Change Detection From Synthetic Aperture Radar Images Based on Channel Weighting-Based Deep Cascade Network

Yunhao Gao, Feng Gao, Junyu Dong, and Shengke Wang

***Abstract*—Deep learning methods have recently demonstrated their significant capability for synthetic aperture radar (SAR) image change detection. However, with the increase of network depth, convolutional neural networks often encounter some negative effects, such as overfitting and exploding gradients. In addition, the existing deep networks employed in SAR change detection tend to produce a lot of redundant features that affect the performance of the network. To solve the aforementioned problems, this article proposed a deep cascade network (DCNet) for SAR image change detection. On the one hand, a very DCNet is established to exploit discriminative features, and residual learning is introduced to solve the exploding gradients problem. In addition, a fusion mechanism is employed to combine the outputs of different hierarchical layers to further alleviate the exploding gradient problem. Moreover, a simple yet effective channel weighting-based module is designed for SAR change detection. Average pooling and max pooling are used to aggregate channel-wise information. Meaningful channel-wise features are emphasized and unnecessary ones are suppressed. Therefore, the similarity in feature maps can be reduced, and then, the classification performance of the DCNet is improved. Experimental results on four real SAR datasets demonstrated that the proposed DCNet can obtain better change detection performance than several competitive methods. Our codes are available at** `https://github.com/summitgao/SAR_CD_DCNet`**.**

***Index Terms*—Change detection, deep cascade network (DCNet), deep learning, residual learning, synthetic aperture radar (SAR).**

I. Introduction

WITH the rapid development of earth observation programs, many synthetic aperture radar (SAR) sensors have been developed for spaceborne systems. A great number of SAR images captured at different times over the same geographical area are available. These images are increasingly important for many scientific applications, such as change detection [1], disaster monitoring [2], urban planning [3], etc. Among these applications, change detection has become an important topic in the SAR community.

SAR change detection is to detect the changed information of the same geographical area by analyzing two images captured in different periods. Since SAR sensors have the capabilities of all-weather and all-time observation, they are widely used in change detection. However, it is very difficult to identify the changed information between multitemporal SAR images, since SAR images generally suffer greatly from the speckle noise. It subdues the visual quality of the images for interpretation and hinders the automatic information extraction by remote sensing image processing software. Therefore, it is essential to develop robust and reliable methods that can effectively suppress the speckle noise.

To solve the problem, many supervised and unsupervised methods have been developed to reduce the speckle noise for SAR change detection in the past few years. Supervised methods require prior knowledge to collect reliable training samples to obtain a robust classifier [4], [5]. On the other hand, unsupervised change detection methods directly compare the input multitemporal SAR images without any additional information [6], [7]. The unsupervised change detection is more popular, since it is sometimes difficult to obtain prior knowledge. Therefore, in this article, we focus on developing change detection method in an unsupervised manner.

Existing unsupervised SAR image change detection methods are generally composed of the following three steps: 1) image preprocessing, 2) difference image (DI) generation, and 3) DI classification. In the first step, geometric registration is generally involved and plays a fundamental role. In DI generation, the log-ratio operator is commonly used, and the operator is capable to transform multiplicative speckle noise into an additive one. Besides the log-ratio operator, Gauss-ratio [8] and neighborhood-based ratio [9] operators are also proposed to generate DI, which is robust to calibration errors [10]. For DI classification, clustering methods [11], [12] are widely used. They are considered to be robust to noise since the contextual information is taken into account. Gong *et al.* [11] proposed a method that detects changed regions by fuzzy c-means (FCM) clustering with an improved Markov random field (MRF) energy function. The method can achieve excellent performance in speckle noise suppression. In [13], a multiple kernels k-means clustering method with neighborhood information was presented to solve the problem of change detection. Besides

Manuscript received August 14, 2019; revised October 11, 2019; accepted November 6, 2019. Date of publication November 24, 2019; date of current version December 30, 2019. This work was supported in part by the National Key R&D Program of China under Grant 2018AAA0100602, in part by the National Natural Science Foundation of China under Grants 41606198 and 41576011, and in part by the Key R&D Program of Shandong Province under Grant 2019GHY11204. *(Corresponding author: Feng Gao.)*

The authors are with the Qingdao Key Laboratory of Mixed Reality and Virtual Ocean, School of Information Science and Engineering, Ocean University of China, Qingdao 266100, China (e-mail: 914283361@qq.com; gaofeng@ouc.edu.cn; dongjunyu@ouc.edu.cn; neverme@ouc.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2019.2953128

these clustering methods, many advanced methods have been employed for speckle noise suppression, such as level-set algorithm [14], joint dictionary learning [15], Bayesian soft fusion [16], curvelet [17], and canonical correlation analysis [18].

Recently, many researchers consider the unsupervised change detection task as an incremental learning problem. Specifically, the change detection process simulates the brain-like pattern recognition mechanism. Children attempt to understand the world through prior knowledge provided by their parents, and the knowledge helps them form their interpretations of the surroundings. Similarly, SAR change detection can be formulated as the following steps [19]: First, an initial change map is created by clustering methods. After that, training samples are selected, and then, fed into a learning system for training. The selected samples can be considered as prior knowledge. Finally, the system provides its interpretations of the changed information and generates a change map. Gao *et al.* [20] presented a hierarchical FCM clustering framework for sample selection, and the extreme learning machine (ELM) is employed as the classifier. Wang *et al.* [21] proposed an imbalanced learning method for SAR change detection, in which samples along the boundary between changed and unchanged regions are selected for training. These methods use shallow models as classifiers, and the feature representation capabilities are limited.

