

HyperNet: Self-Supervised Hyperspectral Spatial–Spectral Feature Understanding Network for Hyperspectral Change Detection

Meiqi Hu, *Graduate Student Member, IEEE*, Chen Wu[✉], *Member, IEEE*, and Liangpei Zhang[✉], *Fellow, IEEE*

Abstract—The fast development of self-supervised learning (SSL) lowers the bar learning feature representation from massive unlabeled data and has triggered a series of researches on change detection of remote sensing images. Challenges in adapting SSL from natural images classification to remote sensing images change detection arise from difference between the two tasks. The learned patch-level feature representations are not satisfying for the pixel-level precise change detection. In this article, we proposed a novel pixel-level self-supervised hyperspectral spatial–spectral feature understanding network (HyperNet) to accomplish pixelwise feature representation for effective hyperspectral change detection. Concretely, not patches but the whole images are fed into the network and the multitemporal spatial–spectral features are compared pixel by pixel. Instead of processing the 2-D imaging space and spectral response dimension in hybrid style, a powerful spatial–spectral attention module is put forward to explore the spatial correlation and discriminative spectral features of multitemporal hyperspectral images (HSIs), separately. Only the positive samples at the same location of bitemporal HSIs are created and forced to be aligned, aiming at learning the spectral difference-invariant features. Moreover, a new similarity loss function named focal cosine is proposed to solve the problem of imbalanced easy and hard positive samples comparison, where the weights of those hard samples are enlarged and highlighted to promote the network training. Six hyperspectral datasets have been adopted to test the validity and generalization of the proposed HyperNet. The extensive experiments demonstrate the superiority of HyperNet over the state-of-the-art algorithms on downstream hyperspectral change detection tasks.

Index Terms—Feature understanding, hyperspectral change detection, pixel level, self-supervised learning (SSL), spatial–spectral attention (SSA).

I. INTRODUCTION

CHANGE detection technology is a significant tool for urban development monitoring and land-cover/land-use change detection [1], [2], [3]. In general, change detection

Manuscript received 19 July 2022; revised 21 September 2022; accepted 13 October 2022. Date of publication 1 November 2022; date of current version 17 November 2022. This work was supported in part by the National Natural Science Foundation of China under Grant T2122014, Grant 61971317, and Grant 42230108. (Corresponding author: Chen Wu.)

The authors are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: meiqi.hu@whu.edu.cn; chen.wu@whu.edu.cn; zlp62@whu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2022.3218795

refers to detect the changes of two remote sensing images acquired on different times at the same location [4], [5]. Hyperspectral image (HSI) is generally composed of dozens and hundreds of spectral bands [6], [7], providing unique spectral curve and abundant spectral information for every kind of ground object and showing great potential for object discrimination and classification [8], [9]. Hyperspectral change detection is beneficial for discriminating the changes hard to notice. The two important tasks are hyperspectral anomalous change detection (HACD-task) and general hyperspectral binary change detection (HBCD), respectively. HACD-task [10], [11] focuses on the changes that usually originate from the displacement, appearance, disappearance, and concealment of small and rare objects. HACD-task is motivated to highlight those rare changes used for air defense and emergency response, etc. And HBCD is targeted at detecting the precise land cover transformation under the embarrassed situation that changes may happen between some pretty similar ground objects [12], [13].

The increase of deep learning has significantly promoted the development of change detection algorithms in recent years [14], [15], [16], [17]. Most of the current hyperspectral change detection methods are mainly based on the deep supervised learning [18], [19], where the training labels are necessary for the model training. The high-quality training data do provide accurate supervision for parameter training and bring about good performance, but are labor-consuming and even elusive to obtain. Thus, alleviating or eliminating the reliance of annotated labels turns into a key to the problem of hyperspectral change detection.

Self-supervised learning (SSL) [20], [21], [22] is a kind of new paradigm in machine learning enabling the model to learn feature representation from massive unlabeled data. And it has intrigued plenty of SSL representation learning algorithms for natural image processing. As a kind of discriminative SSL, contrastive learning (CL) [23], [24] aims at pulling together the positive sample pairs and pushing away the negative sample pairs, and has been applied for multispectral optical and synthetic aperture radar (SAR) remote sensing change detection [25], [26]. However, the SSL for hyperspectral change detection still faces the following limitations.

A. Pixel-Level Feature Representation for Precise Change Detection

The current SSL-based change detection methods are essentially dependent on instance discrimination. Concretely, the multitemporal images are segmented into patches for model input. The patches are mapped into feature vectors in the feature space for feature comparison. As a result, the model can only learn the patch-level feature representation, leading to coarse change detection results.

B. Fully Utilization of Spatial and Spectral Features

The spatial and spectral feature extractions are crucial for hyperspectral change detection since the multitemporal complex ground objects can be confusing due to the variant spectral difference. A common phenomenon is that most deep-learning-based change detection methods adopt 2-D convolutional layers to extract both spatial and spectral features simultaneously. However, with a spatial kernel filter sliding along horizontal and vertical direction, the 2-D convolutional layer focuses on the spatial information along each channel but ignores the relationship between different spectral bands, leading to ineffective exploitation of spectral information.

Given the consideration of problems and concerns mentioned above, a self-supervised hyperspectral spatial-spectral feature understanding network (HyperNet) is proposed. To achieve pixel-level feature for dense prediction task like change detection, we proposed an original fully convolutional SSL framework for convenient pixelwise feature learning. The similarity comparisons of multitemporal HSIs are conducted in pixel level not the patch level for fine-grained alignment from totally unlabeled data. And a powerful spatial-spectral attention (SSA) module is designed to fully exploit the informative spatial and spectral features, including the spatial attention branch and spectral attention branch. The former one is used to explore spatial features and selectively emphasize the most informative spatial objects. The latter one is tailored to make full use of the discriminative spectral information and the relationship between local and global spectral bands. The deep spatial and spectral features are extracted separately from the proposed two branches and then are adaptively fused to produce comprehensive features. Considering the imbalanced easy and hard samples comparison for SSL, a new loss function named focal cosine (cos) is put forward to enlarge the weight of those hard samples and promote the network training. We explore the generalization of the proposed self-supervised HyperNet on two kinds of change detection tasks as HACD-task and HBCD, as SSL is known for learning general features and strong generalization for different downstream tasks. The contribution of this article can be summed up as three parts.

- 1) A self-supervised HyperNet is elaborated for hyperspectral change detection. HyperNet successfully achieves pixel-level self-supervised feature representation learning for accurate dense change detection task. And a spatial attention branch and a spectral attention branch are designed to extract discriminative spatial and spectral information separately, instead of directly processing the HSIs in a hybrid way.

- 2) In order to relieve the imbalanced problem between the easy positive and hard positive samples in SSL training, a novel similarity loss function named focal cos loss is proposed to emphasize the weights of hard positive samples and relieve the dominance of the large majority easy samples. Experimental results tested on six hyperspectral change detection datasets indicate the effectiveness of the proposed focal cos loss function.
- 3) To demonstrate the validity of the proposed HyperNet, we have conducted extensive experiments on HACD-task and HBCD. HyperNet achieves a good performance on both HACD-task and HBCD results and outperforms the state-of-the-art algorithms, demonstrating the effectiveness and generalization of proposed extracting spatial and spectral feature separately for better HSI understanding.

This article is organized as follows. Section II will give a brief introduction of related works. The detailed description of the proposed HyperNet will be exhibited in Section III. The experimental results and analyses will be presented in Section IV. And the discussions on the ablation experiments will be shown in Section V. Finally, Section VI will conclude this article.

II. RELATED WORK

A. Hyperspectral Anomalous Change Detection

HACD-task focuses on the dynamics of small objects, facing the challenges of the violent spectral differences between the unchanged area of multitemporal HSIs, which are mainly caused by the complex atmosphere and illumination conditions, and the motion of sensors, etc. [27], [28]. The predictor-based method is a very classic solution, where the spectral values of a pair pixel of multitemporal HSIs can be mapped by a linear or nonlinear regression model [29], [30]. And the predictive image constructed by the predictor is adjusted to be as close as the other HSI for those unchanged background areas, like the vegetation and building area. Thus, the anomalous change can be separated easily from the residual image. Chronochrome (CC) [31] presented a linear space-invariant observation model to establish a linear mapping between two HSIs, where it is assumed that the observed spectral value is in linear regression with the real primitive spectral curve. However, the inhomogeneous spectral difference filed leaves the linear model with bad performance especially under variant weather conditions. Then, a novel segmented linear prediction method [32] argued to represent multitemporal HSIs by a normal mixture model and enable the linear predictor model to differ between the background subspaces. And hyperspectral anomaly change detection based on autoencoder (ACDA) [33] put forward to establish an effective nonlinear mapping relationship between the multitemporal spectral vectors by auto-encoder with multiple structural layers. It is found that these methods overemphasize the spectral information while neglect the spatial information to detect the anomalous changes.

B. Hyperspectral Binary Change Detection

For HBCD, one of the key problems of HBCD is how to take the advantage of the detailed spectral information, where two subproblems are involved, including the refinement of the redundant adjacent spectral information and the exploration of rich spectral features separating changes from pseudo-changes. The traditional methods can be roughly categorized as three groups [34], [35], where the algebraic-based methods calculate the spectral difference between multitemporal HSIs, the transformation-based approaches compare the transformed information of new feature space, and the classification-based algorithms obtain the change detection results by comparison of the classification maps of multitemporal HSIs. Hou et al. [36] have designed an innovative method to make full use of the spatial information of HSI by multiple morphological profiles, and effective strategy of the attribute profiles is presented to extract change information. Hou et al. [37] proposed a novel three-order tucker decomposition to effectively utilize the spectral and spatial information, aiming at suppressing the differences of unchanged and amplifying the differences of changed. And researchers have attempted to solve the HBCD problem with deep neural networks [38], [39]. General end-to-end 2-D convolutional neural network (GETNET) [40] developed an end-to-end 2-D convolutional neural network framework, where the designed mixed affinity matrix integrated with abundance maps is beneficial for full exploitation of the spectral information of HSIs. A new unsupervised spectral mapping method [41] based on auto-encoder with adversarial learning was proposed for HBCD which utilized the spectral features and minimized the reconstruction loss of input spectral vectors. Li et al. [42] presented a novel strategy which combined two unsupervised model-driven method to generate credible pseudo-labels for proposed 2-D HSI-change detection (CD) framework. And several attention-based methods have been developed for HBCD to enhance the most informative spatial and spectral features from the huge-volume HSI data. A multilevel encoder-decoder attention network (ML-EDAN [19]) was proposed to extract hierarchical features with the designed contextual-information guided attention module. SSA-SiamNet [43] was designed to emphasize discriminative channels and locations and suppress less informative ones. However, the deep-learning-based method mostly driven by annotated samples suffers from expensive samples or even no samples in practical applications. Thus, how to fully explore spatial and spectral information from HSI data with no labels or the coarse-grained labels has turned into a research priority currently [44], [45].

C. Self-Supervised Learning

Self-supervised feature learning is a powerful feature learning paradigm on unlabeled data [20], [21], [23], [46], and has achieved great success on visual image processing, e.g., image classification [47], video representation learning [48], and target recognition [49]. In general, the SSL produces massive labeled data by pretext tasks and then learns the feature representation from the self-produced labeled data [50], [51]. And CL is one kind of discriminative SSL, aiming at attracting the positive sample pairs and pushing away the negative

sample pairs. Data augmentation is one of the common ways to produce abundant positive and negative samples. One single image and its augmented view form a pair of positive samples, while a pair of negative samples refers to augmented views of different images. Concretely, random horizontal flip, vertical flip, random crop and resize, rotation, and noise are important and useful data transformation operators. And a great number of new methods based on Siamese network have been developed to drive feature representation learning. Momentum contrast (MoCo) [52] builds a large dictionary look-up to facilitate contrastive feature learning. Simple framework for contrastive learning of visual representations (SimCLR) [53] uses plenty of negative samples within a batch to balance the feature alignment and the uniformity of the samples embedded into the feature space. Bootstrap your own latent (BYOL) [54] directly removes the negative samples and predicts the view of target network by training the online network, where the target network is updated with a slow-moving average of the online network. Simple siamese network (SimSiam) [55] maximizes the similarity of the prediction target of one augmented view and the embedding vector of another view from the encoder under stop-gradient constraint.

