C_7$, according to descending sort of cluster centres $\xi_1, \xi_2, \ldots, \xi_7$. We noticed that if directly dividing pixels into three classes by FLICM, the intermediate class sometimes occupies a significant part of the difference image, therefore enough representative samples cannot be obtained. The numbers of corresponding pixels in each class are $N_1, N_2, \ldots, N_7$, respectively.

3) The number of CPs is

$$N_{CP} = \begin{cases} \sum_{i=1}^{7} N_i & \text{if } N_{CP} \leqslant N_C \\ N_1 & \text{if } N_1 + N_2 > N_C \end{cases}. \quad (10)$$

The number of classes in CPs is t_C.

4) The number of IPs is

$$N_{IP} = \begin{cases} \sum_{j=1}^{7-t_C} N_{t_C+j} & \text{if } N_C < N_{IP} < N_C \times T \\ N_{t_C+1} & \text{if } N_{t_C+1} + N_{t_C+2} > N_C \times T \end{cases}. \quad (11)$$

The number of classes in IPs is t_I.

5) The number of UPs is

$$N_{UP} = \sum_{k=1}^{7-t_C-t_I} N_{t_C+t_I+k} \quad (12)$$

where $k = \arg\min_k \sum_{k=1}^{7-t_C-t_I} N_{t_C+t_I+k} > N_{CP}$. The number of classes in UPs is t_U.

6) Determine high-confidence unchanged pixels (HUPs). If $t_C + t_I + t_U < 7$, all pixels that have not been assigned labels will be taken as HUPs; if $t_C + t_I + t_U = 7$, HUPs will be merged into UPs.

IV. CHANGE DETECTION NETWORK WITH CLASS-BALANCED ADAPTIVE FOCAL LOSS

To obtain suitable inputs for subsequent detection, two-channel patch pairs are extracted from multi-temporal SAR images by collecting overlapping patches of sizes of $n \times n \times 2$, each of which is assigned a label according to the class of the center pixel in the segmented patch [36]. Dual-channel patches centered at UPs and CPs are considered as samples for subsequent CNN training, and patches centred at HUPs are discarded. This way of constructing inputs for training CNN helps identify CPs with high confidence by decreasing the large gap between the numbers of CPs and UPs and at the same time enhancing the training efficiency and effectiveness.

The focal loss [37] was originally proposed to address the class imbalance problem in computer vision tasks. It extends the cross entropy loss by introducing a modulating term and a balanced factor, formulated by

$$\mathcal{L} = \begin{cases} -\alpha(1-\hat{y})^\gamma \log(\hat{y}) & y = 1 \\ -(1-\alpha)\hat{y}^\gamma \log(1-\hat{y}) & y = 0 \end{cases} \quad (13)$$

where y and $\hat{y}$ represent the class label and the output probability of the network, respectively. α is the balancing factor, and γ denotes the focusing parameter to reduce the impact of samples that are easy to be classified in the training and force the model to pay more attention to those difficult ones to train. Recently, several adaptive focal loss algorithms have been developed through either modifying the focusing parameter γ using decay and scaling strategies, or adding fixed weights to the loss functions [38], [39], [40]. In this article, instead of adjusting the focusing parameter, the balancing factor α is adaptively formulated by the inverse rating of the advantageous and adverse samples obtained by the MH-FLICM algorithm. The class-balanced adaptive focal loss is developed as the objective function in the two-channel network to further handle the sample imbalance problem, where α is set as the ratio of N_{CP} to N_{UP} to adaptively adjust the weight of majority class and minority

class. The loss function then becomes

$$\mathcal{L} = \begin{cases} -\dfrac{N_{\mathrm{CP}}}{N_{\mathrm{UP}}}(1-\hat{y})^{\gamma}\log(\hat{y}) & y = 1 \\ -\left(1-\dfrac{N_{\mathrm{CP}}}{N_{\mathrm{UP}}}\right)\hat{y}^{\gamma}\log(1-\hat{y}) & y = 0 \end{cases}. \quad (14)$$

For samples centered at CPs, $y = 1$, the larger the $\hat{y}$, the smaller the loss. Similarly, for samples centered at UPs, $y = 0$, the smaller the $\hat{y}$, the smaller the loss, corresponding to samples that are easy to be determined. The calculation of the loss function for these samples will slow down the iterative process and make it difficult to reach the optimum.

In the literature, almost all the change detection methods adopt the classification mechanism of predicting only the patches centered at IPs, and acquiesce in the assumption that pre-classification result is 100% correct. To deal with these errors occurring in the pre-classification, in the proposed classification mechanism, the trained network reclassifies all the patches and outputs a probability map [as shown in Fig. 2(f)] with an equal size of the original image. In addition, considering the correlation between the adjacent pixels of the image, the FLICM clustering algorithm is further used as a probabilistic graph binary classifier to obtain the final change detection result.

V. EXPERIMENTS AND ANALYSIS

To validate the effectiveness of the proposed method, experiments were conducted on five datasets. Visual and quantitative comparisons with several state-of-the-art methods are reported, including the FCM algorithm based on LR difference map (LR-FCM), FLICM algorithm based on LR difference map (LR-FLICM), principal component analysis (PCA) and K-means algorithm (PCA-KM) [41], extreme learning machine based on neighborhood ratio (NR-ELM) [42], Gabor PCA network (GaborPCANet) [43], convolutional-wavelet neural networks (CWNN) [44], dual-domain network (DDNet) [45], siamese adaptive fusion network (SAFNet) [46], and robust unsupervised small area change detection method (RUSACD) [32]. A comprehensive investigation of several parameters on the performance of change detection is also presented.

A. Dataset Description and Evaluation Criteria

The original Beijing image, of sizes of approximately 13 000 × 22 000 pixels, were acquired from Gaofen-3 SAR in April 2017 and May 2018, respectively. Seasons for the two acquisition dates are similar, thus vegetation changes have little effect on the backscattering. Several preprocessing steps, such as calibration, registration, geocoding, and cropping were performed. After the following multilook processing of 2 × 3, the azimuth and distance resolution is approximately 7.8 m. Owing to the considerably large size of the entire image, three representative subregions were selected for evaluation, denoted as Beijing-I dataset, Beijing-II dataset, and Beijing-III dataset, respectively. Area A is the Beijing-I dataset of size of 400 × 400 pixels, in which urban construction and displacement of trains are the main changes. Fig. 3 shows exemplar multitemporal images and corresponding reference change map. Area
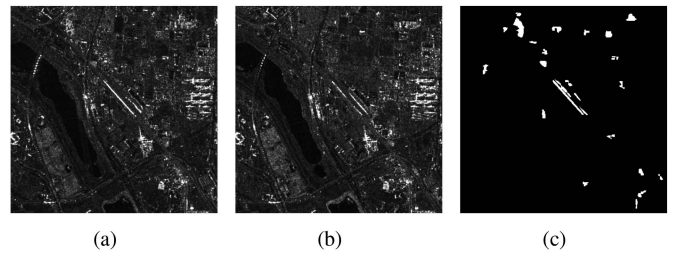

Fig. 3. Beijing-I dataset: (a) image acquired in April 2017, (b) image acquired in May 2018, and (c) reference image.
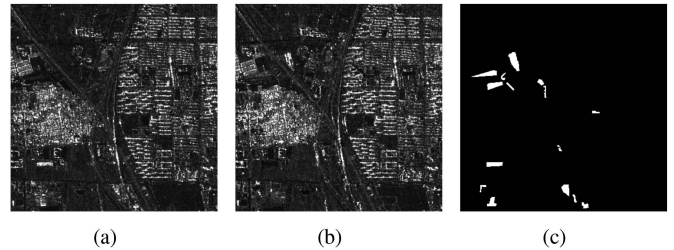

Fig. 4. Beijing-II dataset: (a) image acquired in April 2017, (b) image acquired in May 2018, and (c) reference image.
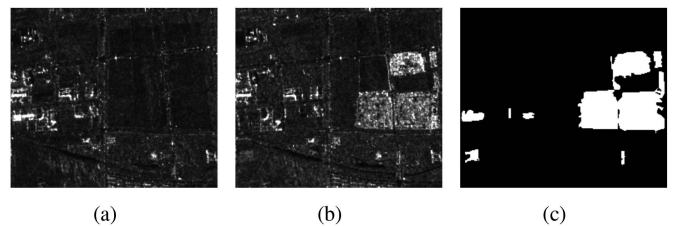
Fig. 5. Beijing-III dataset: (a) image acquired in April 2017, (b) image acquired in May 2018, and (c) reference image.
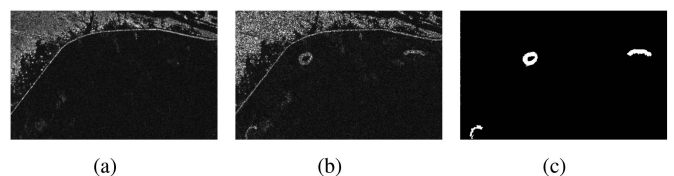
Fig. 6. Coast dataset: (a) image acquired in June 2008, (b) image acquired in June 2009, and (c) reference image.