Deep learning models, which contain many hidden layers, are capable of extracting discriminative features. These models have made dramatic improvements in things like natural language processing [22], [23], saliency detection [24], [25], and objection detection [26], [27]. Recently, deep learning methods have attracted a lot of attention in the remote sensing community. Many deep learning techniques have been proposed to solve the problem of change detection. In [28], an FCM-based classifier was designed for reliable sample selection, and a deep belief network (DBN) was employed for SAR image change detection. Ghosh *et al.* [29] proposed a modified Hopfield type neural network (HTNN) for change detection. Zhang *et al.* [30] utilized a DBN to learn the invariant features from the input images, and then, the clustering algorithm was employed to generate the change map. Planinsic and Gleich [31] presented a change detection algorithm based on stacked autoencoder (SAE). Features were extracted by discrete wavelet transform, and then, these features were fed into SAE to distinguish changed and unchanged pixels. Gong *et al.* [32] utilized a DBN to analyze the changed information using spectral, textual, and spatial features for multispectral image change detection. Hou *et al.* [33] presented a change detection method by combining deep features and saliency computation using a low-rank algorithm. Zhan *et al.* [34] proposed a refined deep siamese convolutional neural network (CNN) model to extract distinct features between changed and unchanged class. In [35], stacked denoising autoencoders are employed to extract features from multitemporal images, and the influence of speckle noise can be alleviated. Su and Cao [36] utilized the fuzzy autoencoder to detect changes from multitemporal images. Later, Liu *et al.* [37] proposed an elegant local restricted convolutional neural networks (LR-CNN) for polarimetric SAR change detection. The spatial constraint of the pixels is formulated as an extra regularization term in the loss function, and it is imposed on the output layer of the CNN. The spatial constraint is demonstrated to be effective in speckle noise suppression.

Recent studies show that the network depth is of considerable significance in classification, and the leading results of the visual recognition tasks [38] [39] all exploit features from very deep networks. However, in the SAR change detection task, discriminant features from very deep architectures have rarely been sufficiently exploited. Therefore, building a very deep network for SAR change detection can capture more discriminant features, and thus, could further enhance the change detection performance.

In this article, we aim to develop an SAR change detection method with very deep architectures. To design such a technique, it involves the following two problems.

1) *Exploding gradients:* With the increase of network depth, error gradients accumulate, and thus, result in substantial gradients. These error gradients result in an unstable network.
2) *Redundant features:* Every convolution operation generates a set of feature maps. Previous studies have shown that deep networks tend to produce a lot of redundant features that are very similar.

These unnecessary features also affect the performance of the network.

To solve the aforementioned problems, we proposed a very deep cascade network (DCNet) to extract invariant and discriminative features for multitemporal SAR image change detection. Different from the cascade method in [40], the feature extractors of different levels are cascaded sequentially in this article. First, the input images are preclassified by the FCM, and pseudolabeled samples are selected. Then, these samples are fed into the DCNet for training. The residual learning is introduced to optimize the convolutional layers, and the degradation problem from increasing depth of the model can be alleviated. Besides, a channel weighting-based residual block is designed to exploit the interchannel relationship of features. Moreover, outputs of different hierarchical layers are fused to form the final feature set. Finally, unlabeled samples from the input multitemporal SAR images can be classified by DCNet. The final change map can be obtained by combining the DCNet classification result and the preclassification result. Experimental results on four real SAR image datasets demonstrate the superiority of the proposed DCNet over several state-of-the-art methods.

In summary, the main contributions of this article are as follows.

- To alleviate the exploding gradients problem, feature fusion mechanism and residual learning are introduced to improve the efficiency of training. Residual learning optimizes the training by optimizing the convolutional layers. Besides, a fusion mechanism is employed to combine the outputs of different hierarchical layers. Such feature fusion can be considered as an extended version of residual learning, and then, the exploding gradient problem is further alleviated.
- A simple yet effective channel weighting-based module is designed to solve the feature redundancy problem. Average