The prevailing of CL has sparked the development of advanced SSL-based change detection methods of remote sensing images [25], [26]. Currently, CL has also been applied for multispectral optical and SAR remote sensing change detection. Under the prior probability of little changes, bitemporal remote sensing images are regarded as different views of the same image acquired at different times on the same location. Image patches of the same location of bitemporal images are saw as positive sample pairs, while image patches of different location of the same training batch size are viewed as negative sample pairs. Without any labels, the network is capable of learning feature representation from input by minimizing the distance of embedding feature vectors of positive samples and pushing apart embedding feature vectors of negative samples. In general, to avoid the collapse of the network, a large batch size is necessary for the satisfaction of enough negative samples of each batch.

Under the situation of suboptimal exploitation of spatial and spectral information and the few and limited training samples for hyperspectral change detection, the proposed method is deeply inspired by the current SSL. We proposed a novel self-supervised spatial-spectral hyperspectral understanding network to accomplish accurate change detection. The proposed method processes the HSIs exactly from the perspective of 2-D imaging space and the spectral dimension, digging the deep and effective spatial and spectral features. Without negative sample pairs, the proposed HyperNet uses only positive sample pairs for similarity comparison. And large training batch size is not necessary for HyperNet, improving the practicability of HyperNet. Rather than patch-based sample pairs, the whole images are fed into the HyperNet to acquire global sense of information comparison and pixelwise features for change detection.

III. METHODOLOGY

This section gives a detailed representation of the proposed approach HyperNet. Fig. 1 shows the flowchart of HyperNet

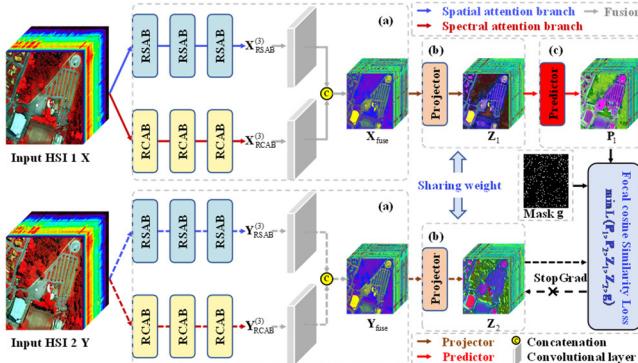


Fig. 1. Structure of the proposed HyperNet for hyperspectral change detection: (a) is the SSA module, which is especially designed to fully exploit the abundant information of HSIs, including the RSABs, RCABs, and the fusion block; (b) is the projector to improve the nonlinear feature mapping ability; and (c) is the predictor, which is only assembled on one of the branches of the HyperNet and is swapped for another branch to alternatively minimize the similarity loss.

for hyperspectral change detection, which is mainly composed of three parts: 1) SSA module; 2) projector; and 3) predictor. The whole bitemporal HSIs are first processed by the SSA module, aiming to fully exploit the spatial relationship and the diverse spectral information. The projector further maps the extracted feature embeddings to acquire deeper and more abundant features. And the predictor which is only equipped with single branch plays a role of creating the prediction target for the other branch. The self-supervised training objective is to force the projected embeddings of one view to be as similar as possible with the prediction target of the other view on the selected mask area. Under the circumstance of absence of labeled data, the predetection strategy provides a coarse pseudo-label for supplement. A novel focal cos similarity loss is designed to boost the contribution of those hard samples for better training. The two branches of HyperNet share the same weight and is devoted to learning the aligned features for bitemporal HSIs without labor-cost labels, where the distance map of the learned features is then used to separate the unchanged from the changed area easily.

A. SSA Module

The SSA module is elaborated to focus on the spatial and spectral features separately. The spatial attention branch concentrates on the spatial relationship from the local and global scale while the spectral attention branch contributes to exploring the useful and informative spectral features. Then, the spatial and spectral features are adaptively fused together to acquire the advantaged information of each pixel.

1) *Spatial Attention Branch*: The spatial attention branch consists of three effective building blocks, the residual spatial attention block (RSAB), which is specially designed for spatial feature representation. The detailed architecture of RSAB is shown in Fig. 2(a). Take the HSI $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ as the input for example, the 3×3 2-D convolutional layer followed by the batch normalization (BN) layer; first, extract the shallow edge and line information within the small inceptive field.

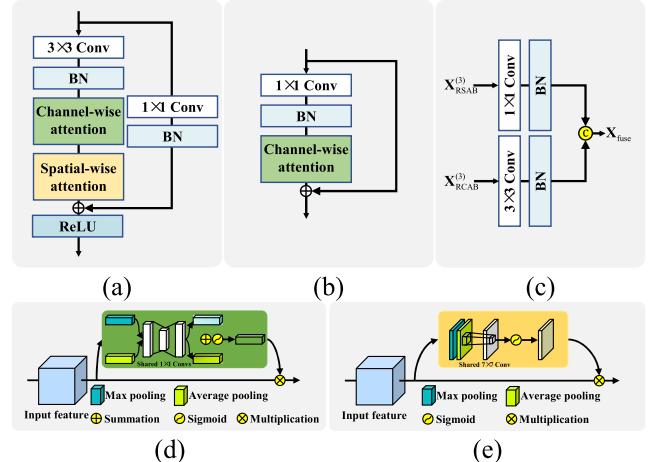


Fig. 2. Architecture of: (a) RSAB; (b) RCAB; (c) adaptive feature fusion block; (d) CA; and (e) SA.

The output can be denoted as $\mathbf{X}^{(1)} \in \mathbb{R}^{H \times W \times n}$ (the following equation), where n refers to the number of feature maps:

$$\mathbf{X}^{(1)} = \text{BN}(f^{3 \times 3}(\mathbf{X})). \quad (1)$$

For the purpose of concentration on the most informative feature, the convolutional block attention module (CBAM) [56] is incorporated after the convolution. The CBAM contains a channelwise attention (CA) [shown in Fig. 2(c)] and a spatialwise attention (SA) [shown in Fig. 2(d)], respectively. The CA squeezes the spatial information and stimulates the most discriminative spectral information from hundreds of spectral bands. Specifically, the global average and max pooling convert each channel of spatial map into a representative statistical value. Then, another two shared 1×1 2-D convolutional layers are targeted to gain inspired channel representations. And the summation of the extracted channel features is converted into a channel attention map \mathbf{M}_{ca} ranging from 0 to 1 by the sigmoid activation function σ . The mathematical definition of CA can be depicted as follows:

$$\begin{aligned} \mathbf{M}_{ca} = & \sigma(f^{1 \times 1}(f^{1 \times 1}(\text{AvgPool}(\mathbf{X}^{(1)}))) \\ & + f^{1 \times 1}(f^{1 \times 1}(\text{MaxPool}(\mathbf{X}^{(1)})))) \end{aligned} \quad (2)$$

where the bigger the weight is, the more attractive the spectral information of the input feature is. The stimulated channel attention feature \mathbf{F}_{ca} is obtained by elemental multiplication \otimes , which can be expressed as formula

$$\mathbf{F}_{ca} = \mathbf{M}_{ca} \otimes \mathbf{X}^{(1)}. \quad (3)$$

Analogously, as shown in Fig. 2(d), SA first reduces the channel number and then concentrates on how to exploit the most informative spatial information. The max pooling layer and average pooling layer of the SA focus on the edge and texture information, separately. The features extracted from the two pooling layers are further transformed by a 2-D convolutional layer with a kernel filter size as 7×7 . Then, the sigmoid activation function transforms the features into the range from 0 to 1 to get a weight constraint map

\mathbf{M}_{sa} , which is finally multiplied by the input features to get the output \mathbf{F}_{sa}

$$\mathbf{M}_{\text{sa}} = \sigma(f^{7 \times 7}([\text{AvgPool}(\mathbf{F}_{\text{ca}}); \text{MaxPool}(\mathbf{F}_{\text{ca}})])) \quad (4)$$

$$\mathbf{F}_{\text{sa}} = \mathbf{M}_{\text{sa}} \otimes \mathbf{F}_{\text{ca}}. \quad (5)$$

Noted that it is exactly the gate mechanism that controls the amount of information. The spatial information is emphasized at the location with the weight closer to 1 and suppressed at the area with the weight closer to 0.

In addition, the spatial attention block is designed in a residual architecture [57] to avoid being overfitted and convenient fusion of deep and shallow features. The input feature here is $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, and the output feature of the SA is $\mathbf{F}_{\text{sa}} \in \mathbb{R}^{H \times W \times n}$. The downsampling on the input is necessary when the input channel is not equal to the output channel. Concretely, a 1×1 2-D convolutional layer and BN are executed on the input to get $\mathbf{X}_d \in \mathbb{R}^{H \times W \times n}$. Thus, the residual output of the first RSAB is described as $\mathbf{X}_{\text{RSAB}}^{(1)} \in \mathbb{R}^{H \times W \times n}$

$$\mathbf{X}_d = \text{BN}(f^{1 \times 1}(\mathbf{X})) \quad (6)$$

$$\mathbf{X}_{\text{RSAB}}^{(1)} = \text{ReLU}(\mathbf{X}_d + \mathbf{F}_{\text{sa}}). \quad (7)$$

Noted that the number of channels of output feature map is set as n for each RSAB. As a result, only the first RSAB is equipped with the downsampling.

2) *Spectral Attention Branch*: For the spectral attention branch, three residual channel attention blocks (RCABs) are specially designed for spectral feature representation. As shown in Fig. 2(b), take the HSI $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ as the input for example, and the 2-D convolutional layer with kernel size 1×1 is selected for concentrating on the spectral space. Another BN layer is used to improve the generalization of the network. The output feature map $\mathbf{X}^{(2)} \in \mathbb{R}^{H \times W \times n}$ with the number set as n can be expressed as

$$\mathbf{X}^{(2)} = \text{BN}(f^{1 \times 1}(\mathbf{X})). \quad (8)$$

The RCAB is assembled with the CA which is tailored to stress the most useful spectral band and get the spectral attention map \mathbf{M}'_{ca} ranging from 0 to 1. The definition is defined as

$$\begin{aligned} \mathbf{M}'_{\text{ca}} = \sigma(f^{1 \times 1}(f^{1 \times 1}(\text{AvgPool}(\mathbf{X}^{(2)}))) \\ + f^{1 \times 1}(f^{1 \times 1}(\text{MaxPool}(\mathbf{X}^{(2)}))). \end{aligned} \quad (9)$$

For the RCAB, the short-cut connection is linked between the input and the output weighted spectral feature when they share the same size. Therefore, the output of first RCAB is exactly the weighted channel attention feature, which is acquired by the following formula. And we set the number of all feature maps acquired by the RCAB as n

$$\mathbf{X}_{\text{RCAB}}^{(1)} = \mathbf{M}'_{\text{ca}} \otimes \mathbf{X}^{(2)}. \quad (10)$$

Compared with those that only a branch is designed to extract both the spatial and spectral information simultaneously, we elaborated two branches focused on spatial and spectral information, separately, for high-dimensional hyperspectral cube. The spatial attention branch stacked three RSABs to extract the high-level semantic features, while the

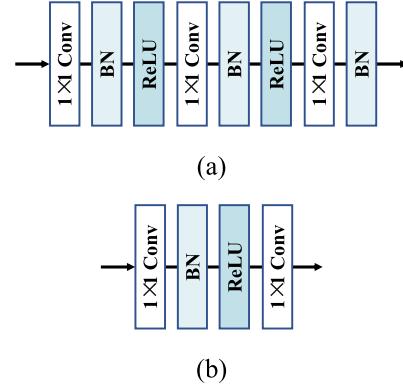


Fig. 3. Architecture of: (a) projector and (b) predictor.

spectral attention branch combined RCABs to extract discriminative spectral features. The spatial attention branch and the spectral attention branch are two independent branches and upgraded individually.

3) *Adaptive Feature Fusion Block*: After acquiring the advanced spatial features $\mathbf{X}_{\text{RSAB}}^{(3)}$ and spectral features $\mathbf{X}_{\text{RCAB}}^{(3)}$, an adaptive spatial-spectral feature fusion block is tailored to get the fused feature. As shown in Fig. 2(c), we adopt the 1×1 2-D convolutional layer and BN to process the extracted spatial features and 3×3 2-D convolutional layer and BN to process the spectral features. The spatial feature is further enhanced, and the spectral features are aggregated by the adaptive fusion block. And these two reorganized features are concatenated together along the channel axis to get the final fused features $\mathbf{X}_{\text{fuse}} \in \mathbb{R}^{H \times W \times (2n)}$, which can be described as follows:

$$\mathbf{X}_{\text{fuse}} = [\text{BN}(f^{1 \times 1}(\mathbf{X}_{\text{RSAB}}^{(3)})); \text{BN}(f^{3 \times 3}(\mathbf{X}_{\text{RCAB}}^{(3)}))]. \quad (11)$$

B. Siamese Projector and Predictor

The projector is designed to promote the extracted spatial and spectral features to get deeper embedding vectors. As presented in Fig. 3(a), the projector is comprised of three dense 1×1 convolutional layers accompanied with BN layers and rectified linear unit (ReLU) activation functions. Noted that these three 2-D convolution layers hold the same output channel number as $2n$.