B is the Beijing-II dataset of size of 400 × 400 pixels. Construction and demolition of artificial structures are the main cause of changes. Examples of Beijing-II are shown in Fig. 4. Area C, the Beijing-III dataset, has a size of 248 × 215 pixels, where changes are primarily caused by the update of construction as displayed in Fig. 5.

Apart from the above three datasets, the Coast and Yellow River datasets, derived from Radarsat-2 in the Yellow River Estuary area in June, 2008 and 2009, were also used. Multitemporal and the reference images in the Coast dataset, as shown in Fig. 6, are of 450 × 280 pixels with areas of relatively small changes. As shown in Fig. 7, the Yellow River dataset contains

(a)   (b)   (c)

Fig. 7. Yellow River dataset: (a) image acquired in June 2008, (b) image acquired in June 2009, and (c) reference image.
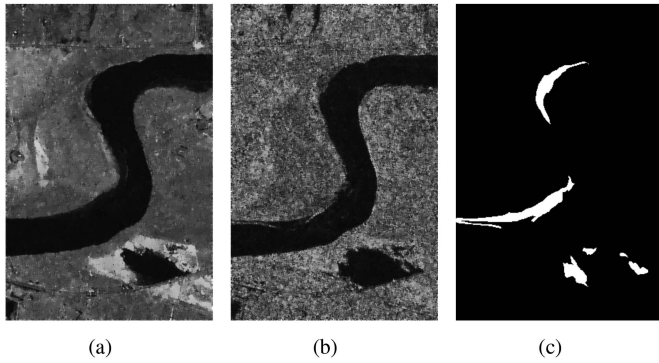
TABLE I
CHANGE DETECTION RESULTS OF VARIOUS METHODS ON BEIJING-I DATASET

Methods	FP	FN	OE	PCC(%)	κ(%)
LR-FCM	18816	576	19392	87.88	15.44
LR-FLICM	6160	560	6720	95.80	37.39
PCA-KM [41]	16192	192	16384	89.76	21.04
NR-ELM [42]	8368	224	8592	94.63	34.80
GaborPCANet [43]	17584	**64**	17648	88.97	20.60
CWNN [44]	30960	112	31072	80.58	11.63
DDNet [45]	16879	160	17029	89.36	20.57
RUSACD [32]	2600	800	3400	97.88	51.79
SAFNet [46]	12293	401	12694	92.07	24.45
Proposed (IPs only)	1075	796	1871	98.83	66.48
Proposed	**691**	905	**1596**	**99.00**	**68.74**

TABLE II
CHANGE DETECTION RESULTS OF VARIOUS METHODS ON BEIJING-II DATASET

Methods	FP	FN	OE	PCC(%)	κ(%)
LR-FCM	16416	368	16768	89.52	17.47
LR-FLICM	3680	448	4112	97.42	47.79
PCA-KM [41]	8048	192	8240	94.85	33.45
NR-ELM [42]	6208	208	6416	95.99	39.55
GaborPCANet [43]	6800	128	6928	95.67	38.32
CWNN [44]	21152	1072	21264	86.71	15.56
DDNet [45]	8608	154	8762	94.52	32.38
RUSACD [32]	317	**861**	**1178**	**99.26**	**72.19**
SAFNet [46]	10523	248	10771	93.27	26.86
Proposed (IPs only)	**184**	1081	1265	99.21	67.51
Proposed	385	867	1252	99.22	70.85

changes occurred mainly along the boundaries of the river with the image size of 291×444 pixels.

Several evaluation criteria were used for assessing accuracy of the detection including, false positives (FP), false negatives (FN), overall error (OE), percentage correct classification (PCC), and Kappa coefficient (κ). Among these, FP is the number of UPs that have been incorrectly detected as changed. FN denotes the number of CPs that have not been detected. OE represents the total number of pixels that were erroneously classified, and is the sum of FP and FN. Detection accuracy is the ratio between the number of pixels detected correctly and the total number of pixels, and κ is the Kappa coefficient.

B. Results on Beijing-I Dataset

Detecting changes in Beijing-I images has been of great difficulty due to complex types of ground objects and their various forms of change, especially in water areas. Visual results of various methods on Beijing-I dataset are shown in Fig. 8, while quantitative results presented in Table I. LR-FCM, LR-FLICM, PCA-KM, NR-ELM, GaborPCANet, CWNN, DDNet, and SAFNet all suffered severely from speckle noise, leading to significantly high FPs and low kappa coefficients. With the incorporation of neighborhood information, the detection result of LR-FLICM was much better than that of LR-FCM algorithm. Despite the noise interference was suppressed in the RUSACD algorithm, many important changes were wrongly classified, resulting in a relatively high FP value. Fig. 8(j) and (k) displays reprediction results of the proposed method on merely the

intermediate class and all the patches, respectively. Although the floating objects on the water and changes in the growing plants resulted in some false detections, it is visually apparent that almost all the change areas have been detected completely, demonstrating strong antinoise ability and high detection accuracy of the method. Meanwhile, it is worth noting that with the incorporation of subtraction operator in the multi-dimensional difference representation, the false alarm problem caused by the low scattering area in the water area has been effectively avoided.

C. Results on Beijing-II Dataset

In Beijing-II dataset, in addition to construction and demolition of buildings, there existed changes between bare land and vegetation. Compared with changes of buildings, differences between bare land and vegetation are less obvious and more scattered, increasing the difficulty of their accurate detection. Visual and quantitative results of Beijing-II dataset are presented in Fig. 9 and Table II, respectively. Reference image is given in Fig. 9(l). Similar to the Beijing-I dataset, CWNN algorithm was significantly affected by noise, producing large amount of false detection and trivial plaques. Detection results of CWNN algorithm is shown in Fig. 9(f) with the lowest κ of 15.56%. As shown in Fig. 9(b), LR-FLICM detected most of the change areas with some resistance to noise. Compared to the above two algorithms, there were fewer trivial plaques in the results of PCA-KM, CWNN, NR-ELM, GaborPCANet, DDNet, and SAFNet algorithms, shown in Fig. 9(c)–(e), with moderate detection rates. In contrast, the proposed method showed strong resistance to noise, though some missed detection occurred. On the quantitative evaluations in Table II, predicting all patches using the proposed algorithm obtained a slightly lower detection accuracy but a higher κ than that of only patches centred at IPs. Among all the methods, RUSACD exhibited the best performance with the highest accuracy and Kappa coefficients, probably due to the usage of generative models for increasing the training samples. It can also be observed that majority of the detection errors was located at the left side of the image due to low backscattering of varying vegetations. Such a small degree of difference may merely be perceptible to LR based algorithms, which are highly sensitive to difference information. Meanwhile, the utilization of neighborhood information in the change detection network help to suppress noise, but may cause false detections and missed detections in boundaries of
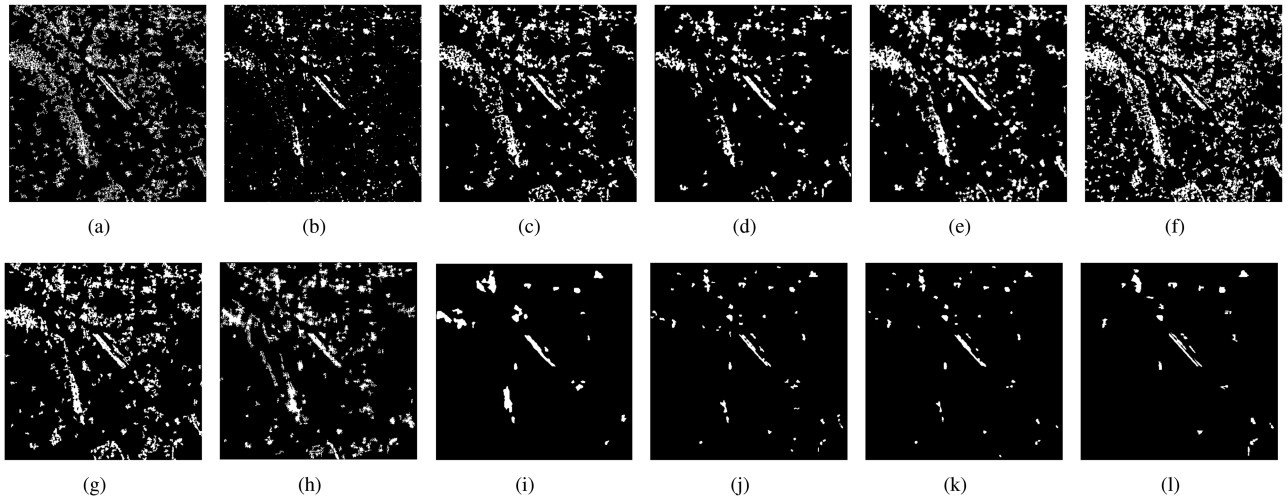
Fig. 8. Visualization of various change detection methods on Beijing-I dataset: (a) LR-FCM, (b) LR-FLICM, (c) PCA-KM, (d) NR-ELM, (e) GaborPCANet, (f) CWNN, (g) DDNet, (h) SAFNet, (i) RUSACD, (j) proposed (IPs only), (k) proposed, and (l) ground truth.