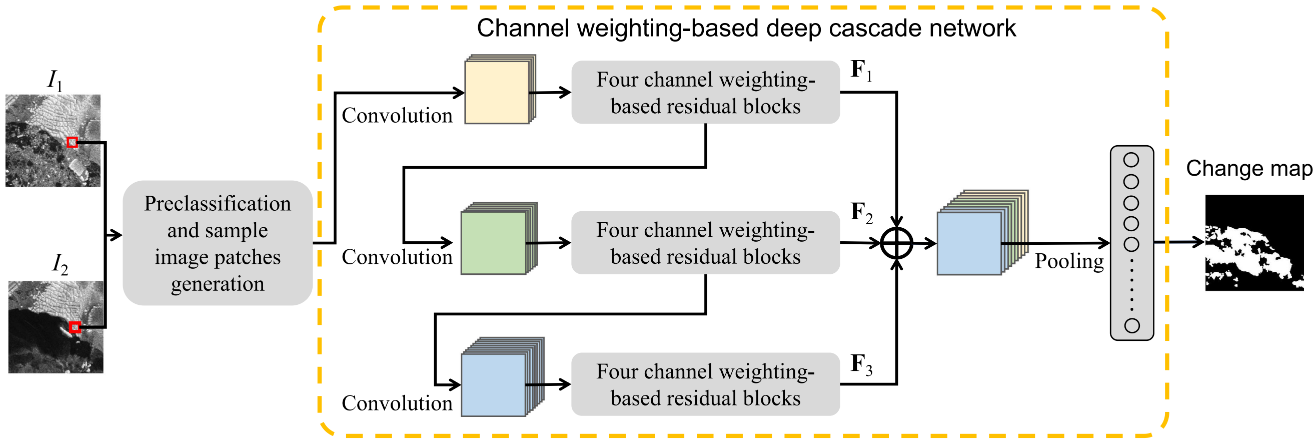


Fig. 1. Schematic illustration of the proposed change detection method based on channel weighting-based DCNet. First, preclassification is implemented by hierarchical FCM, and reliable training samples are selected. Then, sample image patches are generated and fed into the DCNet for training. In the DCNet, channel weighing-based residual learning is adopted to improve the efficiency of training by optimizing the convolutional layers, and the outputs of several hierarchical layers are fused to extract discriminative features. Finally, the change map can be obtained by combining the preclassification result and the DCNet classification result.

pooling and max pooling are used to aggregate channel-wise information. After that, meaningful channel features are emphasized, and unnecessary ones are suppressed. Therefore, the similarity problem in feature maps can be alleviated, and the classification performance of the DCNet is improved.

- We conducted extensive experiments on several real SAR datasets to validate the effectiveness of the proposed DC-Net. In addition, we released our codes and parameters to facilitate future research in SAR image change detection.

The rest of this article is organized as follows. We describe the proposed DCNet in Section II. Section III provides the experimental results on real multitemporal SAR images to verify the effectiveness of the proposed DCNet. Finally, we conclude this article in Section IV.

II. Methodology

Given two coregistered SAR images, I_1 and I_2 captured at different times over the same geographical region, we aim to produce a binary change map that shows the change information between both images. In the binary change map, changed pixels are marked as “1” and unchanged pixels are marked as “0.”

As illustrated in Fig. 1, the proposed method contains the following two parts: 1) preclassification and pseudolabeled samples generation; and 2) classification by the DCNet and final change map generation. In this section, we first describe the process of preclassification, then present the detailed structure of the proposed DCNet. At last, the change map generation process is revealed.

A. Preclassification and Reliable Samples Selection

The log-ratio operator is first adopted to generate a DI. It is widely acknowledged that the influence of speckle noise can be reduced by the log-ratio operator considering the multiplicative nature of speckle in SAR images. Therefore, the speckle noise can be suppressed to some extent. After obtaining the DI, the hierarchical clustering algorithm [20] is utilized for DI classification. In the hierarchical clustering algorithm, the FCM algorithm is used to divide DI into multiple clusters. Then, some clusters will be combined. Finally, three clusters $\{\Omega_c, \Omega_u, \Omega_i\}$ will be generated. Ω_c and Ω_u represent the changed class and unchanged class. Ω_i represents the uncertain class. Pixels belonging to Ω_c and Ω_u are selected as reliable samples for the proposed DCNet. Pixels in Ω_i will be further classified by the DCNet.

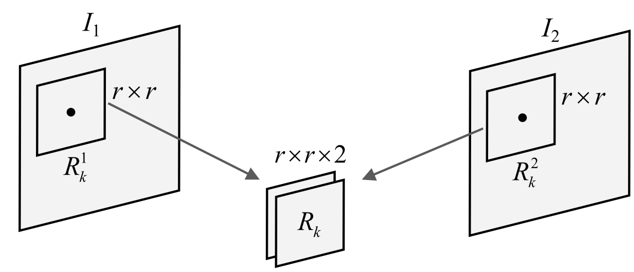


Fig. 2. Process of sample image generation. An image patch R_k^1 of the size $r \times r$ is extracted from I_1 and the corresponding image patch R_k^2 is extracted from I_2. Both image patches are concatenated to a new image R_k with the size of $r \times r \times 2$.

For the task of SAR image change detection, contextual information is essential to achieve robust feature representation [28]. Therefore, image patches centered at selected samples are extracted from the potential changed area of the original SAR images, as shown in Fig. 2. R_k^1 denotes the image patch centered at pixel k in I_1, and R_k^2 denotes the corresponding image patch in I_2. The size of each patch is $r \times r$. Both patches are combined to form a new image R_k with two channels, and the size of R_k is $r \times r \times 2$.

B. Deep Cascade Network (DCNet)

Due to the existence of speckle noise, only 3–5 convolutional layers can hardly extract discriminant features from multitemporal SAR images. However, as the network depth increasing, the

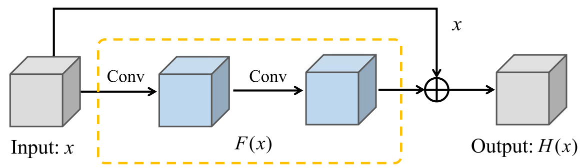


Fig. 3. Illustration of one typical residual block. Fitting the residual mapping $F(x)$ is more easier than fitting $H(x)$. The degradation problem can be effectively alleviated, and deeper network depth can be achieved.

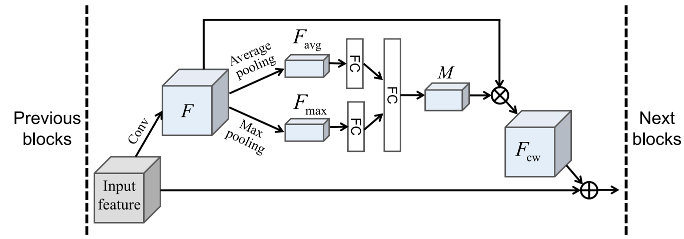


Fig. 4. Illustration of the channel weighting-based residual block.

exploding gradient problem may occur, and the accuracy gets saturated. To alleviate the problem, we introduce the residual learning [41] and batch normalization to optimize the convolutional layers. In addition, a fusing technique is employed to make full use of the features from different layers. Moreover, the channel-weighting module is designed to exploit the interchannel relationship of features. Fig. 1 gives a detailed diagram of the proposed DCNet. The DCNet contains two important components: channel weighting-based residual block and feature fusion module. In the following subsections, we will describe the following two parts in detail.