The predictor with the stop-gradient operation is a key part for the self-supervised training. Only one of the branches of HyperNet is equipped with the predictor. Though the HyperNet learns to extract aligned features of bitemporal HSIs, it may induce identical but meaningless features to force the output features produced by the total Siamese network to be similar. As shown in Fig. 1, the output feature of the predictor in the first branch noted as \mathbf{P}_1 is forced to be similar with the output feature \mathbf{Z}_2 of the projector in the second branch. Meanwhile, no gradient is back propagated in the second branch, where only the HSI \mathbf{X} is responsible for the gradient back propagation. Likewise, the output feature of the projector \mathbf{Z}_1 in the first branch and the output feature \mathbf{P}_2 of the predictor in the second branch are forced to be similar, where

the gradient propagation is stopped in the first branch. The learned feature from predictor may differ from the one from projector. But the similarity constraint and the swap strategy facilitate the features to be more similar with each other alternatively, making the HyperNet learns meaningful aligned spatial-spectral features from the bitemporal HSIs.

Fig. 3(b) shows the structure of the predictor, which is composed of two dense 1×1 convolutional layers with BN as well as ReLU layer. Specifically, the first convolution layer outputs a squeezed feature map with channel number as n . And the final convolutional layer outputs the expanded feature map with channel number as $2n$. The projector and predictor share the same outputted feature map size.

C. Focal Cos Loss Function

Cos similarity distance is a commonly used measuring index to compare the similarity of two high-dimensional vectors. The more similar the two vectors are, the closer the cos similarity distance value is to 1. And the similarity value between the bitemporal vectors of large difference is close to 0 or even -1 . The natural imaging conditions may induce different spectral values over large areas. However, most of the spectral differences of the unchanged area are easy to be optimized. In other words, these positive samples are easy to be pulled together. Consequently, the large majority easy samples dominate during the training and the difficult samples are harder to be optimized. In order to solve the easy-sample-dominant problem, a novel similarity loss function named focal cos is proposed.

Given $\mathbf{Z}_1 \in \mathbb{R}^{H \times W \times (2n)}$ gained from the embedding space of first view and $\mathbf{P}_2 \in \mathbb{R}^{H \times W \times (2n)}$ from the prediction target of the other view, the proposed loss function named as focal cos \mathcal{L}_{fc} is defined as follows:

$$\mathcal{L}_{fc}(\mathbf{Z}_1, \mathbf{P}_2) = -(2 - \cos(\mathbf{Z}_1, \mathbf{P}_2)) \otimes \cos(\mathbf{Z}_1, \mathbf{P}_2) \quad (12)$$

$$\cos(\mathbf{Z}_1, \mathbf{P}_2) = \frac{\mathbf{Z}_1}{\|\mathbf{Z}_1\|_2} \cdot \frac{\mathbf{P}_2}{\|\mathbf{P}_2\|_2} \quad (13)$$

where we call the $(2 - \cos(\mathbf{Z}_1, \mathbf{P}_2))$ as an adjustive factor. For the hard samples, $\cos(\mathbf{Z}_1, \mathbf{P}_2) \rightarrow -1$, then the $(2 - \cos(\mathbf{Z}_1, \mathbf{P}_2)) \rightarrow 3$. In such case, the focal cos loss function greatly increases the weight of hard samples. As for the easy samples, $\cos(\mathbf{Z}_1, \mathbf{P}_2) \rightarrow 1$, $(2 - \cos(\mathbf{Z}_1, \mathbf{P}_2)) \rightarrow 1$. As a result, the weight of the easy samples keeps the same. As shown in Fig. 4, the blue line refers to the original cos similarity loss function and the red line represents the proposed focal cos loss function. For the original loss function painted in blue, the loss achieves the minimum when the two vectors are similar enough and hold zero angles. And proposed focal cos loss function shares the same minimum with the cos similarity loss function. Moreover, when two vectors hold a big angle, which is obviously a pair of hard positive samples, the proposed focal cos loss of these two vectors move up to a very high value compared with the original cos similarity loss. When the two vectors hold a small angle, which is more a pair of easy positive samples, the proposed focal cos loss of these two vectors drops and is lower than the value gained from the original cos similarity loss function.

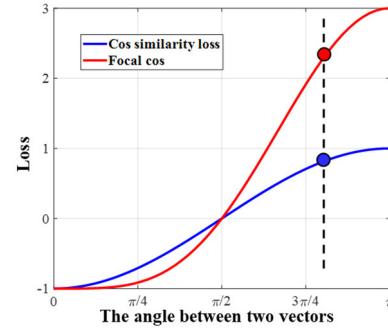


Fig. 4. Comparison of cos similarity loss function in blue and proposed focal cos loss function in red.

Considering extensive change may happen among multi-temporal HSIs, it heavily hampers the performance of the network to compute similarity loss for all pixels. We proposed a predetection strategy to get a pseudo-mask \mathbf{g} , where a limited number of the pixels with high probability to be unchanged are opted from the predetection result of a certain classic method.

The final objective function is defined as \mathcal{L}

$$\mathcal{L} = \frac{1}{2} \cdot \mathbf{g} \otimes (\mathcal{L}_{fc}(\text{StopGrad}(\mathbf{Z}_1), \mathbf{P}_2) + \mathcal{L}_{fc}(\text{StopGrad}(\mathbf{Z}_2), \mathbf{P}_1)). \quad (14)$$

The parameters of the whole network are upgraded by minimizing the loss function using gradient back propagation. And the proposed HyperNet is capable of learning powerful pixel-level deep features from the bitemporal HSIs. Moreover, the bitemporal features are aligned with each other in the feature space for those unchanged areas and scattered for the changed area. As a result, the learned features of the SSA module are quite appropriate for separating the unchanged from the changed. For the postprocessing, Diff-RX [58] is used as anomaly detector to detect anomalous changes for three HBCD datasets. And the cos similarity distance is adopted with K -means [59] that is employed for thresholding for binary change maps.

IV. EXPERIMENTS AND ANALYSIS

To test the effectiveness of the proposed method on hyperspectral change detection, extensive hyperspectral experiments have been conducted on HACD-task and HBCD. In this section, the descriptions of the datasets are first presented. Next, the details of the experimental setting are given. Then, the analyses of the experimental result tested on HACD-task and HBCD are exhibited, respectively.

A. Hyperspectral Datasets

The datasets used for HACD-task include the benchmark dataset “Viareggio 2013” [60] and another simulated HyMap dataset.

B. “Viareggio 2013” Dataset

In general, the “Viareggio 2013” dataset involves three HSIs, namely, D1F12H1, D1F12H2, and D2F22H2, as shown in Fig. 5. All of the three HSIs are shot in Viareggio, Italy,

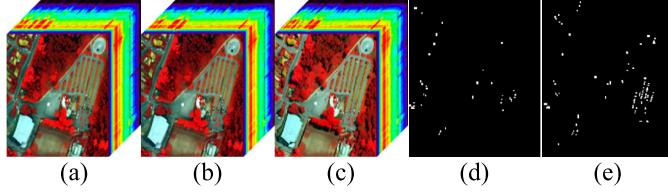


Fig. 5. (From left to right) Pseudo-color representation of “Viareggio 2013” dataset for HACD-task: (a) D1F12H1; (b) D1F12H2; and (c) D2F22H2 for “Viareggio 2013” dataset; (d) reference map of EX-1: D1F12H1-D1F12H2; and (e) reference map of EX-2: D1F12H1-D2F22H2.

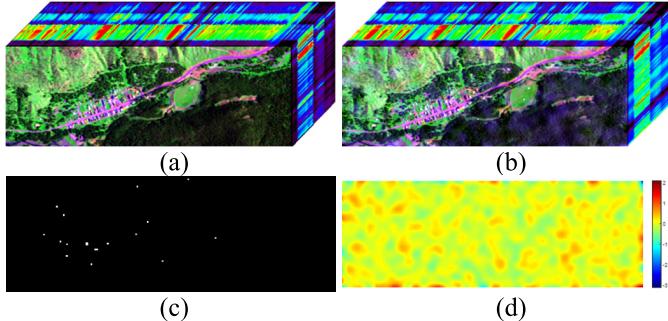


Fig. 6. Pseudo-color map of simulated HyMap dataset for HACD-task: (a) original HyMap image; (b) simulated HyMap image; (c) reference map; and (d) simulation noise map of random uniform distribution processed by a low-pass filter.

by an airborne hyperspectral sensor SIM.GA. The spectral information covers spectrum ranging from 400 to 1000 nm, with a spectral resolution as 1.2 nm approximately. And the spatial resolution is 0.6 m. The image sizes are all 450×375 , with 127 bands. The data are processed by destriped [61] and spectrally binned and available at <http://rsipg.dii.unipi.it/>. The D1F12H1 and D1F12H2 are acquired on the same day, May 8, 2013, making up the EX-1: D1F12H1-D1F12H2. And EX-2: D1F12H1-D2F22H2 is composed of D1F12H1 and D2F22H2, which is acquired on May 9, 2013. Two HSIs of EX-1 share very similar imaging conditions, and the anomalous changes mainly come from the movement of vehicles. But the case of EX-2 is quite different. There is severe shift between the gesture of the shooting platforms for EX-2, casting extensive distortion and great challenge for pixel-to-pixel anomalous change detection. Moreover, the weather condition of D2F22H2 has also changed compared with D1F12H1, where shadows appear around the trees and buildings.

1) Simulated HyMap Dataset: The simulated HyMap dataset (shown in Fig. 6) is designed to test the impact of noise on the proposed method; the first image is a hyperspectral target detection dataset acquired at Cooke City, MT, USA, on July 4, 2006. And it is available at <http://dirsapps.cis.rit.edu/blindtest/>. The image sizes are $280 \times 800 \times 126$. The spectrum ranges cover from 453 to 2486 nm. To simulate different imaging conditions, a noise map of random uniform distribution ranging from -10 to 10 is first created and then is processed by a low-pass filter with a standard deviation of ten pixels. The simulated HSI is created by an addition of the noise map and the original HyMap data. In addition, an offset of one

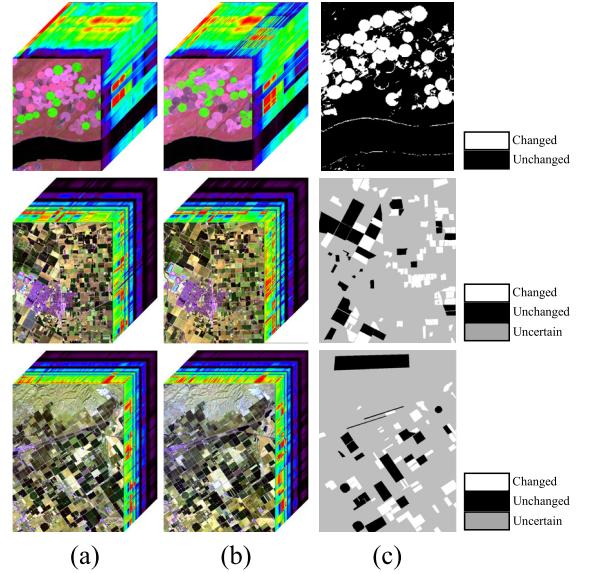


Fig. 7. (From top to bottom) Pseudo-color map of Hermiston dataset, Bay dataset, and Santa Barbara dataset for HBCD: (a) HSI acquired at time 1; (b) HSI acquired at time 2; and (c) reference map.

pixel on horizontal and vertical directions is applied to the simulated HSI. And the anomalous changes are obtained by the replacement of objects at other locations.