Fig. 9. Visualization of various change detection methods on Beijing-II dataset: (a) LR-FCM, (b) LR-FLICM, (c) PCA-KM, (d) NR-ELM, (e) GaborPCANet, (f) CWNN, (g) DDNet, (h) SAFNet, (i) RUSACD, (j) proposed (IPs only), (k) proposed, and (l) ground truth.

small change areas, leading to possible deterioration of change detection accuracy.

D. Results on Beijing-III Dataset

Fig. 10 and Table III, respectively, show the visual and quantitative results on the Beijing-III dataset. Compared with the previous two datasets, types of changes in Beijing-III area are less complex and more obvious, mainly caused by transitions between buildings and bare land. Therefore, all the algorithms except for LR-FCM achieved good detection results with accuracies over 85%. The reference image is shown in Fig. 10(l). Different from the previous results, on this dataset, the detection results of CWNN and NR-ELM were superior to those of LR-FLICM, PCA-KM, and GaborPCANet. Due to the uneven backscattering in the building area, many subtle change regions were eliminated as noise, leading to a large number of holes in the detection

TABLE III
CHANGE DETECTION RESULTS OF VARIOUS METHODS ON BEIJING-III DATASET

Methods	FP	FN	OE	PCC(%)	κ(%)
LR-FCM	469	1343	1812	96.60	83.15
LR-FLICM	261	1242	1503	97.18	85.87
PCA-KM [41]	192	1120	1312	97.54	87.72
NR-ELM [42]	459	538	10027	98.12	91.18
GaborPCANet [43]	501	608	1109	97.92	90.21
CWNN [44]	959	**219**	1178	97.79	90.16
DDNet [45]	619	343	962	98.20	91.72
RUSACD [32]	**106**	1015	1121	97.90	89.54
SAFNet [46]	1054	344	1398	97.38	88.30
Proposed (IPs only)	280	510	790	98.52	92.97
Proposed	275	488	**763**	**98.57**	**93.22**

results and high FN values in LR-FCM, LR-FLICM, PCA-KM, and RUSACD. It is also worth noting that CWNN, DDNet and SAFNet produced many false detections with high FP values,

Fig. 10. Visualization of various change detection methods on Beijing-III dataset: ((a) LR-FCM, (b) LR-FLICM, (c) PCA-KM, (d) NR-ELM, (e) GaborPCANet, (f) CWNN, (g) DDNet, (h) SAFNet, (i) RUSACD, (j) proposed (IPs only), (k) proposed, and (l) ground truth.
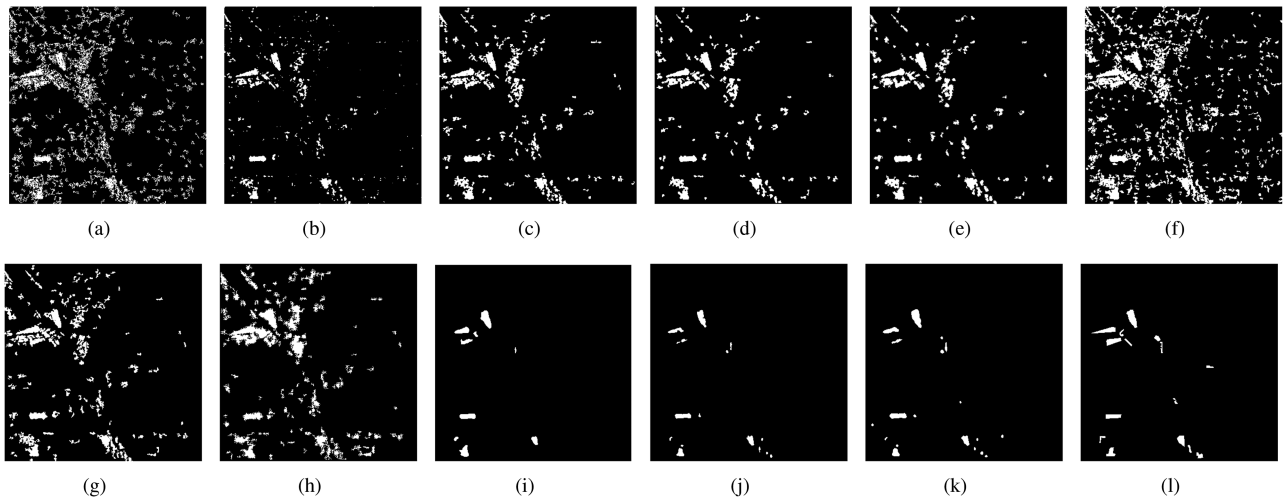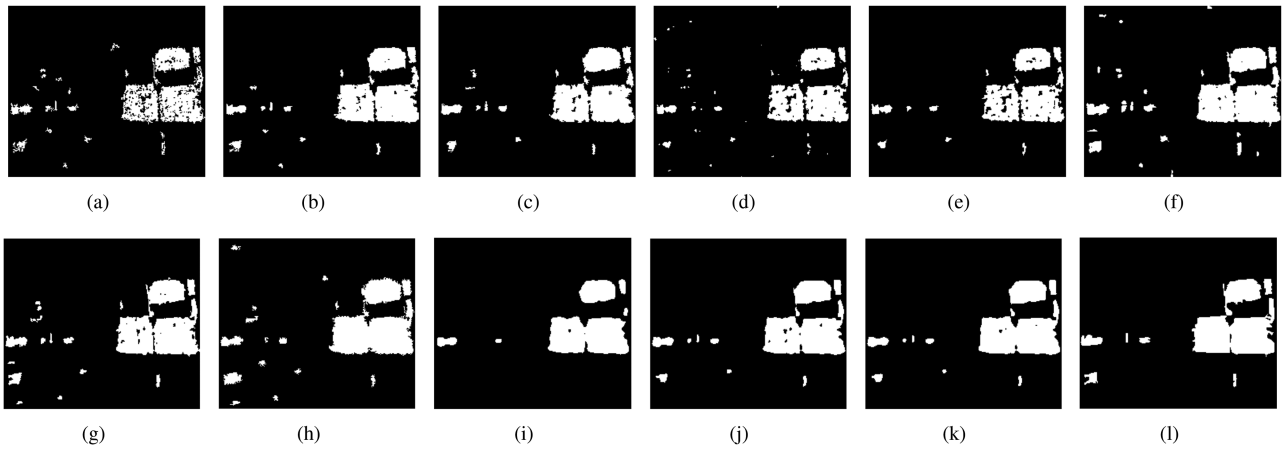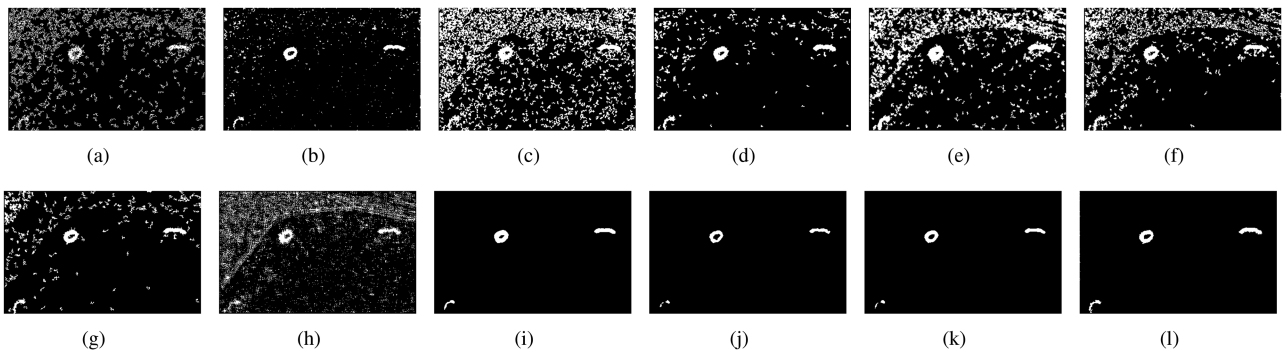