1) Channel Weighting-Based Residual Block: Residual learning presented good performance in the degradation problems. As reported in [41], the residual networks are much easier to be trained, since most values of the residual image are likely to be zero or tend to be small. One typical residual block is illustrated in Fig. 3. It directly connects the input and the output by a shortcut pathway within the block. Mathematically, let $H(x)$ denotes the desired underlying mapping, the residual mapping $F(x)$ can be denoted by

$$F(x) := H(x) - x. \tag{1}$$

Therefore, the mapping of the residual block equals to the output of a typical CNN. Fitting a residual mapping $F(x)$ is easier than fitting the original $H(x)$. When $H(x)$ is a near identity mapping, the processing of training is very efficient in the residual block.

Recently, the attention mechanism is incorporated in large-scale classification tasks. Humans do not process the whole scene at a glance. Instead, humans can selectively focus on the most salient parts of the scene, and therefore, obtain a better understanding of the visual structure. Hu *et al.* [42] designed a squeeze-and-excitation (SE) module that can exploit the interchannel relationship. In the module, global average-pooled features are used to achieve attention. Inspired by Hu's work, max pooling is also introduced to achieve better channel-wise attention in this article. We design a channel weighting-based residual block for the DCNet as shown in Fig. 4. The input feature map $x \in \mathbb{R}^{w \times w \times c}$ is fed into the convolutional layer to obtain F. And, F is squeezed in the spatial domain by average pooling and max pooling operators. Then, local statistical information from local receptive fields is aggregated. Two different spatial feature descriptors $F_{\text{avg}} \in \mathbb{R}^{1\times 1\times c}$ and $F_{\max} \in \mathbb{R}^{1\times 1\times c}$ are generated.

Both feature descriptors are then fed into two fully connected (FC) layers, respectively. The fully connected (FC) layers are utilized to parameterize the gating mechanism. $W_0 \in \mathbb{R}^{\frac{c}{t}\times c}$ represents the weights in the FC layer handling F_{avg}, and $W_1 \in \mathbb{R}^{\frac{c}{t}\times c}$ represents the weights in the FC layer handling $F_{\max}$. t is the dimensionality-reduction ratio to limit model complexity. In our experiments, we set t to 4.

In the last FC layer, the weight $W_2 \in \mathbb{R}^{c\times \frac{c}{t}}$ is shared between the average-pooling features and the max-pooling features. After the shared FC layer is applied, the output feature vectors are merged using element summation. The channel weighting-based vector M is computed as

$$M = \sigma(W_2\delta(W_0F_{\text{avg}}) + W_2\delta(W_1F_{\max})) \tag{2}$$

where M is a learnable one-dimensional vector $M \in \mathbb{R}^{1\times 1\times c}$, which represents a nonmutually-exclusive relationship ($M = [M_1, M_2, \ldots M_c]$). It should be noted that the $ReLU$ function δ is used in the first FC layer. σ denotes the $sigmoid$ function, and the output from the last FC layer is mapped to the range of 0–1. Thus, opposed to one-hot activation, multiple channels are allowed to be emphasized. Larger M_c means that the information of the cth feature map in F is emphasized. Finally, the channel weighting-based feature F_{cw} can be denoted as

$$F_{\text{cw}} = M \otimes F \tag{3}$$

where F is the feature before pooling. $\otimes$ denotes channel-wise multiplication to broadcast the attention values. Finally, the output of the channel weighting-based residual block is calculated by elements-wise summation of F_{cw} and the input feature x.

The proposed method achieves a deeper network depth by cascaded channel weighting-based residual blocks with different parameters. Cascaded residual blocks can fully extract the feature information of SAR images, and is not prone to degradation and overfitting. In addition, there are three groups of channel weighting-based residual blocks in the proposed DCNet as illustrated in Fig. 1. They are utilized to extract low-, mid-, and high-level features, respectively. The first group of residual blocks can extract minor details of the image, such as small lines or dots. Then, the second group of residual blocks can extract mid-level features that correspond to a combined output of low-level features. Then, the third group of the residual block can capture the structured information and semantic context of the input data. Each group is comprised of four channel weighting-based residual blocks that have the same parameters.

2) Feature Fusion: Recently, feature fusion has attracted wide attention due to its excellent performance in computer vision tasks. Zhao *et al.* [43] proposed an end-to-end network that fused a variety of features for classification. In [44], deep features extracted from two fully connected layers were fused for remote sensing scene classification. Song *et al.* [45] proposed a deep learning model in which the outputs of different layers were fused for the hyperspectral image classification. Inspired by these studies, we introduce the feature fusion strategy to investigate the complementary information in three groups of channel weighting-based residual blocks for SAR image change detection. Moreover, the feature fusion strategy works in similar ways as residual learning. Specifically, residual learning combines features of the same scale, while the feature fusion strategy combines features from different scales. Therefore, the feature fusion strategy can be viewed as an extended version of residual learning, and thus, the exploding gradient problem can be further alleviated.