Moreover, we have also tested the validity of HyperNet on another three HBCD datasets.

a) Hermiston dataset: As the first row of Fig. 7 presents, two HSIs are acquired on May 1, 2004, and May 8, 2007 by Hyperion in Hermiston, OR, USA, respectively. And Fig. 7(c) shows the reference map of HBCD. The scene covers a wide range of irrigated fields, river, and cultivated land, with size as 307×241 pixels and 154 spectral bands.

b) Bay dataset: The second row of Fig. 7 gives 3-D pseudo-color cube representation of Bay dataset, which are taken on 2013 and 2015, individually, with the AVIRIS sensor surrounding the city of Patterson, CA, USA. Bay dataset is largely covered by farm lands and buildings, with a spatial size as 600×500 pixels and 224 spectral bands. Noted that only the labeled changed and unchanged area are adopted for assessment.

c) Santa Barbara dataset: The last one for HBCD is exhibited as the third row of Fig. 7, where (a) and (b) are shot on the years 2013 and 2014 with the AVIRIS sensor on the Santa Barbara region. The spatial dimensions are 984×740 pixels and both have 224 spectral bands. The two HSIs have recorded the urban evolution and dynamic changes of farmland, and HBCD provides a powerful tool to detect the accurate dynamics changes of urban development.

C. Implementation Details

1) Experimental Settings: We implemented our method by Pytorch and conducted experiments on a single NVIDIA RTX 3090 GPU. And the parameters of the network are initialized by He-normal way [62]. The optimizer is stochastic gradient descent (SGD) with momentum as 0.9 and L2 normalization efficient as 0.0001. We set the initial learning rate as

TABLE I

SELECTED SAMPLES FROM PREDETECTION RESULTS FOR SIX DATASETS

Dataset	Image size	Selected numbers	Ratio (%)	n
D1F12H1-D1F12H2	450*375	8192	4.85	64
D1F12H1-D2F22H2	450*375	8192	4.85	64
Simulated Hymap	280*800	8192	3.66	64
Hermiston	307*241	8192	11.07	72
Bay	600*500	8192	2.73	112
Santa Barbara	492*740	8192*2	4.50	112

0.05, and a cos decay [63] strategy is adopted. The number of total epochs for training is set 200. Since the proposed method is a full-convolutional neural network, the whole image is fed into the network without the need of batch size. As for the pre-detection strategy, Diff-RX is opted for predetection methods for HACD-task and change vector analysis (CVA) is opted for HBCD. The pseudo-samples are selected from the predetection result. The selected numbers of pseudo-samples for all datasets are listed in Table I. Noted that Santa Barbara dataset is segmented into two due to the large size of 984×740 for HyperNet. The parameter settings of feature num n are summarized in Table I.

2) *Comparison Methods*: To testify the effectiveness of HyperNet, several methods are opted for comparison. For HACD-task, another eight methods, namely, CC [31], unsupervised slow feature analysis (USFA) [64], Diff-RX [58], hyperbolic anomalous change detector (HACD) [11], simple difference hyperbolic anomalous change detector (SDHACD) [11], simple difference anomalous change detector (SDACD) [11], multi-temporal spatial: spectral comparison network for hyperspectral anomalous change detection (MTC-NET) (patch = 13), and MTC-NET (patch = 31) [65]. In order to stress the effect of the patch size on the performance of change detection, MTC-NET (patch = 13) with a patch size set as 13×13 and MTC-NET (patch = 31) with a patch size set as 31×31 are opted for comparison. For HBCD, another six algorithms, CVA, iterative slow feature analysis (ISFA) [66], self-supervised multisensor change detection (MSCD) [26], self-supervised change detection in multiview remote sensing images (OSCD) [25], MTC-NET (patch = 13), and MTC-NET (patch = 31) are selected for comparison. Noted that OSCD and MSCD are two patch-based SSL methods for multisensor change detection, and the latter has been tested on multispectral image change detection. The batch size of MSCD is set as 64 according to [26], while the OSCD adopts patch size as 16 for Bay and Santa Barbara datasets, and 8 for Hermiston dataset. Except for OSCD, the rest comparative deep-learning-based methods all set the batch size as 256. And OSCD employs the multiview contrastive loss of one positive sample and another $N - 1$ negative sample within a simple batch, where the batch size is set as N . Therefore, the batch size of OSCD is set as 1024. The details of parameter settings of various methods are presented in Table II.

3) *Evaluation Criteria*: Moreover, the receiver operating characteristic (ROC) and area under curve (AUC) are computed for quantitative evaluation of HACD-task, while overall accuracy (OA), Kappa, F1 score, Precision, and Recall rate

TABLE II
PARAMETER SETTINGS OF VARIOUS METHODS
FOR BOTH HACD-TASK AND HBCD

Method	Batch size	Patch size	Stride
MSCD	256	64	32
OSCD	1024	16	8
MTC-NET (patch size=13)	256	13	13
MTC-NET (patch size=31)	256	31	31

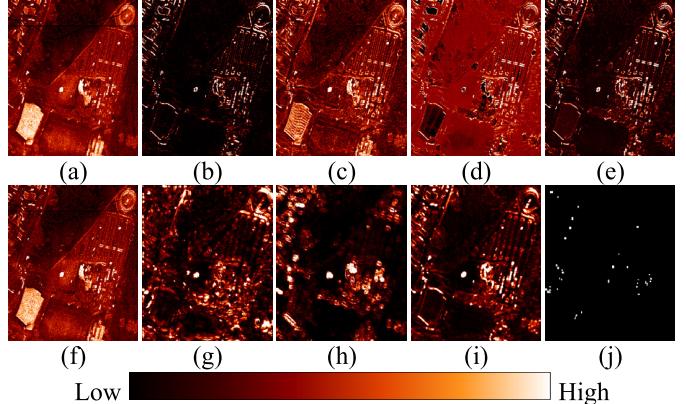


Fig. 8. Anomalous change detection maps of EX-1: D1F12H1-D1F12H2 on: (a) CC; (b) USFA; (c) Diff_RX; (d) HACD; (e) SDHACD; (f) SDACD; (g) MTC-NET (patch = 13); (h) MTC-NET (patch = 31); (i) proposed HyperNet; and (j) reference change map.

are calculated for the quantitative assessment of HBCD. The detection maps are segmented into four parts, true positive (TP), true negative (TN), false negative (FN), and false positive (FP).

D. HACD-Task Results

The anomalous change detection results of eight comparative methods and proposed HyperNet on EX-1: D1F12H1-D1F12H2 are represented in Fig. 8. The darker the detection map is, the more likely the area is unchanged and vice versa. It can be found that the backgrounds of USFA, SDHACD, MTC-NET (patch = 31), and proposed HyperNet are nearly in black, indicating a good performance on the compression of the background. However, for MTC-NET (patch = 31), the large input patch size induces the wide range of information shared and the decrease of location accuracy. As a result, there is some omissions of anomalous changes in the detection map of MTC-NET (patch = 31). Although a smaller input patch size reduces those effect, much more noise appears in the detection map of the MTC-NET (patch = 13). Based on the fully convolutional neural network, the proposed HyperNet balances the need of spatial information and location accuracy with the design of spatial and spectral attention module, where the local spatial information, pixel-level spectral information, and the attention from global scale are integrated. As shown in Fig. 8(i), the anomalous changes are almost highlighted and the edges of the change area are preserved in the detection map of HyperNet.

Fig. 9 shows the anomalous detection maps of all methods on EX-2: D1F12H1-D2F22H2. Due to the violent shift of the sensor, obvious brightness can be observed at the edge of the

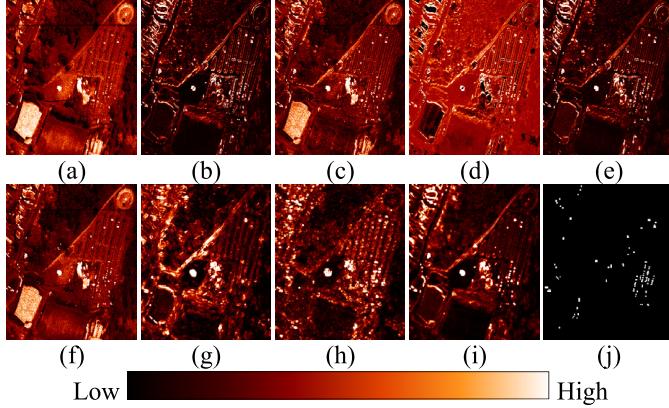


Fig. 9. Anomalous change detection maps of EX-2: D1F12H1-D2F22H2 on: (a) CC; (b) USFA; (c) Diff_RX; (d) HACD; (e) SDHACD; (f) SDACD; (g) MTC-NET (patch = 13); (h) MTC-NET (patch = 31); (i) proposed HyperNet; and (j) reference change map.

pine trees area in all results, extending from the top right to left middle. Most of the backgrounds of the results of USFA, HACD, and proposed HyperNet are at low value, showing nice performance on reducing the impact of pervasive spectral difference. Lots of noise can be observed on the results of MTC-NET (patch = 13) and MTC-NET (patch = 31). For EX-2, the anomalous changes mainly hail from the appearance and disappearance of the vehicles, and are concentrated in the right center, which is a parking lot. Compared with the reference change map, HACD, SDHACD, and HyperNet are capable of detecting nearly all anomalous changes, where the anomalous change values of the result of HyperNet seem to be brighter than the rest. In contrast with the maps of MTC-NET (patch = 13) and MTC-NET (patch = 31), there is no sharp edge expansion in the change region and the omission and noise are less on the result of HyperNet, indicating the better performance of the proposed pixel-level HyperNet. The two-branch SSA module is tailored to focus on spectral and spatial information, separately, where sufficient features of two distinct dimensions are abstracted.

The anomalous change detection results of all methods on the simulated Hymp dataset are presented in Fig. 10. Due to the offset on both horizontal and vertical directions, there are massive misplacements on the center of the image, which is a dense mixture of buildings, vehicles, and roads. All methods gain dense high values on the center area except for the HACD. Moreover, the noise also has a negative impact on detecting the anomalous changes. Obvious noise speckles can be found in the detection maps of all comparative methods except USFA. And the noise impact is magnified especially in the detection map of MTC-NET (patch = 31). However, the noise has little influence on the detection result of HyperNet, indicating strong noise resistance of the proposed method. In addition, HyperNet detects most simulated anomalous changes, gaining good performance on anomalous change detection.

Quantitative assessment is also provided to test the validity of the proposed method. The first column of Fig. 11 shows the ROC curves of all methods on three datasets, where the horizontal axis refers to the false alarm rate and the vertical

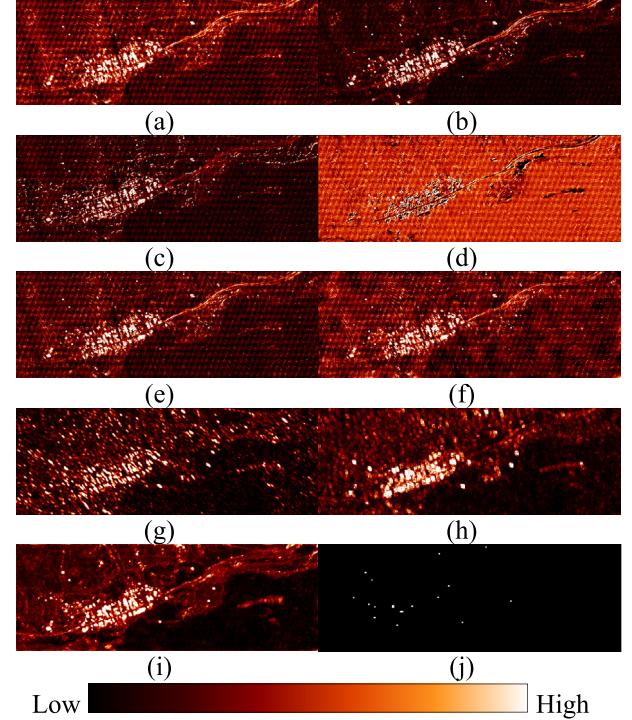


Fig. 10. Anomalous change detection maps of simulated Hymp dataset on: (a) CC; (b) USFA; (c) Diff_RX; (d) HACD; (e) SDHACD; (f) SDACD; (g) MTC-NET (patch = 13); (h) MTC-NET (patch = 31); (i) proposed HyperNet; and (j) reference change map.

axis is the probability of detection, respectively. And the method that gains high probability of detection at low false alarm rate is a terrific anomalous change detector. The ROC curve of the proposed HyperNet in red overtops the rest comparative methods for all three datasets. And the MTC-NET (patch = 13) in purple obtains a good performance on the EX-1: D1F12H1-D1F12H2 and simulated Hymp dataset; HACD in light green also acquires comparative effect on the EX-1: D1F12H1-D1F12H2 and EX-2: D1F12H1-D2F22H2.