Fig. 11. Visualization of various change detection methods on Coast dataset: (a) LR-FCM, (b) LR-FLICM, (c) PCA-KM, (d) NR-ELM, (e) GaborPCANet, (f) CWNN, (g) DDNet, (h) SAFNet, (i) RUSACD, (j) proposed (IPs only), (k) proposed, and (l) ground truth.

confirming its sensitivity to subtle changes. In comparison to test results of only the intermediate patches in Fig. 10(j), repredictions of all patches [displayed in Fig. 10(k)] produced less noisy and more complete detection in the main change area. Highest κ value of 93.22% was achieved with the proposed method. In particular, unlike the Beijing-II dataset, although the degree of difference between bare ground and vegetation is small, due to the large area of occurrence, the proposed method could still detect all the change areas with a relatively high accuracy.

E. Results on Coast Dataset

Change maps generated by various methods on the Coast dataset are shown in Fig. 11, while the corresponding evaluation results are listed in Table IV. Types of changes in the Coast dataset are relatively simple, all of which are located in the small area in water. Judging from the visual results, LR-FCM, PCA-KM, GaborPCANet, CWNN, and SAFNet were all seriously influenced by speckle noise. Large numbers of false detections exhibited not only in the land area, but also in the water, resulting in extremely high FP values and kappa coefficient lower than 20%. For LR-FLICM, NR-ELM, and DDNet, the FP values

TABLE IV
CHANGE DETECTION RESULTS OF VARIOUS METHODS ON COAST DATASET

Methods	FP	FN	OE	PCC (%)	κ (%)
LR-FCM	13978	257	14235	88.70	11.55
LR-FLICM	2590	174	2764	97.81	45.07
PCA-KM [41]	32299	25	32324	74.35	5.63
NR-ELM [42]	5142	21	5163	95.90	32.76
GaborPCANet [43]	23517	**7**	23524	81.33	8.37
CWNN [44]	16191	50	16241	87.11	12.03
DDNet [45]	5663	59	5722	95.46	29.80
RUSACD [32]	130	199	**329**	**99.74**	**87.34**
SAFNet [46]	15236	230	15466	87.73	10.87
Proposed (IPs only)	**5**	416	421	99.67	81.41
Proposed	27	352	379	99.70	83.87

were also relatively high, and hence the detection accuracies were all below 50%. Among all the methods, RUSACD had the best PCC value of 99.74% and the highest κ of 87.34%, and yielded a clean change map similar to the ground truth. With the integration of LMR, LR, and SUB difference operators, the proposed method could make full use of various types of difference information to alleviate the impact of noise to a great extent, effectively detecting changed areas and producing a competitive performance to RUSACD.
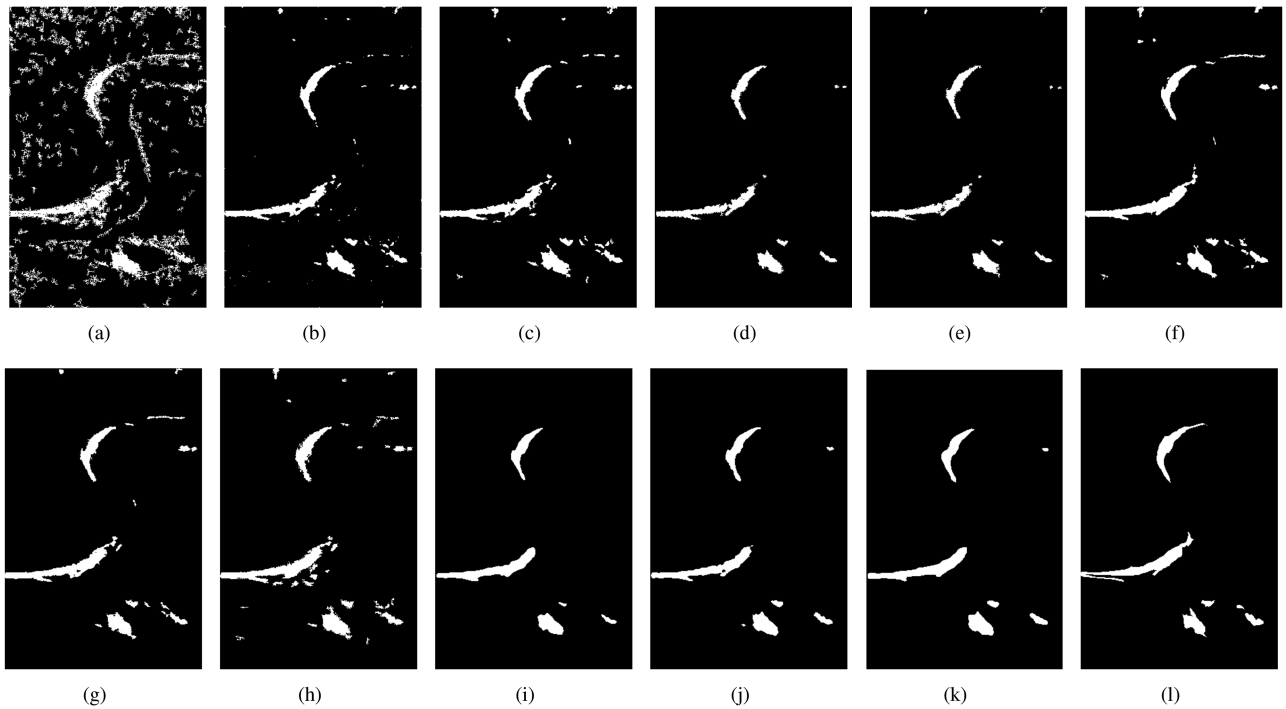
Fig. 12. Visualization of various change detection methods on Yellow River dataset: (a) LR-FCM, (b) LR-FLICM, (c) PCA-KM, (d) NR-ELM, (e) GaborPCANet, (f) CWNN, (g) DDNet, (h) SAFNet, (i) RUSACD, (j) proposed (IPs only), (k) proposed, and (l) ground truth.

TABLE V
CHANGE DETECTION RESULTS OF VARIOUS METHODS ON YELLOW
RIVER DATASET

Methods	FP	FN	OE	PCC(%)	κ(%)
LR-FCM	12265	517	12782	90.11	33.47
LR-FLICM	1344	788	2132	98.35	75.63
PCA-KM [41]	673	1693	2366	98.17	74.21
NR-ELM [42]	599	1229	1828	98.59	76.08
GaborPCANet [43]	646	1219	1865	98.56	75.76
CWNN [44]	1534	**496**	2030	98.43	77.93
DDNet [45]	1344	648	1992	98.46	77.56
RUSACD [32]	650	1190	1840	98.58	76.18
SAFNet [46]	2284	579	2863	97.78	70.85
Proposed (IPs only)	**640**	883	1523	98.82	80.97
Proposed	805	647	**1452**	**98.88**	**82.67**

F. Results on Yellow River Dataset

Fig. 12 and Table V show the detected change maps and their quantitative evaluation metrics of various state-of-the-art methods. In this Yellow River dataset, most of changes occur in the water area, making the change detection relatively straightforward. Therefore, except for LR-FCM, which had almost no resistance to noise, all methods exhibited some noise resilient capability in the detected change maps with Kappa coefficients higher than 70%. From visual comparisons, LR-FLICM, CWNN, DDNet, and SAFNet produced some noisy regions along the river boundary, and hence suffered from high FPs. On the contrary, for PCA-KM, NR-ELM, GaborPCANet, and RUSACD, many subtle change areas were eliminated, and FNs values were relatively high. Especially, CWNN, which was sensitive to changing differences, achieved the least number of missed detections of 496, and RUSACD with the strongest anti-noise ability yielded the least number of false detections of 640. The proposed method outperformed all these methods on both

accuracy and Kappa coefficient, and generated balanced FP and FN values, further demonstrating its effectiveness and usefulness in suppressing the influence of speckle noise in change detection.