Different levels of cascade residual blocks may capture information of different scales, including coarse or fine scales. Therefore, fusion of different level features of cascade residual blocks is very important for accurate change detection. The features from three channel weighting-based residual block groups are denoted as $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$, respectively. $\mathbf{F}_1$ contains 16 feature maps, $\mathbf{F}_2$ contains 32 feature maps, and $\mathbf{F}_3$ contain 64 feature maps. The most important issue in feature fusion is dimension matching. In order to achieve this, 64 kernels are used to convolve $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$. The size of each kernel is 1×1. After such convolution, the numbers of three groups of output all become 64. After that, the fusion process can be achieved by pixel-wise summation as follows:

$$\mathbf{F} = g_1(\mathbf{F}_1) + g_2(\mathbf{F}_2)) + g_3(\mathbf{F}_3) \tag{4}$$

where $\mathbf{F}$ represents the fused features, and g_1, g_2, and g_3 are the operation of dimension matching.

C. Final Change Map Generation

After obtaining the fused features $\mathbf{F}$, they are transformed into a vector through several fully connected layers. Then, the feature vector is fed into a $softmax$ layer to calculate the possibility to be changed or unchanged. The output of the fully connected layer is denoted as $(\mathbf{F}_u, \mathbf{F}_c)$, and it is hard to express the probability distribution of output. Therefore, the $softmax$ layer is employed to calculate the possibility

$$p_c = \frac{e^{\mathbf{F}_c}}{e^{\mathbf{F}_c} + e^{\mathbf{F}_u}} \tag{5}$$

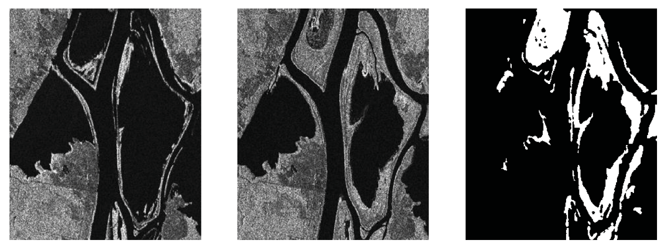
$$p_u = \frac{e^{\mathbf{F}_u}}{e^{\mathbf{F}_c} + e^{\mathbf{F}_u}} \tag{6}$$

where p_c and p_u denote the possibility to be changed and unchanged, respectively. $\mathbf{F}_u$ represents the first output node of the proposed DCNet, and $\mathbf{F}_c$ represents the second output node of the proposed DCNet. If $p_c > p_u$, the pixel is considered to be changed, otherwise, the pixel is considered to be unchanged.

As mentioned before, pixels belonging to Ω_c and Ω_u are treated as training samples for the proposed DCNet. After obtaining a trained model, pixels in Ω_i can be classified by the model. Finally, the DCNet classification result and the preclassification result can be combined together to form a map. In the map, changed pixels are marked as "1" and unchanged pixels are marked as "0."

TABLE I
DETAILS OF THE PROPOSED DCNET

Type	Kernel size	Layers	Output	Output size
Conv	3×3	1-9	16	28×28×16
Conv	3×3	10	32	14×14×32
Conv	3×3	11-18	32	14×14×32
Conv	3×3	19	64	7×7×64
Conv	3×3	20-27	64	7×7×64
Pooling	—	28	—	1×1×64
FC	—	29	2	—



(a) (b) (c)

Fig. 5. Ottawa dataset. (a) Image captured in May 1997. (b) Image captured in August 1997. (c) Ground-truth image.

Table I shows the implementation details of the proposed DCNet (where the input is resized as 28×28, and 27 convolutional layers are implemented). In the pooling layer, we use global average pooling, and FC means the fully connected layer.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we first describe the datasets used in our experiments. After that, the evaluation criteria are introduced in detail. Next, an exhaustive investigation of several important parameters on the change detection performance is presented. Finally, the proposed method is compared with several excellent methods.

A. Dataset and Evaluation Criteria

In order to verify the effectiveness of the proposed DCNet, we employed the proposed method on four multitemporal SAR datasets acquired by different sensors. Coregistration and geometric corrections have been done on these datasets. Since the ground-truth data are crucial for the accuracy assessment, the ground-truth change maps were manually annotated carefully with expert knowledge.

The first dataset is the Ottawa dataset, which is captured by the Radarsat sensor. As shown in Fig. 5, it presents two SAR images captured over the city of Ottawa. The size of the images is 290×350 pixels. The dataset was provided by the National Defense Research and Development Canada. Both images are acquired in May 1997 and August 1997, respectively. The images show changed regions that were afflicted by floods.

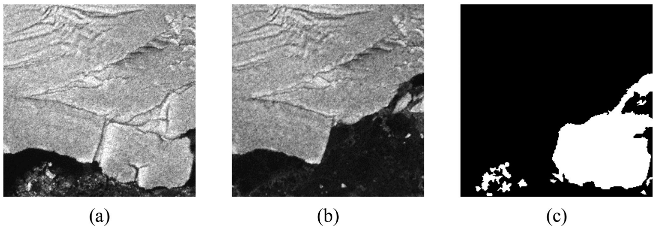


Fig. 6. Sulzberger I dataset. (a) Image captured in March 11 in 2011. (b) Image captured in March 16 in 2011. (c) Ground-truth image.