The second column of Fig. 11 is the separability maps of all methods on three datasets. The output value of anomalous changes and background are collected, respectively, to test the ability of separating anomalous change and background. The values are normalized to [0, 1] for convenience of visualization. The box in orange refers to the anomalous change, while the box in green denotes the background. The top and bottom of each box represent the maximum and minimum, individually, and the main part of the box holds values from the smallest 25% to the largest 75%. For EX-1: D1F12H1-D1F12H2, USFA, SDHACD, and MTC-NET (patch = 31) can suppress the background values at low range, but their separability with the anomalous change values is less perfect. And there is larger intersection between the anomalous change box and background box of CC, Diff_RX, HACD, and SDACD. By contrast, MTC-NET (patch = 13) and HyperNet gain large gap between corresponding boxes, where the values of background of HyperNet fall into a lower range. The situation of EX-2: D1F12H1-D2F22H2 is very similar to EX-1. However, MTC-NET (patch = 13) and MTC-NET (patch = 31) both

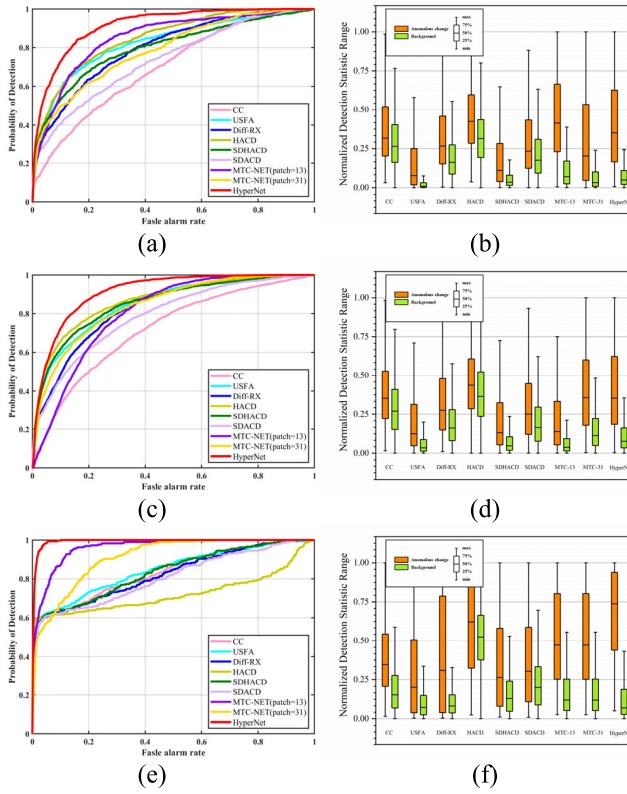


Fig. 11. First column is the ROC curves of all methods on: (a) EX-1: D1F12H1-D1F12H2; (c) EX-2: D1F12H1-D2F22H2; and (e) simulated Hymap dataset, and the second column is separability maps of all methods on: (b) EX-1: D1F12H1-D1F12H2; (d) EX-2: D1F12H1-D2F22H2; and (f) simulated Hymap dataset.

acquire larger background values, which intersect with corresponding anomalous change values to some extent. HyperNet gains good separability between the anomalous change box and background box. As for the simulated Hymap dataset, there is a large distance separating the anomalous change from the background for the result of HyperNet. It is observed that the result of USFA and Diff-RX hold low anomalous change values, but they are intersected with the anomalous change values set to a great degree. In addition, HACD gains low separability, which induces a dropped ROC curve shown in Fig. 11(e). Although it is a simulated dataset, the noise from a random uniform distribution processed with low-pass filtering and the offset on whole image cast great challenge for anomalous change detection. HyperNet obtains a good performance, discriminating the anomalous change with the complex background. Table III gives a comparison of AUC performance of all method on three datasets. The AUC means the area under the ROC value and is the comprehensive index for an anomalous change detector. The highest is in bold, and the second best is underlined. For all three datasets, HyperNet outperforms other methods with best AUC of 0.9147 for EX-1, 0.9126 for EX-2, and 0.9927 for Hymap dataset, which is much higher than the second best. HACD ranks second with AUC as 0.8489 for EX-1: D1F12H1-D1F12H2 and 0.8666 for EX-2: D1F12H1-D2F22H2. MTC-NET (patch = 13) gains

TABLE III
AUC COMPARISON OF ALL METHODS ON THREE HACD-TASK DATASETS

Method	EX-1	EX-2	Simulated Hymap
CC	0.6991	0.7196	0.8327
USFA	0.8220	0.8479	0.8437
Diff-RX	0.7993	0.8277	0.8249
HACD	<u>0.8489</u>	<u>0.8666</u>	0.7232
SDHACD	0.7978	0.8525	0.8343
SDACD	0.7372	0.7896	0.8069
MTC-NET (patch=13)	0.8346	0.8019	0.9589
MTC-NET (patch=31)	0.7776	0.8411	0.9276
HyperNet	0.9147	0.9126	0.9927

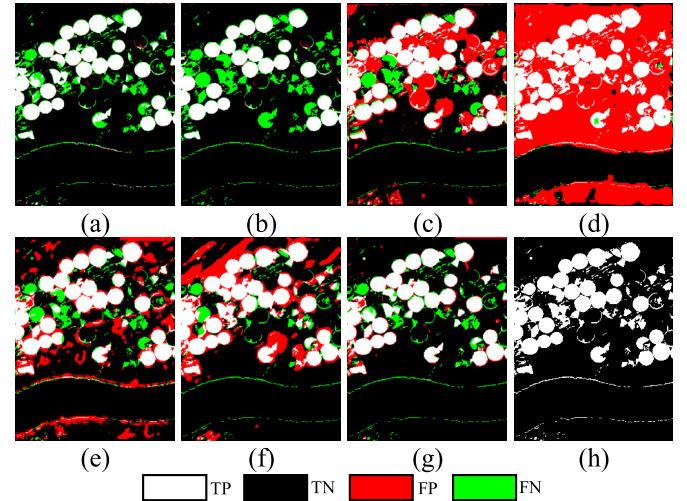


Fig. 12. Binary change detection maps on Hermiston dataset on of: (a) CVA; (b) ISFA; (c) MSCD, (d) OSCD; (e) MTC-NET (patch = 13); (f) MTC-NET (patch = 31); (g) proposed HyperNet; and (h) reference change map.

TABLE IV
QUANTITATIVE EVALUATION OF ALL METHODS ON HERMISTON DATASET

Method	OA	Kappa	F1	Precision	Recall
CVA	0.9272	0.7670	0.8103	0.9819	0.6898
ISFA	0.9023	0.6716	0.7262	0.9852	0.5750
MSCD	0.7851	0.4788	0.6201	0.5154	<u>0.7782</u>
OSCD	0.4332	0.1307	0.4344	0.2802	0.9656
MTC-NET (patch size=13)	0.7960	0.4741	0.6011	0.6111	0.6581
MTC-NET (patch size=31)	0.8343	0.5399	0.6468	0.6370	0.6733
HyperNet	<u>0.9206</u>	<u>0.7613</u>	0.8112	0.8740	0.7569

second best for simulated Hymap dataset with AUC equals 0.9589.

E. HBCD Results

The binary change detection results of all comparative methods and proposed HyperNet on the Hermiston dataset are shown in Fig. 12. Concretely, the TP is in white; TN is in black; FP is in red, and FN is in green. There are massive

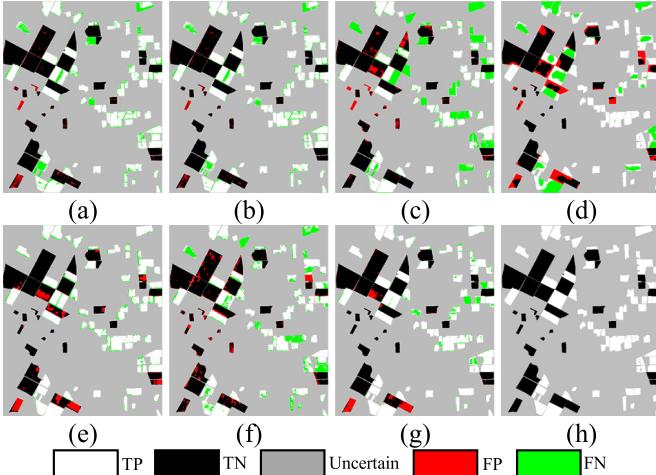


Fig. 13. Binary change detection maps on Bay dataset on of: (a) CVA; (b) ISFA; (c) MSCD; (d) OSCD; (e) MTC-NET (patch = 13); (f) MTC-NET (patch = 31); (g) proposed HyperNet; and (h) reference change map.

false detections in red in the detection map of the OSCD [Fig. 12(d)], which is consistent with the low precision as 0.2802 presented in Table IV. It is analyzed that large batch size is necessary for the training of OSCD while the Hermiston dataset is a small size image with 307×241 . And there are also some red false detections in the results of MSCD, MTC-NET (patch = 13), and MTC-NET (patch = 31), especially in the left bottom corner, where some buildings are located. And some missing detection in green can be found in the results of CVA and ISFA. By contrast, HyperNet is capable of detecting most of the changes with few false detections. Table IV lists the quantitative assessment of all method on Hermiston dataset. The maximum is in bold and the second largest is underlined. CVA gains the best OA and Kappa as 0.9272 and 0.7670, separately. HyperNet obtains the comparative OA as 0.9206 and Kappa as 0.7613. The highest F1 score is 0.8112 acquired by HyperNet. In addition, the ISFA gains the largest Precision and the CVA ranks second, indicating that the detected changes are almost really changed area. The OSCD detects all the changes with the cost of a low Kappa and Precision. On the whole, HyperNet gets the best comprehensive performance on Hermiston dataset. The proposed double-branch SSA module is targeted at extracting spatial objects relation and the discriminative spectral features, beneficial to reduce the false detection and improve the change detection, making a difference in the change detection result of Hermiston dataset.

Fig. 13 represents the binary change detection results on Bay dataset. The uncertain area which is unlabeled is in gray. MSCD and OSCD gain a little more omitted changes in green than other methods especially for the left farmland area. And these two methods are originally designed for multispectral remote sensing change detection, facing challenges dealing with the HSIs. Compared with the MTC-NET (patch = 31), more false detections in red can be found in the detection map of MTC-NET (patch = 13). However, MTC-NET (patch = 31) omits more changes in green than MTC-NET (patch = 13)

TABLE V
QUANTITATIVE EVALUATION OF ALL METHODS ON BAY DATASET

Method	OA	Kappa	F1	Precision	Recall
CVA	0.8723	0.7462	0.8708	<u>0.9474</u>	0.8057
ISFA	<u>0.8917</u>	<u>0.7848</u>	0.8905	0.9695	0.8234
MSCD	0.7503	0.5089	0.7260	0.8777	0.6190
OSCD	0.7264	0.4533	0.7339	0.7644	0.7065
MTC-NET (patch size=13)	0.8821	0.7627	<u>0.8910</u>	0.8807	<u>0.9016</u>
MTC-NET (patch size=31)	0.8108	0.6240	0.8082	0.8818	0.7459
HyperNet	0.9079	0.8152	0.9129	0.9224	0.9037

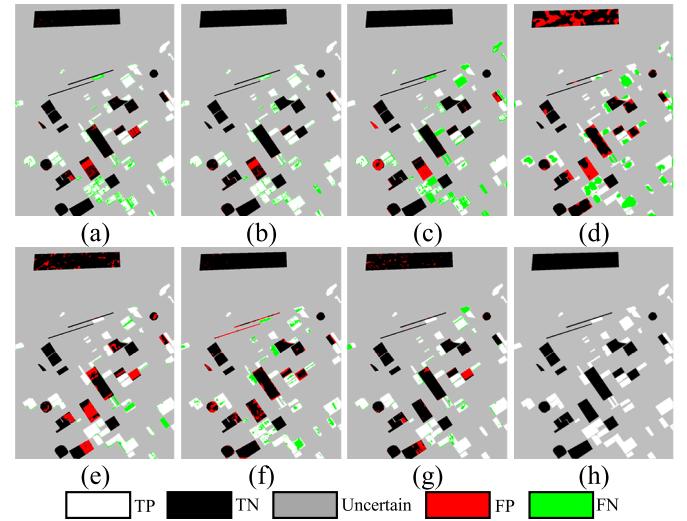


Fig. 14. Binary change detection maps on Santa Barbara dataset on of: (a) CVA; (b) ISFA; (c) MSCD; (d) OSCD; (e) MTC-NET (patch = 13); (f) MTC-NET (patch = 31); (g) proposed HyperNet; and (h) reference change map.

does. The selection of batch size plays an import role in the detection performance. By contrast, HyperNet with a patch-free architecture avoids this problem and gains better change detection result with less false alarms. Table V shows the quantitative evaluation results of Bay dataset. HyperNet acquires the best OA as 0.9079, Kappa as 0.8152, F1 score as 0.9129, and Recall rate as 0.9037 among all method, and outperforms the second largest one. And ISFA obtains the highest Precision as 0.9695 with least false detection.