G. Ablation Studies

Several aspects are explored and discussed in this section: 1) impact of number of training samples; 2) analysis of preclassification performance; 3) analysis of different types of difference maps and clustering algorithms; 4) impact of input patch size; 5) analysis of different classification strategies; 6) analysis of different types of loss functions; and 7) analysis on different types of classifiers.

1) Impact of Number of Training Samples: The number of training samples greatly affects the change detection performance and thus is a critical parameter for model training. In this article, we have developed the MH-FLICM algorithm based on visual saliency difference map for change detection. To test its effectiveness, the Beijing-I dataset was chosen to evaluate the number of training samples, and the proposed method was compared to 15 combinations of various difference maps and clustering algorithms, including three-class FCM based on neighborhood-ratio difference map (NR-FCM3), three-class FCM based on LMR difference map (LMR-FCM3), three-class FCM based on LR difference map (LR-FCM3), three-class FCM based on the visual saliency difference map (VSDM-FCM3), H-FCM based on NR difference map (NR-HFCM), H-FCM based on LMR difference map (LMR-HFCM), and H-FCM based on LR difference map (LR-HFCM), H-FCM based on the visual saliency difference map (VSDM-HFCM), three-class FLICM based on neighborhood ratio difference map (NR-FLICM3), three-class FLICM based on LMR difference map
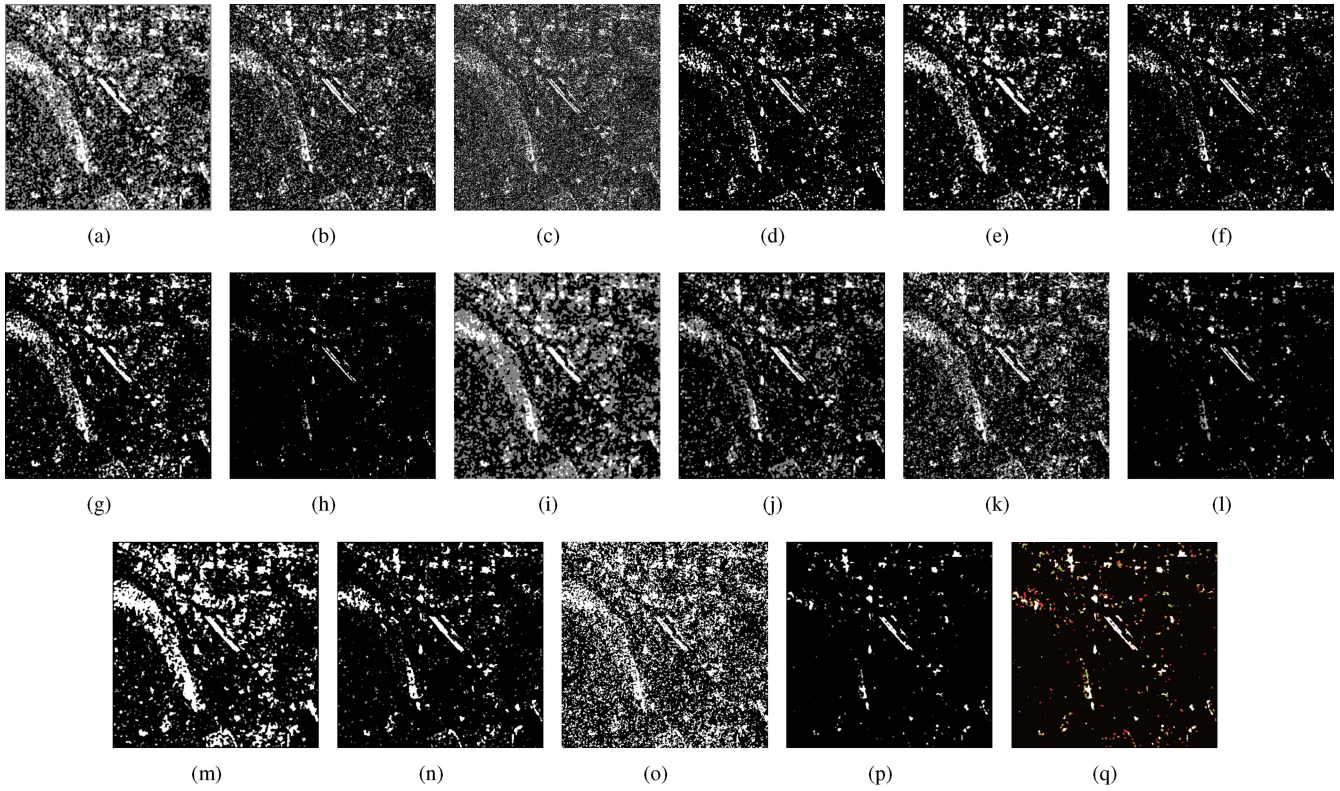
Fig. 13. Pre-classification results of (a) NR-FCM3, (b) LMR-FCM3, (c) LR-FCM3, (d) VSDM-FCM3, (e) NR-HFCM, (f) LMR-HFCM, (g) LR-HFCM, (h) VSDM-HFCM, (i) NR-FLICM3, (j) LMR-FLICM3, (k) LR-FLICM3, (l) VSDM-FLICM3, (m) NR-MHFLICM, (n) LMR-MHFLICM, (o) LR-MHFLICM, (p) proposed after merging HUPs into UPs, and (q) proposed before merging HUPs into UPs.

(LMR-FLICM3), three-class FLICM based on LR difference map (LR-FLICM3), three-class FLICM based on the visual saliency difference map (VSDM-FLICM3), MH-FLICM based on NR difference map (NR-MHFLICM), MH-FLICM based on LMR difference map (LMR-MHFLICM), and MH-FLICM based on LR difference map (LR-MHFLICM). As displayed in Fig. 13, the proposed method produced stable and accurate pre-classification results both before and after merging HUPs with UPs. In Fig. 13(a)–(p), the black, grey, and white pixels represent UPs, IPs, and CPs, respectively. In Fig. 13(q), the black, red, green, and white pixels represent HUPs, UPs, IPs, and CPs, respectively. Parameters UU, UI, UC, CU, CI, and CC were calculated for comparison, where the first character represents the class in the reference image ("C" is the changed class, "I" represents intermediate class and "U" denotes the unchanged class), and the second character denotes the pixel class in selected samples. These quantitative results of various methods are presented in Table VI. The proportion of different classes in the pre-classification result is visualized in Fig. 14. UU values below 40% were omitted. It can be observed that among all the methods, largest numbers of samples have been obtained using the proposed method.

2) Analysis of Pre-classification Performance: Apart from the number of training samples, selecting samples that would achieve as many correctly classified pixels as possible is also a significant factor. The proportion of pixels that have been
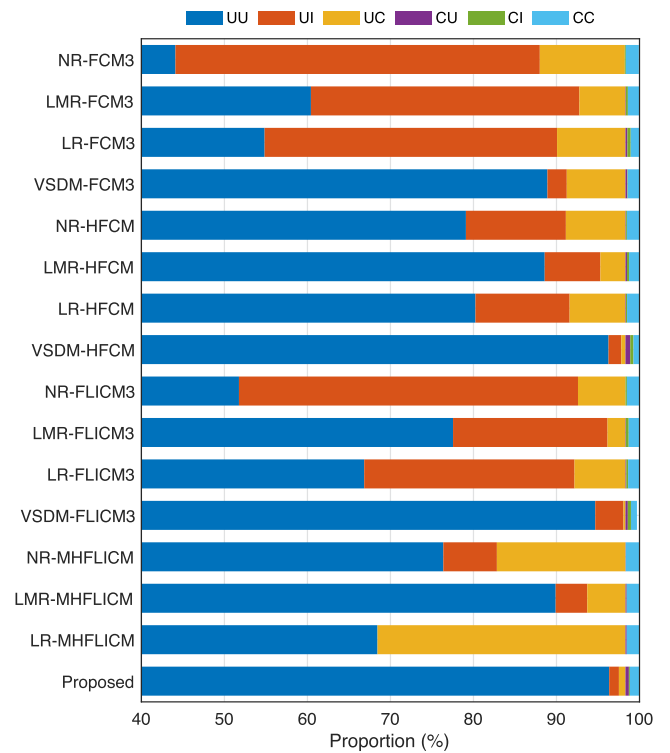

Fig. 14. Visualization of volume proportion (%) in pre-classification on Beijing-I dataset.