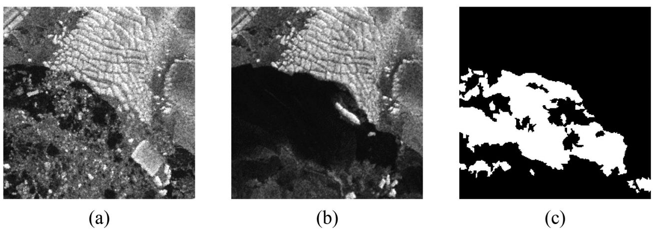


Fig. 7. Sulzberger II dataset. (a) Image captured in March 11 in 2011. (b) Image captured in March 16 in 2011. (c) Ground-truth image.

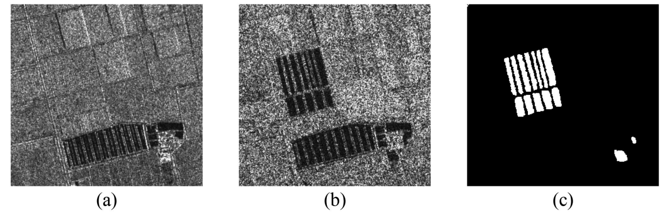


Fig. 8. Farmland dataset. (a) Image captured in June 2008. (b) Image captured in June 2009. (c) Ground-truth image.

The available ground-truth change map was manually annotated by experts with rich knowledge in photointerpretation.

The next two datasets (Sulzberger I and II) are selected from two large SAR images from one ice shelf. Both images were taken by the Envisat satellite of the European Space Agency on March 11 and 16, 2011, respectively. The two images show the process of the sea ice breakup. In March 2011, the Tohoku Tsunami was triggered in the Pacific Ocean, and massive ocean waves caused the Sulzberger Ice Shelf to flex and break. The original size of the two SAR images is 2263×2264 pixels. It is difficult to display detailed information in such huge images. Therefore, we chose two typical regions (256×256 pixels in each area). The available ground-truth images are generated by integrating prior knowledge and photointerpretation. The Sulzberger I dataset is shown in Fig. 6, and the Sulzberger II dataset is shown in Fig. 7.

The last dataset is the Farmland dataset, which is selected from two SAR images captured over the Yellow River Estuary in China. Both SAR images were acquired by Radarsat-2 in June 2008 and June 2009, respectively. The original size of the two images is 7666×7692 pixels. Similar to the Sulzberger datasets, one typical region is chosen to demonstrate the efficacy of the proposed DCNet. The changed areas show newly cultivated farmland. The dataset is shown in Fig. 8. It should be noted that one image is a single-look image, and the other is a four-look image. Therefore, the speckle noise is much stronger. It is very challenging to perform change detection on this dataset.

The performance evaluation of change detection methods is critical. In this article, false positives (FP), false negatives (FN), percentage correct classification (PCC), overall errors (OE), and Kappa coefficient (KC) are used as the evaluation criteria. The FP denotes the number of pixels that are unchanged in the ground-truth image but falsely classified into the changed class in the change detection result. The FN denotes the number of pixels that are changed in the ground-truth image but falsely classified into the unchanged class in the change detection result. Then, the OE can be computed by using $\mathrm{OE} = \mathrm{FP} + \mathrm{FN}$. The PCC can be computed by

$$\mathrm{PCC} = \frac{N_t - OE}{N_t} \times 100\% \tag{7}$$

where N_t represents the total pixels in the ground-truth image. KC is defined as

$$\mathrm{KC} = \frac{\mathrm{PCC} - \mathrm{PRE}}{1 - \mathrm{PRE}} \tag{8}$$

$$\mathrm{PRE} = \frac{(N_c + \mathrm{FP} - \mathrm{FN}) \times N_c + (N_u + \mathrm{FN} - \mathrm{FP}) \times N_u}{N_t \times N_t} \tag{9}$$

where N_c denotes the number of changed pixels in the ground truth image, and N_u denotes the number of unchanged pixels in the ground-truth image. It should be noted that KC is more persuasive than the PCC, since more detailed information should be taken into account to obtain good KC values in change detection.

B. Parameters Analysis of the Proposed DCNet

1) Analysis of the Sample Image Size: In the proposed DCNet, the size of the sample image is an important parameter. The contextual information in classification is sensitive to neighborhood noise. The first experiment tests the parameter r. Here, r denotes the size of a sample image patch from the input SAR image. We evaluate the performance of change detection by taking $r = 5$, 7, 9, 11, 13, and 15, respectively. The PCC value is investigated as the validation criterion on different datasets.

Fig. 9 shows the sensitivity analysis result of the parameter r. We can observe that when $r = 9$, the proposed DCNet achieves the best results on most of the dataset. If we use larger patches, the sample image may not be representative of the pixel in the center. In addition, larger patches will increase the computational burden. Hence, r is set to 9 in our following experiments for sample image generation.

2) Analysis of Training Sample Numbers: The number of training samples affects the change detection performance, since a large number of samples is usually essential for deep neural networks to optimize parameters. In this subsection, we investigate the relationship between the PCC and training sample numbers on four SAR datasets.

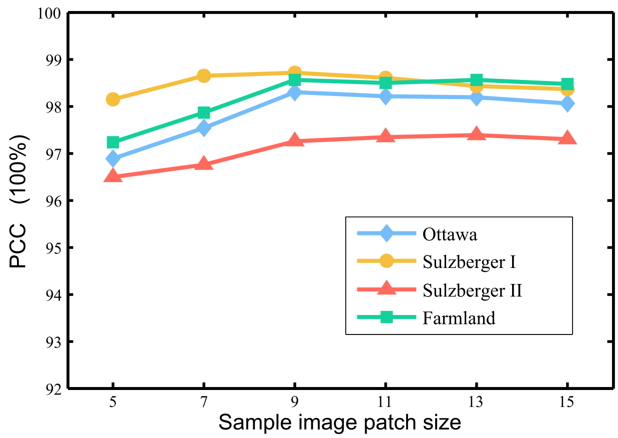


Fig. 9. Relationship between the classification accuracy and sample image patch size.