The binary change detection results of Santa Barbara dataset are shown in Fig. 14. It is observed that plenty of false detections in red appears in the detection maps of OSCD and MTC-NET (patch = 13), corresponding to low Precision shown in Table VI. And obvious missing detection can be found in the results of CVA, MSCD, and OSCD, consistent with low recall rate shown in Table VI. HyperNet is able to detect most of the changes under low false detection rate. As Table VI presents, HyperNet gains the best OA as 0.9114, Kappa as 0.8148, F1 score as 0.8880, and Recall as 0.8925 among all comparative methods, while the ISFA gains the highest Precision with 0.9071. To conclude, HyperNet is

TABLE VI
QUANTITATIVE EVALUATION OF ALL METHODS ON
SANTA BARBARA DATASET

Method	OA	Kappa	F1	Precision	Recall
CVA	0.8780	0.7403	0.8376	0.8792	0.7997
ISFA	<u>0.8912</u>	<u>0.7675</u>	<u>0.8535</u>	0.9071	0.8059
MSCD	0.7868	0.5313	0.6872	0.8127	0.5952
OSCD	0.6869	0.3384	0.5914	0.6137	0.5751
MTC-NET (patch size=13)	0.8805	0.7550	<u>0.8566</u>	0.8288	<u>0.8920</u>
MTC-NET (patch size=31)	0.8886	0.7641	0.8539	0.8801	0.8301
HyperNet	0.9114	0.8148	0.8880	<u>0.8836</u>	0.8925

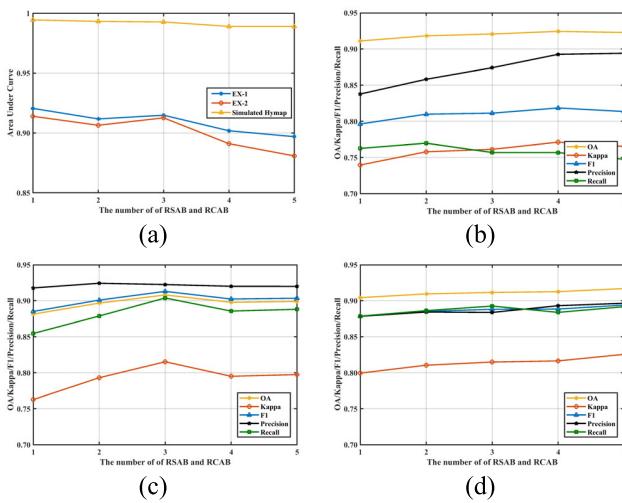


Fig. 15. HACD-task result and HBCD results regard to the number of RSAB and RCAB in the SSA module: (a) AUC performance for three datasets; the five quantitative evaluation indices (OA, Kappa, F1, Precision, and Recall) on the binary change detection results on: (b) Hermiston dataset; (c) Bay dataset; and (d) Santa Barbara dataset.

able to detect more precise change detection result, capturing subtle changes of the farmland in an SSL way.

V. DISCUSSION

A. Sensitivity and Stability Analysis of the Parameters

The depth of network is always a primary factor, which usually affects the balance between feature extraction ability, over-fitting problem, and the computation cost. The proposed SSA module is designed to dig the spatial relationship and the discriminative spectral features from the HSIs. To test the stability and sensitivity of the number of the RSAB and RCAB in the SSA module, experiments are conducted on both HBCD and HACD-task for all datasets.

Concretely, based on the proposed HyperNet, both the number of RSAB and RCAB are set as 1, 2, 3, 4, 5, respectively, for analysis. Fig. 15 shows the experimental results. It is observed that the AUC value decreases with the increase of the number of RSAB and RCAB for all three datasets of HACD-task [shown in Fig. 15(a)]. However, for the HBCD, as the number of RSAB and RCAB for Hermiston dataset increases

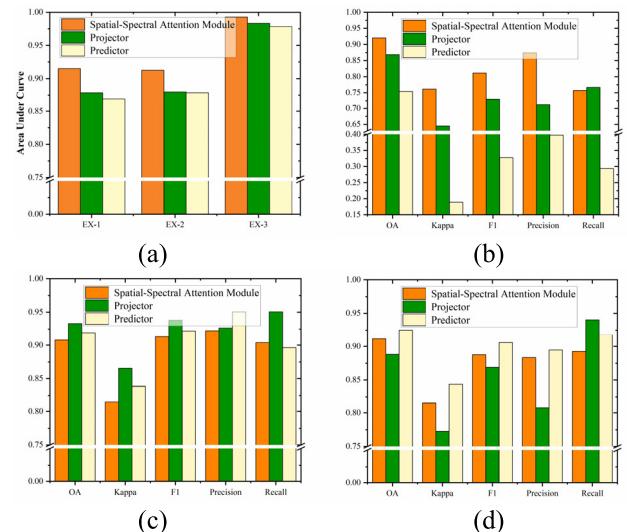


Fig. 16. Evaluation of the performance of the representation from SSA module, projector, and predictor for HACD-task and HBCD tasks: (a) AUC performance for three datasets of HACD-task; the five quantitative evaluation indices (OA, Kappa, F1, Precision, and Recall) of HBCD on: (b) Hermiston dataset; (c) Bay dataset; and (d) Santa Barbara dataset.

from 1 to 4, the OA, Kappa, F1, and Precision increase, and these four indicators drop when the number turns 5; the Recall rate decreases when the number increases from 2 to 5. For Bay dataset, the five indices rise initially with the increase of number from 1 to 3, followed by a drop when the number is larger than 3. For Santa Barbara dataset, the five indices gradually get bigger as the number climbs.

It is obvious that the performance with the number of RSAB and RCAB varies from HACD-task and HBCD. The proposed HyperNet is designed to learn feature representation by the semantic feature alignment of the selected pseudo-unchanged area of multitemporal HSIs. With the increase of the blocks of RSAB and RCAB, deeper features are acquired and the receptive field of the final feature map increases. Concretely, the anomalous changes are actually small and rare for HACD-task; thus, the selected pseudo-unchanged sample is randomly distributed throughout the image. More blocks lead to larger range of semantic feature alignment and even the area of anomalous changes, and more missing problems detected, which correspond to smaller AUC. For HBCD, the five indices increase first and then decline for both Hermiston and Bay datasets, and increase with the number of RSAB and RCAB for Santa Barbara dataset. It is likely that the deeper architecture promotes the semantic features alignment and especially for those selected pseudo-unchanged areas. However, the changed areas around the selected pseudo-unchanged area are involved with larger and larger receptive field, bringing about more unchanged areas and less changed areas, corresponding to falling recall rate for Hermiston and Bay datasets.

The number of blocks of RASB and RCAB plays an important role of balancing feature extraction ability and the over-fitting problem. The experimental results and analysis indicate that the number set as 3 is a suitable choice for comprehensive consideration of HACD-task and HBCD tasks.

TABLE VII

AUC COMPARISON OF HYPERNET WITH AND WITHOUT PROPOSED FOCAL COS LOSS ON THREE HACD-TASK DATASETS

Method	EX-1	EX-2	Simulated Hymap
Base	0.8562	0.8966	0.9886
Base + SSA	<u>0.9081</u>	<u>0.9112</u>	<u>0.9923</u>
HyperNet	0.9147	0.9126	0.9927

B. Evaluation of the Representations

We then study and compare the representations from SSA module, projector, and predictor on the performance of change detection. As shown in Fig. 16(a) and (b), it is observed that the representation of SSA module is better than projector and predictor for all three HACD-task datasets and the Hermiston dataset of HBCD. Concretely, 1%–4% increase of AUC value can be observed and sharp improvement of 47%–57% over Kappa, F1, Precision, and Recall. In addition, the representation of projector gets even a little better result for Bay dataset and the representation of predictor gains a bit better performance for Barbara dataset than the representation of SSA module does [shown in Fig. 16(c) and (d)]. On the whole, the representation of SSA module is able to achieve comparative even superior results than the representation of projector and predictor do.

Fig. 1 gives the exact visualization of the feature map learned from the SSA module, projector, and predictor, respectively. The HyperNet demands to learn feature representations invariant to the natural data augmentation, specifically the multitemporal HSI acquisition conditions. It is argued that the learned features from the projector and predictor can be less informative since the projected feature embeddings are forced to be as similar as the predicted targets. By contrast, the representation learned from SSA module maintains more sound and comprehensive information and Siamese feature patterns, which is more favorable for better change detection results. Therefore, the representation of SSA module is selected for postprocessing for both HACD-task and HBCD considering the overall performance and analysis.

C. Ablation Experiments of HyperNet

To test the performance of proposed SSA module and the focal cos loss function on change detection, extensive ablation experiments have been designed. Concretely, the “Base” is designed using basic normal 2-D convolutional layers with the cos similarity function. And “Base + SSA” denotes that the SSA module is adopted compared with “Base.”

Table VII gives AUC performance of ablation experiments of HyperNet on three HACD-task datasets. The largest one is in bold and the second largest is underlined. As shown in Table VII, the “Base + SSA” gains better AUC than “Base” does for all three datasets. And HyperNet acquires higher AUC than “Base + SSA” does. For intuitional representation of the validity of focal cos loss function, Fig. 17 exhibits the ROC comparison of the ablation experiments for three datasets. The “Base + SSA” in blue gains higher probability of detection

TABLE VIII

QUANTITATIVE COMPARISON OF HYPERNET WITH AND WITHOUT PROPOSED FOCAL COS LOSS ON THREE HBCD DATASETS

Dataset	Method	OA	Kappa	F1	Precision	Recall
Hermiston	Base	0.9140	0.7427	0.7969	0.8530	0.7489
	Base + SSA	<u>0.9197</u>	<u>0.7579</u>	<u>0.8083</u>	0.8751	<u>0.7516</u>
	HyperNet	0.9206	0.7613	0.8112	<u>0.8740</u>	0.7569
	Base	0.8605	0.7211	0.8649	0.8964	0.8355
Bay	Base + SSA	<u>0.9075</u>	<u>0.8144</u>	<u>0.9123</u>	0.9234	<u>0.9016</u>
	HyperNet	0.9079	0.8152	0.9129	<u>0.9224</u>	0.9037
	Base	0.8906	0.7730	0.8647	0.8441	0.8867
Barbara	Base + SSA	0.9155	0.8234	0.8934	0.8869	0.9001
	HyperNet	0.9114	0.8148	0.8880	0.8836	0.8925

than “Base” in black does. And more anomalous changes are highlighted in the results of “Base + SSA” (shown in the second row of Fig. 18) than the maps of “Base” (shown in the first row of Fig. 18), representing the effectiveness of the proposed SSA module. The ROC curve of HyperNet in red outperforms than the “Base + SSA” in blue for all three datasets (shown in Fig. 17), where the focal cos loss function is encouraged to put emphasis on hard positive samples promoting the training of network than cos similarity loss function does.

Moreover, the qualitative and quantitative evaluations of ablation experiments on HBCD have been implemented on another three datasets. As presented in Table VIII, with the SSA module, “Base + SSA” gains higher OA, Kappa, F1 score, Precision, and recall rate than “Base” does for all three datasets. And it is validated that less false alarm in red and less missing change in green can be found in the detection maps of “Base + SSA” (shown in the second row of Fig. 19) than that of “Base” (shown in the first row of Fig. 19). In addition, HyperNet with cos loss function gains better OA, Kappa, F1 score, and recall rate than “Base + SSA” does for Hermiston and Bay datasets. However, HyperNet with cos loss function gains better result on Barbara dataset than HyperNet with focal cos loss function does, where a little bit more false alarm in red can be found in the result of HyperNet (presented in third row of Fig. 19). The reason may be the hard-positive samples are in a low percentage and the focal cos loss function plays a less important role compared with the case of Hermiston and Bay datasets since the “Base” gains higher comprehensive Kappa as 0.7730 for Barbara dataset than for the Hermiston as 0.7427 and Bay as 0.7211.

Overall, HyperNet is capable of obtaining better change detection results, where the SSA module makes a difference in fully exploitation of the spatial relationship and the discriminative spectral features, and the focal cos plays an important role in amplifying the attention on area with complex spectral difference.