TABLE VI
PROPORTION (%) OF DIFFERENT CLASSES IN PRE-CLASSIFICATION RESULTS ON
BEIJING-I DATASET

Methods	UU	UI	UC	CU	CI	CC
NR-FCM3	44.19	43.89	10.24	0.00	0.12	1.57
LMR-FCM3	60.42	32.34	5.56	0.05	0.23	1.41
LR-FCM3	54.84	35.23	8.24	0.23	0.40	1.06
VSDM-FCM3	88.91	2.33	7.07	0.21	0.06	1.42
NR-HFCM	79.09	12.04	7.18	0.05	0.12	1.52
LMR-HFCM	88.58	6.71	3.02	0.18	0.25	1.26
LR-HFCM	80.24	11.34	6.73	0.06	0.13	1.51
VSDM-HFCM	96.27	1.54	0.50	0.58	0.38	0.72
NR-FLICM3	51.75	40.84	5.73	0.00	0.16	1.53
LMR-FLICM3	77.54	18.61	2.16	0.05	0.34	1.30
LR-FLICM3	66.84	25.33	6.15	0.09	0.22	1.37
VSDM-FLICM3	94.67	3.40	0.24	0.28	0.38	0.72
NR-MHFLICM	76.37	6.45	15.50	0.03	0.00	1.66
LMR-MHFLICM	89.90	3.81	4.61	0.13	0.02	1.54
LR-MHFLICM	68.44	0.00	29.87	0.15	0.00	1.53
Proposed	96.35	1.17	0.79	0.41	0.12	1.16

TABLE VII
PRE-CLASSIFICATION PERFORMANCE WITH VARIOUS DIFFERENCE MAPS AND
CLUSTERING ALGORITHMS ON BEIJING-I DATASET

Methods	ICV	SA
NR-FCM3	45.76	81.72
LMR-FCM3	61.83	91.68
LR-FCM3	55.90	86.84
VSDM-FCM3	90.33	92.54
NR-HFCM	80.61	91.77
LMR-HFCM	89.83	96.56
LR-HFCM	81.75	92.33
VSDM-HFCM	96.99	98.89
NR-FLICM3	53.07	90.29
LMR-FLICM3	78.85	97.28
LR-FLICM3	68.21	91.62
VSDM-FLICM	95.35	99.46
NR-MHFLICM	78.03	83.41
LMR-MHFLICM	91.44	95.07
LR-MHFLICM	69.97	69.98
Proposed	97.51	98.78

correctly classified in the pre-classification results (initial correct volume, ICV) and the sample accuracy (SA) are used to evaluate the accuracy of the selected pixels, given as

$$ICV = CC + UU \tag{15}$$

$$SA = \frac{CC + UU}{CC + UU + CU + UC}. \tag{16}$$

From Table VII, it can be seen that 97.51% of the pixels have been correctly classified in the pre-classification. While the accuracy of the selected samples reached 98.78%. Although the VSDM-FLICM algorithm was comparable with other sample selection algorithms, the visual saliency difference map based MH-FLICM algorithm achieved the best overall pre-classification result, further verifying the superior ability of the proposed sample selection method for selecting valid samples in terms of the number of samples and classification accuracy.

3) Analysis of Different Types of Difference Maps and Clustering Algorithms: In a change detection framework, sample selection performance depends on how well the difference map extracts and highlights difference information, as well as how well the clustering algorithm identifies potential changed areas for pre-classification. Also the sample selection and pre-classification accuracies would directly influence the final detection results. Therefore, to analyze the usefulness of proposed

TABLE VIII
CHANGE DETECTION RESULTS WITH VARIOUS DIFFERENCE MAPS AND
CLUSTERING ALGORITHMS ON BEIJING-I DATASET

Methods	FP	FN	OE	PCC (%)	κ (%)
NR-FCM3	33477	33	33510	79.06	10.94
LMR-FCM3	20210	135	20345	87.28	17.66
LR-FCM3	10169	469	10638	93.35	27.56
VSDM-FCM3	9323	242	9565	94.02	32.10
NR-HFCM	16726	119	16845	89.47	21.13
LMR-HFCM	7058	369	7427	95.36	36.93
LR-HFCM	12990	157	13147	91.78	25.78
VSDM-HFCM	498	1359	1857	98.84	58.56
NR-FLICM3	21131	67	21198	86.75	17.41
LMR-FLICM3	7360	264	7624	95.24	37.35
LR-FLICM3	17224	160	17384	89.14	20.26
VSDM-FLICM3	451	1297	1748	98.91	61.12
NR-MHFLICM	31775	22	31797	80.13	11.66
LMR-MHFLICM	7868	240	8108	94.93	36.07
LR-MHFLICM	9229	134	9363	94.15	33.60
Proposed	691	905	**1596**	**99.00**	**68.74**

visual saliency difference map and MH-FLICM algorithm, we conducted additional experiments, in which different types of difference maps and various clustering algorithms were utilized and compared, while the subsequent processes stayed the same. Final change detection results on the Beijing-I dataset are reported in Table VIII. The proposed visual saliency difference map based MH-FLICM algorithm outperformed all other methods, obtaining the highest accuracy of 99.00% and Kappa coefficient of 68.74%. It is worth mentioning that compared with the NR, LMR, and LR difference maps, the proposed visual saliency difference map achieved superior results no matter what clustering algorithm was adopted. We can also observe that although certain types of difference maps (NR and LMR) worked well with HFCM and FLICM, the proposed MH-FLICM clustering algorithm produced considerably good results on all types of difference maps, especially the visual saliency difference map, confirming the effectiveness of both the visual saliency difference map and the MH-FLICM clustering algorithm.

4) Impact of Input Patch Size: For shallow CNNs, the size of input patch has an important impact on feature extraction and learning. The proposed dual-channel network was constructed to learn the mapping relationship from input patches to labels of center pixels, and in this case, when the input patch is too large, too much interference may be brought into the feature extraction. While if the input patch is too small, along the network the feature maps would reduce in size or fill with too many zeros, hence extracting meaningless and useless feature maps. To determine its optimal size, we varied the input patch size from 9×9 to 17×17 on all five datasets. As displayed in Fig. 15 and Table IX, with increase in patch size, false positive rate showed a tendency of gradually increasing, while false negative rate decreases. On the Beijing-III, Coast and Yellow River datasets, a moderate patch size seemed more suitable.

5) Analysis of Different Classification Strategies: To confirm the efficacy of reclassifying all the patches, we adopted two classification strategies and compared their subsequent performance in terms of Kappa coefficients. The first mechanism supposed that there were no errors in the pre-classification, and then a suitable threshold was chosen to predict patches centred at IPs. In
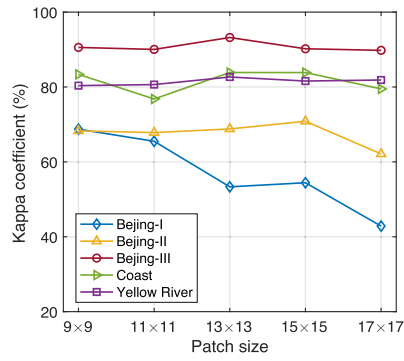
Fig. 15. Comparisons of Kappa coefficients with varying patch sizes.