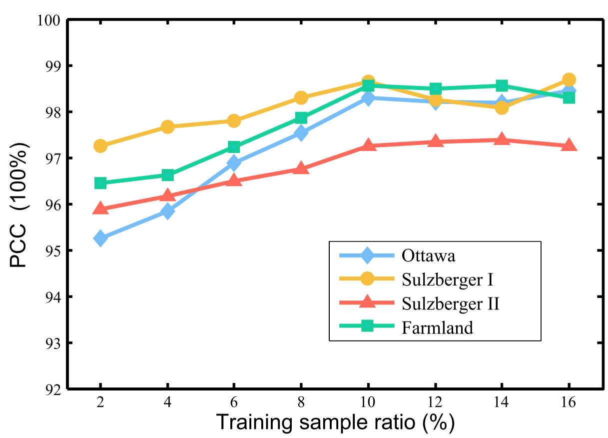


Fig. 10. Relationship between PCC and the number of training samples.

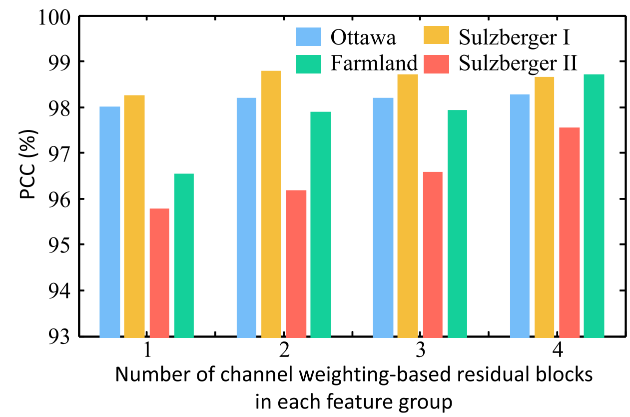


Fig. 11. Relationship between PCC and the residual block numbers N_r in each feature group.

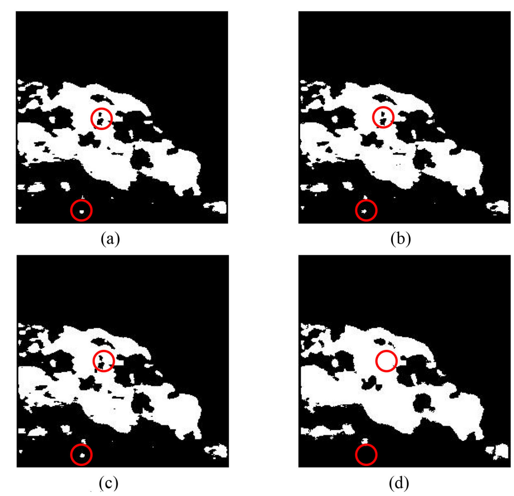


Fig. 12. Change detection results on the Sulzberger II dataset with different N_r values. (a) $N_r = 1$. (b) $N_r = 2$. (c) $N_r = 3$. (d) $N_r = 4$.

As mentioned in Section II, training samples are randomly selected from Ω_c and Ω_u. We randomly selected 2%, 4%, 6%, 8%, 10%, 12%, 14%, and 16% pixels from Ω_c and Ω_u as training samples, respectively. Fig. 10 shows the change detection performance on four datasets when different number of training samples are employed. From these curves, it can be observed that there is a sharp increase in the PCC value when the number of training samples ranges from 2% to 10%. The PCC value tends to be stable when the training sample number is 10% or larger on most datasets. Therefore, in our experiments, we select 10% pixels from Ω_c and Ω_u as training samples. Such a relatively small ratio is selected since the hardware resources are often limited.

3) Analysis of Channel Weighting-Based Residual Block Numbers: In this article, we are devoted to constructing a deep neural network for the interpretation of multitemporal SAR images. From Fig. 1, we can observe that there are three groups of channel weighting-based residual blocks in the proposed DCNet. N_r denotes the number of channel weighting-based residual blocks in each group. A larger value of N_r means a deeper neural network. For instance, $N_r = 1$ means six convolution layers in residual blocks for feature extraction, and so on. In this analysis, N_r is set to 1, 2, 3, 4. The relationship between PCC and the number of residual blocks is presented in Fig. 11.

We can observe that the proposed DCNet achieves the best performance when $N_r = 4$ for the Ottawa, Sulzberger II, and Farmland datasets. There are many small noisy regions in the change maps of the Ottawa, Sulzberger II, and Farmland datasets, the classification model needs strong discriminative power to detect these regions correctly. It is evident that deeper networks can increase the discriminative power of the proposed DCNet. On the Sulzberger I dataset, there are fewer noise regions in the change map, and the PCC values tend to be stable when $N_r \geq 2$. It means that two residual blocks are enough for discriminant feature extraction on this dataset. Therefore, in our implementations, N_r is set to 2 for the Sulzberger I dataset. On the other datasets, $N_r = 4$ is selected.