D. Runtime Cost Analysis

To test the runtime efficiency of all methods, Table IX shows the runtime cost comparison of eight comparative algorithms

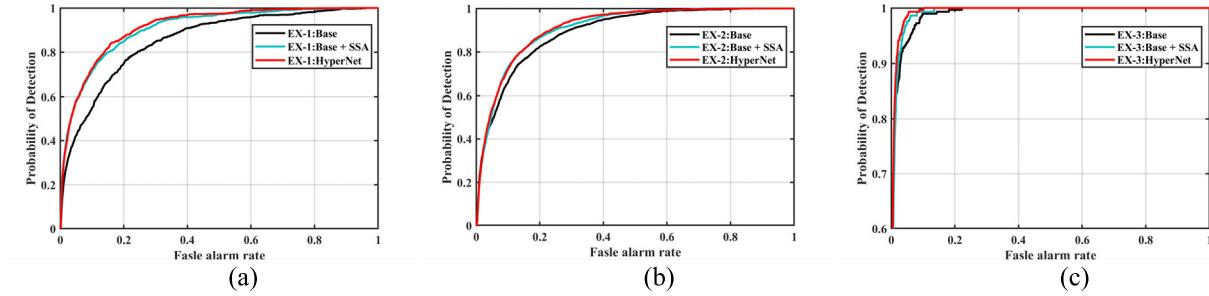


Fig. 17. ROC comparison of ablation experiments of HyperNet on three HACD-task datasets of: (a) EX1: D1F12H1-D1F12H2; (b) EX-2: D1F12H1-D2F22H2; and (c) simulated Hymap dataset.

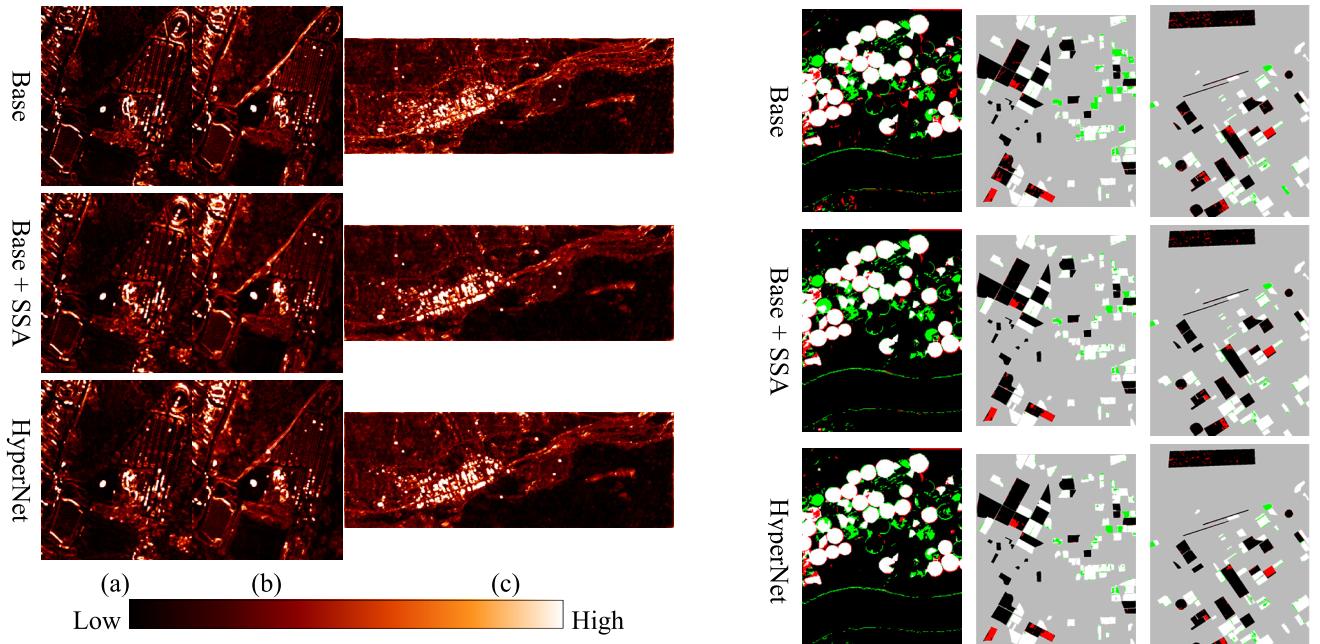


Fig. 18. Anomalous change detection map comparison of ablation experiments of: (a) EX1: D1F12H1-D1F12H2; (b) EX-2: D1F12H1-D2F22H2; and (c) simulated Hymap dataset. (From top to bottom) Result from Base, Base + SSA, and HyperNet.

TABLE IX
RUNTIME COST OF HYPERSPECTRAL ANOMALY CHANGE DETECTION
ON THREE DATASETS (UNIT: SECOND)

Method	EX-1	EX-2	Simulated Hymap
CC	6.4648	6.1926	4.7479
USFA	6.2399	5.7220	4.2144
Diff-RX	5.2346	5.2716	3.9411
HACD	5.8634	5.5663	4.0575
SDHACD	5.2340	5.2752	3.5482
SDACD	<u>5.1161</u>	<u>5.2610</u>	<u>3.4776</u>
MTC-NET (patch=13)	325.5863	323.9787	377.7896
MTC-NET (patch=31)	357.7804	711.4080	1114.5716
HyperNet	53.1382	53.4293	55.0648

and HyperNet on three datasets of HACD-task. The highest time cost is in bold, and the lowest time cost is underlined.

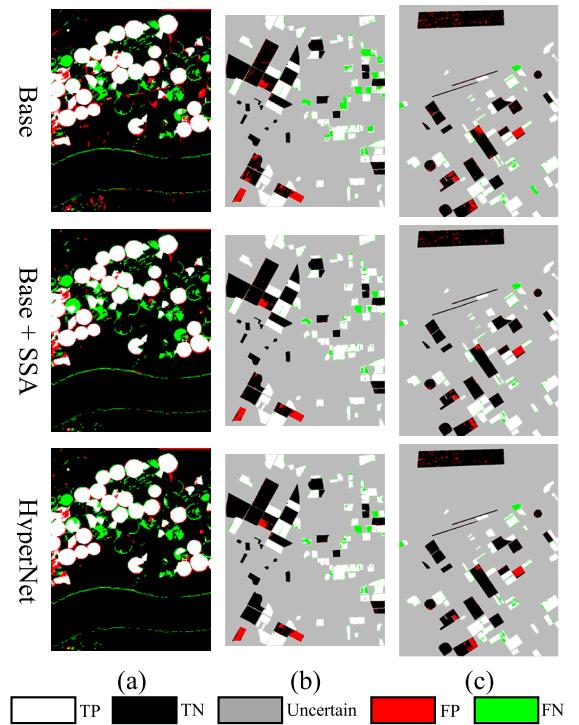


Fig. 19. Comparison of binary change detection maps of ablation experiments: (a) Hermiston dataset; (b) Bay dataset; and (c) Santa Barbara dataset.

It is obvious that the MTC-NET (patch = 31) suffers from largest time consumption than the others for all three datasets. And the SDACD gains the smallest time cost for HACD-task. In addition, the first six traditional approaches all run fast for all three datasets. The HyperNet spends around 55 s for all three datasets, which is acceptable for greater performance of HACD-task.

Table X exhibits the runtime cost of all approaches for three datasets on HBCD task. The OSCD runs much slower than the other methods for three datasets for large consumption of sample CL. And CVA obtains the smallest time expense for all datasets. Although the HyperNet takes a little bit more time than the CVA does, HyperNet runs fastest among the last five deep-learning based methods. Thus, the increase in time cost is acceptable in view of better comprehensive change detection result.

TABLE X
RUNTIME COST OF HBCD ON THREE DATASETS (UNIT: SECOND)

Method	Hermiston	Bay	Barbara
CVA	4.0485	9.2552	24.2482
ISFA	5.2825	18.9484	41.3809
MSCD	258.1106	1230.0600	3395.0281
OSCD	577.4696	3485.8024	9934.2895
MTC-NET (patch=13)	213.2759	602.3598	1291.5176
MTC-NET (patch=31)	489.0661	658.9878	1200.4907
HyperNet	36.8909	143.7598	261.8970

VI. CONCLUSION

In this article, we have attempted to put forward a self-supervised spatial-spectral understanding network for hyperspectral change detection. HyperNet employs an SSL mode and accomplishes multitemporal spatial and spectral feature comparison in fully convolutional way for pixelwise feature representation learning. The designed spatial attention branch focuses on the spatial correlation from 2-D imaging space, while the spectral attention branch concentrates on the discriminative spectral features for various ground surface objects. And a novel similarity loss function named focal cos is created for better adjustment of the imbalanced easy and hard positive samples in SSL training. The performance of extensive experiments on six hyperspectral datasets demonstrates the effectiveness and generalization of the proposed HyperNet on both HACD-task and HBCD tasks. In addition, it is proven that the designed SSA branch and focal cos loss function are capable of detecting more anomalous changes for HACD-task, reducing the false detection and improve the detection rate for HBCD.