TABLE IX
CHANGE DETECTION RESULTS WITH VARYING PATCH SIZES

| Dataset	Patch	FP	FN	OE	PCC (%)	κ (%)
Beijing-I	9×9	691	905	1596	99.00	68.74
	11×11	1537	606	2143	98.66	65.50
	13×13	3822	250	4072	97.46	53.54
	15×15	3547	392	3866	97.58	54.43
	17×17	6242	155	6397	96.00	42.85
Beijing-II	9×9	150	1076	1226	99.23	68.28
	11×11	237	1044	1281	99.20	67.82
	13×13	429	911	1340	99.16	68.80
	15×15	385	867	1252	99.22	70.85
	17×17	1710	528	2238	98.60	62.13
Beijing-III	9×9	181	877	1028	98.07	90.57
	11×11	161	918	1079	97.98	90.04
	13×13	275	488	763	98.57	93.22
	15×15	173	891	1064	98.01	90.21
	17×17	162	940	1104	97.93	89.79
Coast	9×9	17	389	406	99.68	82.37
	11×11	10	498	508	99.60	76.80
	13×13	27	352	379	99.70	83.87
	15×15	41	342	383	99.70	83.68
	17×17	22	441	463	99.63	79.49
Yellow River	9×9	622	936	1558	98.79	80.37
	11×11	687	872	1559	98.79	80.65
	13×13	805	647	1452	98.88	82.67
	15×15	764	753	1517	98.83	81.59
	17×17	695	782	1477	98.86	81.87
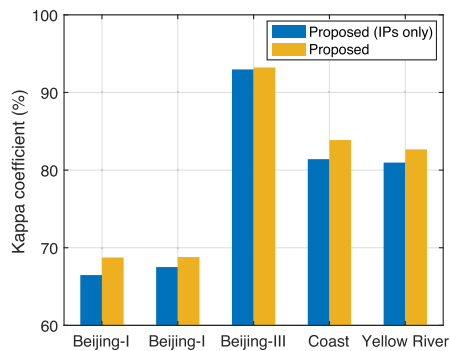
Fig. 16. Comparisons of Kappa coefficients with different classification strategies.

the second mechanism, patches corresponding of all pixels were re-classified to obtain a probability map, and FLICM clustering algorithm was then used to distinguish between unchanged and changed categories. Kappa coefficient κ was chosen as the evaluation criterion and results were compared on the Beijing-I, Beijing-II, Beijing-III, Coast, and Yellow River datasets. As shown in Fig. 16, the mechanism of repredicting all the patches produced higher κ values on all five datasets, confirming its effectiveness.

TABLE X
CHANGE DETECTION RESULTS WITH DIFFERENT TYPES OF LOSS FUNCTIONS

Dataset	Loss	FP	FN	OE	PCC (%)	κ (%)
Beijing-I	BCEL	1241	806	2047	98.72	64.30
	NAWL	1078	795	1873	98.83	66.47
	CBAWL	691	905	1596	99.00	68.74
Beijing-II	BCEL	244	1102	1346	99.16	65.76
	NAWL	368	916	1284	99.20	69.66
	CBAWL	385	867	1252	99.22	70.85
Beijing-III	BCEL	174	877	1051	98.03	90.34
	NAWL	320	570	890	98.33	92.07
	CBAWL	275	488	763	98.57	93.22
Coast	BCEL	12	459	471	99.63	78.88
	NAWL	17	414	431	99.66	81.09
	CBAWL	27	352	379	99.70	83.87
Yellow River	BCEL	592	1040	1632	98.74	79.11
	NAWL	588	887	1475	98.86	81.45
	CBAWL	805	647	1452	98.88	82.67

6) Analysis of Different Types of Loss Functions: In the change detection network, class-balanced adaptive focal loss is employed to address the class imbalance issue during training. According to the number of samples in each class, it applies an adaptive balancing factor and a modulating term to the cross entropy loss in order to focus training on misclassified examples. To evaluate its efficacy, change detection results of binary cross entropy loss, nonadaptive focal loss and class-balanced adaptive focal loss were compared, terms as BCEL, NAWL, and CBAWL, respectively. As presented in Table X, with class-balanced adaptive focal loss, higher accuracies and Kappa coefficients were obtained on all datasets. It can be observed that the model trained using the binary cross entropy loss produced large false negatives in most cases, indicating the network biased toward learning more representations of the disadvantaged class and neglecting the advantaged class. The nonadaptive focal loss extends the binary cross entropy by introducing an additional modulating term to avoid producing large gaps between numbers of false negative and false positives. The proposed class-balanced adaptive focal loss further incorporates a class-balanced weight factor, pushing the network to adapt to imbalanced classes.

7) Analysis of Different Types of Classifiers: Deep learning based change detection methods commonly use the softmax function as the final classifier to determine the predictions from the probability map. In this article, we proposed to adopt FLICM algorithm to produce satisfactory change detection results. To validate the usefulness of the FLICM, experiments were performed by comparing three different types of classifiers: softmax, FCM and FLICM. Quantitative results on the five datasets are given in Table XI. Compared with softmax, by employing the FLICM algorithm to cluster pixels according to the neighborhood information, the proposed method achieved higher accuracies and Kappa coefficients on four dataset. In Beijing-II dataset, although the detection accuracy of FLICM was marginally worse than that of softmax, higher κ was obtained with more balanced FP and FN values. Furthermore, fusing with local neighborhood information, FLICM yields stable and appreciable results with stronger capability of suppressing noise, illustrating its usefulness in distinguishing changed, and unchanged regions.

TABLE XI
CHANGE DETECTION RESULTS WITH DIFFERENT TYPES OF CLASSIFIERS

Dataset	Classifier	FP	FN	OE	PCC (%)	κ (%)
Beijing-I	Softmax	1019	811	1830	98.86	66.81
	FCM	694	929	1623	98.99	68.09
	FLICM	691	905	1596	99.00	68.74
Beijing-II	Softmax	257	1020	1277	99.20	68.26
	FCM	287	970	1257	99.21	69.35
	FLICM	385	867	1252	99.22	70.85
Beijing-III	Softmax	282	509	791	98.52	92.96
	FCM	206	638	844	98.42	92.38
	FLICM	275	488	763	98.57	93.22
Coast	Softmax	19	359	378	99.70	83.81
	FCM	14	417	431	99.66	81.04
	FLICM	27	352	379	99.70	83.87
Yellow River	Softmax	713	800	1513	98.83	81.43
	FCM	572	979	1551	98.80	80.24
	FLICM	805	647	1452	98.88	82.67

VI. CONCLUSION

In this article, we have developed an approach to extract difference map based on visual saliency, capable of highlighting authentic change areas with robust suppression of false alarms. An unsupervised multi-hierarchical clustering algorithm, MH-FLICM, is then proposed to effectively detect potential changed areas and select high-quality samples with accurate pseudolabels. These pre-classified samples can be constructed as inputs to the change detection network, in which class-balanced adaptive focal loss is incorporated for diminishing the gap between numbers of samples in unchanged and changed classes. FLICM clustering algorithm is employed to yield satisfactory predictions from the network output with strong resistance to noise.

Experimental results and comparisons with state-of-the-art algorithms on various datasets have verified the advantages of the proposed SAR change detection framework, reaching averaged accuracy of 99.07% and Kappa coefficient of 79.87% over five benchmark datasets. The proposed method combines the respective strengths of multi-hierarchical clustering and deep models to produce training samples of high-confidence in an unsupervised way, and to yield stable and exceeding change detection performances, particularly in images with changes in small areas. It is worth mentioning that in spite of complex types of changes involved, improved performances over the existing methods have been achieved both visually and quantitatively.

REFERENCES

[1] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.

[2] A. Reigber et al., "Very-high-resolution airborne synthetic aperture radar imaging: Signal processing and applications," *Proc. IEEE*, vol. 101, no. 3, pp. 759–783, Mar. 2013.

[3] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[4] J. Lu, J. Li, G. Chen, L. Zhao, B. Xiong, and G. Kuang, "Improving pixel-based change detection accuracy using an object-based approach in multitemporal SAR flood images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3486–3496, Jul. 2015.

[5] Y. Li, S. Martinis, S. Plank, and R. Ludwig, "An automatic change detection approach for rapid flood mapping in sentinel-1 SAR data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 73, pp. 123–135, 2018.

[6] H. Gokon et al., "A method for detecting buildings destroyed by the 2011 tohoku earthquake and tsunami using multitemporal terraSAR-X data," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1277–1281, Jun. 2015.

[7] M. Wieland, W. Liu, and F. Yamazaki, "Learning change from synthetic aperture radar images: Performance evaluation of a support vector machine to detect earthquake and tsunami-induced changes," *Remote Sens.*, vol. 8, no. 10, 2016, Art. no. 729.

[8] L. Li, C. Wang, H. Zhang, B. Zhang, and F. Wu, "Urban building change detection in SAR images using combined differential image and residual U-Net network," *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 1091.

[9] S. Saha, F. Bovolo, and L. Bruzzone, "Building change detection in VHR SAR images via unsupervised deep transcoding," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 1917–1929, Mar. 2021.

[10] R. S. Lunetta, J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy, "Land-cover change detection using multi-temporal MODIS NDVI data," *Remote Sens. Environ.*, vol. 105, no. 2, pp. 142–154, 2006.

[11] D. C. Zanotta and V. Haertel, "Gradual land cover change detection based on multitemporal fraction images," *Pattern Recognit.*, vol. 45, no. 8, pp. 2927–2937, 2012.

[12] S. H. Khan, X. He, F. Porikli, and M. Bennamoun, "Forest change detection in incomplete satellite images with deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5407–5423, Sep. 2017.

[13] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858–867, Jul. 1997.

[14] M. Roy, S. Ghosh, and A. Ghosh, "A neural approach under active learning mode for change detection in remotely sensed images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1200–1206, Apr. 2014.

[15] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005.

[16] H.-C. Li, H.-C. Li, G. Yang, W. Yang, Q. Du, and W. J. Emery, "Deep non-smooth nonnegative matrix factorization network with semi-supervised learning for SAR image change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 160, pp. 167–179, 2020.

[17] J. Wang, X. Yang, L. Jia, and S. Fang, "Unsupervised change detection between SAR images based on hypergraphs," *ISPRS J. Photogramm. Remote Sens.*, vol. 164, pp. 61–72, 2020.

[18] Y. Bazi, L. Bruzzone, and F. Melgani, "Automatic identification of the number and values of decision thresholds in the log-ratio image for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 349–353, Jul. 2006.

[19] B. Hou, Q. Wei, Y. Zheng, and S. Wang, "Unsupervised change detection in SAR image based on gauss-log ratio image fusion and compressed projection," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3297–3317, Aug. 2014.

[20] M. Gong, Y. Cao, and Q. Wu, "A neighborhood-based ratio approach for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 307–311, Mar. 2012.

[21] Y. Zheng, X. Zhang, B. Hou, and G. Liu, "Using combined difference image and k-means clustering for SAR image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 691–695, Mar. 2014.

[22] B. Cui, Y. Zhang, L. Yan, J. Wei, and Q. Huang, "A SAR change detection method based on the consistency of single-pixel difference and neighbourhood difference," *Remote Sens. Lett.*, vol. 10, no. 5, pp. 488–495, 2019.

[23] J. Geng, X. Ma, X. Zhou, and H. Wang, "Saliency-guided deep neural networks for SAR image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7365–7377, Oct. 2019.

[24] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[25] C. Gao, H. Zhang, C. Wang, and F. Wu, "SAR change detection based on generalized gamma distribution divergence and auto-threshold segmentation," *J. Remote Sens.*, vol. 14, no. 4, pp. 710–724, 2010.

[26] Y. Bazi, L. Bruzzone, and F. Melgani, "Image thresholding based on the em algorithm and the generalized Gaussian distribution," *Pattern Recognit.*, vol. 40, no. 2, pp. 619–634, 2007.

[27] S. Krinidis and V. Chatzis, "A robust fuzzy local information C-means clustering algorithm," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1328–1337, May 2010.

[28] M. Gong, L. Su, M. Jia, and W. Chen, "Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar images," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 1, pp. 98–109, Feb. 2014.

[29] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2141–2151, Apr. 2012.

[30] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jia, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.

[31] Y. Gao, F. Gao, J. Dong, and S. Wang, "Transferred deep learning for Sea Ice change detection from synthetic-aperture radar images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1655–1659, Oct. 2019.

[32] X. Zhang, H. Su, C. Zhang, X. Gu, X. Tan, and P. M. Atkinson, "Robust unsupervised small area change detection from SAR imagery using deep learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 79–94, 2021.

[33] Y. Zheng, L. Jiao, H. Liu, X. Zhang, B. Hou, and S. Wang, "Unsupervised saliency-guided SAR image change detection," *Pattern Recognit.*, vol. 61, pp. 309–326, 2017.

[34] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. ACM Int. Conf. Multimedia*, 2006, pp. 815–824.

[35] R. Cong, J. Lei, H. Fu, M. M. Cheng, W. Lin, and Q. Huang, "Review of visual saliency detection with comprehensive information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 10, pp. 2941–2959, Oct. 2019.

[36] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 4353–4361.

[37] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[38] Y. Zhao, M. Jiang, J. Kong and S. Li, "Paralleled attention modules and adaptive focal loss for siamese visual tracking," *IET Image Process.*, vol. 15, no. 6, pp. 1345–1358, 2021.

[39] Y. Liang, Y. Long, Y. Li, J. Liang and Y. Wang, "Joint framework with deep feature distillation and adaptive focal loss for weakly supervised audio tagging and acoustic event detection," *Digit. Signal Process.*, vol. 123, pp. 103446–103455, 2022.

[40] X. Zhao, J. Yao, W. Deng, M. Jia and Z. Liu, "Normalized conditional variational auto-encoder with adaptive focal loss for imbalanced fault diagnosis of bearing-rotor system," *Mech. Syst. Signal Process.*, vol. 170, 2022, Art. no. 108826.

[41] 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.

[42] 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, 2016, Art. no. 046019.

[43] 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.

[44] F. Gao, X. Wang, Y. Gao, J. Dong, and S. Wan, "Sea Ice change detection in SAR images based on convolutional-wavelet neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1240–1244, Aug. 2019.

[45] X. Qu, F. Gao, J. Dong, Q. Du, and H.-C. Li, "Change detection in synthetic aperture radar images using a dual-domain network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021, Art no. 4013405.

[46] Y. Gao, F. Gao, J. Dong, Q. Du, H.-C. Li, "Synthetic aperture radar image change detection via siamese adaptive fusion network," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10748–10760, 2021.

Bin Cui received the B.Eng. degree in geomatics engineering from Henan Polytechnic University, Jiaozuo, China, in 2011, the M.Eng. degree in geomatics engineering from the Beijing University of Civil Engineering and Architecture, Beijing, China, in 2015, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2020.

He is currently a Lecturer with the School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing, China. And he is also enrolled as a Postdoctoral Researcher with the School of Geography and Ocean Science, Nanjing University. His current research interests include remote sensing image processing, change detection, multimodality image fusion and their applications in remote sensing.

Hujun Yin (Senior Member, IEEE) received the B.Eng. degree in electronic engineering and the M.Sc. degree in signal processing, both from Southeast University, Nanjing, China, in 1983 and 1986, respectively, and the Ph.D. degree in neural networks from the University of York, Heslington, U.K., in 1997.

Since 1996, he has been with the Department of Electrical and Electronic Engineering, The University of Manchester, Manchester, U.K., where he is currently a Professor of artificial intelligence. He has supervised more than 25 Ph.D. students and authored or coauthored more than 200 peer-reviewed articles. His main research interests include neural networks, self-organizing and unsupervised learning, deep learning, image classification and hyperspectral image processing, face recognition, data analytics, time series modelling and prediction, bio-/neuro-informatics, and interdisciplinary applications.

Prof. Yin was a recipient of more than £4 million research funding from UK research councils, EPSRC, BBSRC, Innovate UK and industries across 25 funded projects. Many of his projects involve industries in developing cutting edge AI solutions. He had served as an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS, 2006–2010 and the *International Journal of Neural Systems* 2005–2020. Since 2015, he has been serving as an Associate Editor for the IEEE TRANSACTIONS ON CYBERNETICS. He has also served as the General Chair or Programme Chair for a number of international conferences in AI, machine learning and data analytics. He is a member of the EPSRC Peer Review College (since 2006), a senior member of the IEEE (since 2003) and a Turing Fellow of the Alan Turing Institute (since 2018).

Yonghong Zhang received the bachelor's degree in mathematics from Wuhan University, Wuhan, China, in 1994, and the Ph.D degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2001.

He is with the Chinese Academy of Surveying and Mapping, Beijing, China. He has authored more than 120 publications. His research interests include algorithm development and applications associated with SAR, SAR interferometry, and SAR polarimetry.

Yao Peng (Member, IEEE) received the B.Eng. degree in detection guidance and control techniques from the Nanjing University of Science and Technology, Nanjing, China, in 2013, the M.Sc. degree in digital image and signal processing from the University of Manchester, Manchester, U.K., in 2014, and the Ph.D. degree in electrical and electronic engineering from the University of Manchester, U.K., in 2019.

She is currently enrolled as a Lecturer with the School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing, China. Her current research interests include generative adversarial learning, data-independent feature leaning, self-organizing and unsupervised learning, and their applications to face recognition, image generation, hyperspectral image analysis, remote sensing image processing.

Peijun Du (Senior Member, IEEE) received the Ph.D. degree in geodesy and survey engineering from the China University of Mining and Technology, Xuzhou, China, in 2001.

He is currently a Professor of remote sensing and geographic information science with Nanjing University, Nanjing, China. He has authored or coauthored more than 150 articles in international peer-reviewed journals, and more than 220 papers in international conferences and Chinese journals. His research interests focus on remote sensing image intelligent processing and geographic analysis, urban remote sensing for smart cities, and integrated applications of geospatial information technologies.

Dr. Du was an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS from 2009 to 2019.