REFERENCES

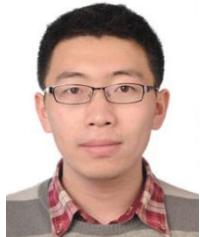
- [1] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, Jun. 2004, doi: [10.1080/0143116031000139863](https://doi.org/10.1080/0143116031000139863).
- [2] R. S. Lunetta, J. F. Knight, J. Ediriywickrema, J. G. Lyon, and L. D. Worthy, "Land-cover change detection using multi-temporal MODIS NDVI data," *Remote Sens. Environ.*, vol. 105, no. 2, pp. 142–154, 2006, doi: [10.1016/j.rse.2006.06.018](https://doi.org/10.1016/j.rse.2006.06.018).
- [3] Z. Zhu, "Change detection using Landsat time series: A review of frequencies, preprocessing, algorithms, and applications," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 370–384, Aug. 2017, doi: [10.1016/j.isprsjprs.2017.06.013](https://doi.org/10.1016/j.isprsjprs.2017.06.013).
- [4] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989, doi: [10.1080/01431168908903939](https://doi.org/10.1080/01431168908903939).
- [5] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, Mar. 2005, doi: [10.1109/TIP.2004.838698](https://doi.org/10.1109/TIP.2004.838698).
- [6] W. Dong, T. Zhang, J. Qu, S. Xiao, J. Liang, and Y. Li, "Laplacian pyramid dense network for hyperspectral pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, doi: [10.1109/TGRS.2021.3076768](https://doi.org/10.1109/TGRS.2021.3076768).
- [7] W. Dong, S. Hou, S. Xiao, J. Qu, Q. Du, and Y. Li, "Generative dual-adversarial network with spectral fidelity and spatial enhancement for hyperspectral pansharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 10, 2021, doi: [10.1109/TNNLS.2021.3084745](https://doi.org/10.1109/TNNLS.2021.3084745).
- [8] L. Zhang and B. Du, "Recent advances in hyperspectral image processing," *Geo-Spatial Inf. Sci.*, vol. 15, no. 3, pp. 143–156, Sep. 2012, doi: [10.1080/10095020.2012.719684](https://doi.org/10.1080/10095020.2012.719684).
- [9] P. Ghamisi et al., "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017, doi: [10.1109/MGRS.2017.2762087](https://doi.org/10.1109/MGRS.2017.2762087).
- [10] J. Theiler, "Quantitative comparison of quadratic covariance-based anomalous change detectors," *Appl. Opt.*, vol. 47, no. 28, p. F12, Oct. 2008, doi: [10.1364/AO.47.000F12](https://doi.org/10.1364/AO.47.000F12).
- [11] N. Acito, M. Diani, G. Corsini, and S. Resta, "Introductory view of anomalous change detection in hyperspectral images within a theoretical Gaussian framework," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 32, no. 7, pp. 2–27, Jul. 2017, doi: [10.1109/MAES.2017.160155](https://doi.org/10.1109/MAES.2017.160155).
- [12] Q. Du, N. Younan, and R. King, "Change analysis for hyperspectral imagery," in *Proc. Int. Workshop Anal. Multi-Temporal Remote Sens. Images*, Jul. 2007, pp. 1–4, doi: [10.1109/MULTITEMP.2007.4293052](https://doi.org/10.1109/MULTITEMP.2007.4293052).
- [13] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019, doi: [10.1109/MGRS.2019.2898520](https://doi.org/10.1109/MGRS.2019.2898520).
- [14] J. Shi, Z. Zhang, T. Wu, X. Zhang, and Y. Lei, "Collaborative self-perception network architecture for hyperspectral image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: [10.1109/LGRS.2022.3157916](https://doi.org/10.1109/LGRS.2022.3157916).
- [15] C. Shi, Z. Zhang, W. Zhang, C. Zhang, and Q. Xu, "Learning multiscale temporal-spatial-spectral features via a multipath convolutional LSTM neural network for change detection with hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022, doi: [10.1109/TGRS.2022.3176642](https://doi.org/10.1109/TGRS.2022.3176642).
- [16] Y. Yang, J. Qu, S. Xiao, W. Dong, Y. Li, and Q. Du, "A deep multiscale pyramid network enhanced with spatial-spectral residual attention for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, doi: [10.1109/TGRS.2022.3161386](https://doi.org/10.1109/TGRS.2022.3161386).
- [17] T. Bai et al., "Deep learning for change detection in remote sensing: A review," *Geo-Spatial Inf. Sci.*, early access, pp. 1–27, Jul. 2022, doi: [10.1080/10095020.2022.2085633](https://doi.org/10.1080/10095020.2022.2085633).
- [18] L. Wang, L. Wang, Q. Wang, and L. Bruzzone, "RSCNet: A residual self-calibrated network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022, doi: [10.1109/TGRS.2022.3177478](https://doi.org/10.1109/TGRS.2022.3177478).
- [19] J. Qu, S. Hou, W. Dong, Y. Li, and W. Xie, "A multilevel encoder-decoder attention network for change detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, doi: [10.1109/TGRS.2021.3130122](https://doi.org/10.1109/TGRS.2021.3130122).
- [20] L. Ericsson, H. Gouk, and T. M. Hospedales, "Why do self-supervised models transfer? Investigating the impact of invariance on downstream tasks," 2021, *arXiv:2111.11398*.
- [21] L. Ericsson, H. Gouk, and T. M. Hospedales, "How well do self-supervised models transfer?" in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 5410–5419, doi: [10.1109/CVPR46437.2021.00537](https://doi.org/10.1109/CVPR46437.2021.00537).
- [22] Y. Wang, C. M. Albrecht, N. A. A. Braham, L. Mou, and X. X. Zhu, "Self-supervised learning in remote sensing: A review," 2022, *arXiv:2206.13188*.
- [23] Y. Tian, C. Sun, B. Poole, D. Krishnan, C. Schmid, and P. Isola, "What makes for good views for contrastive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6827–6839.
- [24] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A survey on contrastive self-supervised learning," *Technologies*, vol. 9, no. 1, p. 2, Dec. 2020, doi: [10.3390/technologies9010002](https://doi.org/10.3390/technologies9010002).
- [25] Y. Chen and L. Bruzzone, "Self-supervised change detection in multiview remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022, doi: [10.1109/TGRS.2021.3089453](https://doi.org/10.1109/TGRS.2021.3089453).
- [26] S. Saha, P. Ebel, and X. X. Zhu, "Self-supervised multisensor change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–10, 2022, doi: [10.1109/TGRS.2021.3109957](https://doi.org/10.1109/TGRS.2021.3109957).
- [27] N. Acito, M. Diani, A. Rossi, and S. U. De Ceglie, "Hyperspectral anomalous change detection in the presence of non-stationary atmospheric/illumination conditions," in *Proc. SPIE*, Amsterdam, The Netherlands, Oct. 2014, p. 92440, doi: [10.1117/12.2067314](https://doi.org/10.1117/12.2067314).
- [28] S. Resta, N. Acito, M. Diani, G. Corsini, T. Opsahl, and T. V. Haavardsholm, "Detection of small changes in airborne hyperspectral imagery: Experimental results over urban areas," in *Proc. 6th Int. Workshop Anal. Multi-Temporal Remote Sens. Images (Multi-Temp)*, Jul. 2011, pp. 5–8, doi: [10.1109/Multi-Temp.2011.6005033](https://doi.org/10.1109/Multi-Temp.2011.6005033).
- [29] J. Theiler and S. Perkins, "Proposed framework for anomalous change detection," in *Proc. ICML Workshop Mach. Learn. Algorithms Survill. Event Detection*, 2006, p. 8.
- [30] P. Schaum and A. Stocker, "Hyperspectral change detection and supervised matched filtering based on covariance equalization," in *Proc. Defense Secur.*, Orlando, FL, USA, Aug. 2004, p. 77, doi: [10.1117/12.544026](https://doi.org/10.1117/12.544026).

- [31] S. Alan and S. Alan, "Long-interval chronochrome target detection," in *Proc. Int. Symp. Spectral Sens. Res.*, 1998, pp. 1760–1770.
- [32] M. T. Eismann, J. Meola, and R. C. Hardie, "Hyperspectral change detection in the presence of diurnal and seasonal variations," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 237–249, Jun. 2008, doi: [10.1109/TGRS.2007.907973](https://doi.org/10.1109/TGRS.2007.907973).
- [33] M. Hu, C. Wu, L. Zhang, and B. Du, "Hyperspectral anomaly change detection based on autoencoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3750–3762, 2021, doi: [10.1109/JSTARS.2021.3066508](https://doi.org/10.1109/JSTARS.2021.3066508).
- [34] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Review of digital change detection methods in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, 2004, doi: [10.1080/0143116031000101675](https://doi.org/10.1080/0143116031000101675).
- [35] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, Jun. 2013, doi: [10.1016/j.isprsjprs.2013.03.006](https://doi.org/10.1016/j.isprsjprs.2013.03.006).
- [36] Z. Hou, W. Li, L. Li, R. Tao, and Q. Du, "Hyperspectral change detection based on multiple morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5507312.
- [37] Z. Hou, W. Li, R. Tao, and Q. Du, "Three-order tucker decomposition and reconstruction detector for unsupervised hyperspectral change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6194–6205, 2021, doi: [10.1109/JSTARS.2021.3088438](https://doi.org/10.1109/JSTARS.2021.3088438).
- [38] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jan. 2016, doi: [10.1109/MGRS.2016.2540798](https://doi.org/10.1109/MGRS.2016.2540798).
- [39] W. Shi, M. Zhang, R. Zhang, S. Chen, and Z. Zhan, "Change detection based on artificial intelligence: State-of-the-art and challenges," *Remote Sens.*, vol. 12, no. 10, p. 1688, May 2020, doi: [10.3390/rs12101688](https://doi.org/10.3390/rs12101688).
- [40] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019, doi: [10.1109/TGRS.2018.2849692](https://doi.org/10.1109/TGRS.2018.2849692).
- [41] J. Lei, M. Li, W. Xie, Y. Li, and X. Jia, "Spectral mapping with adversarial learning for unsupervised hyperspectral change detection," *Neurocomputing*, vol. 465, pp. 71–83, Nov. 2021, doi: [10.1016/j.neucom.2021.08.130](https://doi.org/10.1016/j.neucom.2021.08.130).
- [42] Q. Li et al., "Unsupervised hyperspectral image change detection via deep learning self-generated credible labels," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9012–9024, 2021, doi: [10.1109/JSTARS.2021.3108777](https://doi.org/10.1109/JSTARS.2021.3108777).
- [43] L. Wang, L. Wang, Q. Wang, and P. M. Atkinson, "SSA-SiamNet: Spectral-spatial-wise attention-based Siamese network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2022, doi: [10.1109/TGRS.2021.3095899](https://doi.org/10.1109/TGRS.2021.3095899).
- [44] C. Wu, B. Du, and L. Zhang, "Fully convolutional change detection framework with generative adversarial network for unsupervised, weakly supervised and regional supervised change detection," 2022, *arXiv:2201.06030*.
- [45] C. Tao, J. Qia, G. Zhang, Q. Zhu, W. Lu, and H. Li, "TOV: The original vision model for optical remote sensing image understanding via self-supervised learning," 2022, *arXiv:2204.04716*.
- [46] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 4037–4058, Nov. 2021, doi: [10.1109/TPAMI.2020.2992393](https://doi.org/10.1109/TPAMI.2020.2992393).
- [47] S. Azizi et al., "Big self-supervised models advance medical image classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 3458–3468, doi: [10.1109/ICCV48922.2021.00346](https://doi.org/10.1109/ICCV48922.2021.00346).
- [48] T. Han, W. Xie, and A. Zisserman, "Self-supervised co-training for video representation learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 5679–5690.
- [49] Y. Li, D. Huang, D. Qin, L. Wang, and B. Gong, "Improving object detection with selective self-supervised self-training," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2020, pp. 589–607, doi: [10.1007/978-3-030-58526-6_35](https://doi.org/10.1007/978-3-030-58526-6_35).
- [50] C. Doersch, A. Gupta, and A. A. Efros, "Unsupervised visual representation learning by context prediction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1422–1430, doi: [10.1109/ICCV.2015.167](https://doi.org/10.1109/ICCV.2015.167).
- [51] P. Sermanet, C. Lynch, J. Hsu, and S. Levine, "Time-contrastive networks: Self-supervised learning from multi-view observation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 486–487, doi: [10.1109/CVPRW.2017.69](https://doi.org/10.1109/CVPRW.2017.69).
- [52] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 9726–9735, doi: [10.1109/CVPR42600.2020.00975](https://doi.org/10.1109/CVPR42600.2020.00975).
- [53] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. 37th Int. Conf. Mach. Learn.*, Nov. 2020, pp. 1597–1607.
- [54] J.-B. Grill et al., "Bootstrap your own latent—A new approach to self-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 21271–21284.
- [55] X. Chen and K. He, "Exploring simple Siamese representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 15745–15753, doi: [10.1109/CVPR46437.2021.01549](https://doi.org/10.1109/CVPR46437.2021.01549).
- [56] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
- [57] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [58] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
- [59] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-means clustering algorithm," *J. Roy. Stat. Soc. C, Appl. Statist.*, vol. 28, no. 1, pp. 100–108, 1979, doi: [10.2307/2346830](https://doi.org/10.2307/2346830).
- [60] N. Acito, S. Matteoli, A. Rossi, M. Diani, and G. Corsini, "Hyperspectral airborne 'Viareggio 2013 Trial' data collection for detection algorithm assessment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2365–2376, Jun. 2016, doi: [10.1109/JSTARS.2016.2531747](https://doi.org/10.1109/JSTARS.2016.2531747).
- [61] N. Acito, M. Diani, and G. Corsini, "Residual striping reduction in hyperspectral images," in *Proc. 17th Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2011, pp. 1–7, doi: [10.1109/ICDSP.2011.6005002](https://doi.org/10.1109/ICDSP.2011.6005002).
- [62] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1026–1034, doi: [10.1109/ICCV.2015.123](https://doi.org/10.1109/ICCV.2015.123).
- [63] M. Caron, P. Bojanowski, J. Mairal, and A. Joulin, "Unsupervised pre-training of image features on non-curated data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2959–2968.
- [64] C. Wu, L. Zhang, and B. Du, "Hyperspectral anomaly change detection with slow feature analysis," *Neurocomputing*, vol. 151, pp. 175–187, Mar. 2015, doi: [10.1016/j.neucom.2014.09.058](https://doi.org/10.1016/j.neucom.2014.09.058).
- [65] M. Hu, C. Wu, and B. Du, "Multi-temporal spatial-spectral comparison network for hyperspectral anomalous change detection," 2022, *arXiv:2205.11395*.
- [66] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2858–2874, May 2014, doi: [10.1109/TGRS.2013.2266673](https://doi.org/10.1109/TGRS.2013.2266673).



Meiqi Hu (Graduate Student Member, IEEE) received the B.S. degree in surveying and mapping engineering from the School of Geoscience and Info-Physics, Central South University, Changsha, China, in 2019. She is currently pursuing the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

Her research interests include deep learning and multitemporal remote sensing image change detection.



Chen Wu (Member, IEEE) received the B.S. degree in surveying and mapping engineering from Southeast University, Nanjing, China, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2015.

He is currently a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research interests include multitemporal remote sensing image change detection and analysis in multispectral and hyperspectral images.



Liangpei Zhang (Fellow, IEEE) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He was a Principal Scientist for the China State Key Basic Research Project from 2011 to 2016 appointed by the Ministry of National Science and Technology of China to lead the Remote Sensing Program in China. He is currently a Chair Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University. He has authored more than 700 research articles and five books. He is the Institute for Scientific Information (ISI) highly cited author. He holds 30 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a fellow of the Institution of Engineering and Technology (IET). He was a recipient of the 2010 Best Paper Boeing Award, the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing (ASPRS), and 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics (SPIE). His research teams won the top three prizes of the IEEE Geoscience and Remote Sensing Society (GRSS) 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) Student Paper Contest in recent years. He is the Founding Chair of the IEEE GRSS Wuhan Chapter. He also serves as an associate editor or an editor for more than ten international journals. He serves as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.