Visual comparisons of different N_r values on the Sulzberger II and Farmland datasets are illustrated in Figs. 12 and 13, respectively. It can be observed that there are many noisy falsely detected changed regions in the generated change map when $N_r \leq 3$. As the value of N_r increases, this phenomenon can

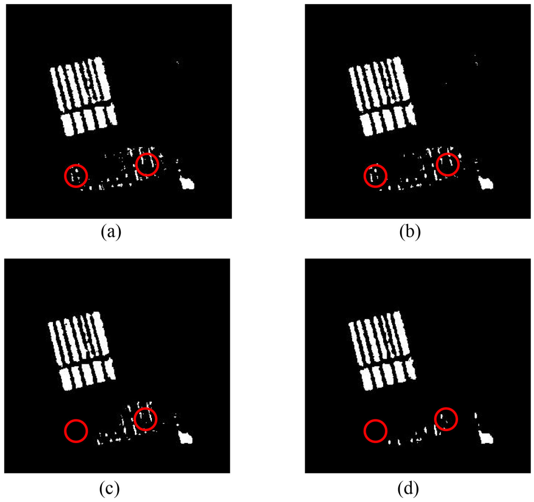


Fig. 13. Change detection results on the Farmland dataset with different N_r values. (a) $N_r = 1$. (b) $N_r = 2$. (c) $N_r = 3$. (d) $N_r = 4$.

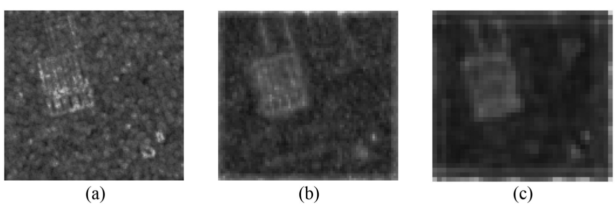


Fig. 14. Feature images of the Farmland dataset. (a) Fused feature images extracted from low level. (b) Fused feature images extracted from midlevel. (c) Fused feature images extracted from high level.

be alleviated. Careful observation can identify the improvement when N_r increases, such as regions marked by the red circles in Figs. 12 and 13. Deeper architecture can effectively extract discriminative features, and therefore, improve the change detection performance.

Furthermore, we fuse the feature images of the Farmland dataset extracted from different levels, as shown in Fig. 14. We can observe that information on coarse scale (stripes and edges) is extracted from low-level and midlevel. However, the noise is still obvious. In Fig. 14(c), it is clear that meaningful features can be extracted so that the impact of speckle noise is alleviated. These features effectively improve the change detection performance, as shown in Fig. 11. Therefore, we can conclude that the proposed DCNet has good feature learning ability in terms of feature visualization and PCC values.

C. Analysis of Different Feature Fusion Strategy

In this subsection, different levels of features are fused to evaluate the feature fusion strategy. Table II illustrates the PCC values on four datasets by employing different fusion strategies. DCNet-H denotes that feature fusion strategy is not employed in classification, and only high-level features are used. DCNet-LH represents that the low-level and high-level features are fused in classification. Correspondingly, DCNet-MH means that the midlevel and high-level features are fused in classification.

TABLE II
PCC VALUES BY DIFFERENT FUSION STRATEGIES ON FOUR DATASETS

	DCNet-H	DCNet-LH	DCNet-MH	DCNet
Ottawa	98.10	98.11	98.13	98.30
Sulzberger I	98.72	98.72	98.68	98.80
Sulzberger II	95.50	95.68	95.22	97.25
Farmland	97.85	97.88	97.96	98.71

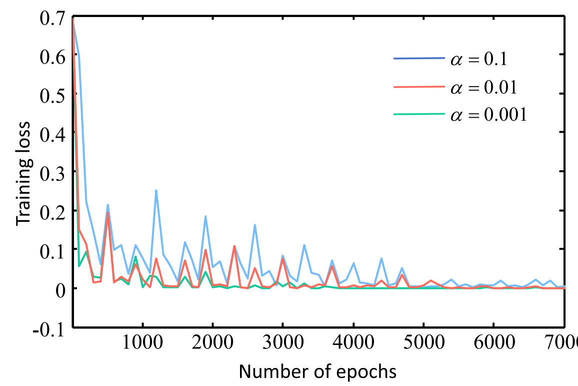


Fig. 15. Relationship between the training loss and the number of epochs with different learning rate α on the Ottawa Dataset.

It should be noted that the proposed DCNet combines the low-level, midlevel, and high-level features.

From Table II, we can observe that feature fusion can improve change detection performance. DCN-H exhibits the lowest PCC values because information embedded in different levels is not fully exploited. The classification performance can be improved by fusing the features from lower or middle levels. It indicates that the complementary information in different levels of features can be exploited by the simple yet effective fusion strategy employed in this article. Particularly, the proposed DCNet achieves the best performance, and we can draw the conclusion that the fusion strategy used in the proposed DCNet is indeed effective.

D. Analysis of the Learning Rate

In this subsection, we analyze the effect of learning rate α on network convergence and the effectiveness of the residual learning based on the Ottawa dataset. As shown in Fig. 15, we can observe that α plays a key role in network convergence. As mentioned, batch normalization is included in the proposed DCNet. It allows a relatively high learning rate while accelerating the convergence of the network. As illustrated in Fig. 15, there are less fluctuations in the loss when $\alpha = 0.001$. Therefore, the proposed DCNet is optimized by $\alpha = 0.001$, which provides a relatively steady convergence.

In Fig. 16, comparison of the training loss between DCNet with residual learning and DCNet without residual learning is presented. It can be observed that the DCNet with residual learning can converge steadily and efficiently. On the contrary, the DCNet without residual learning exhibits the degradation

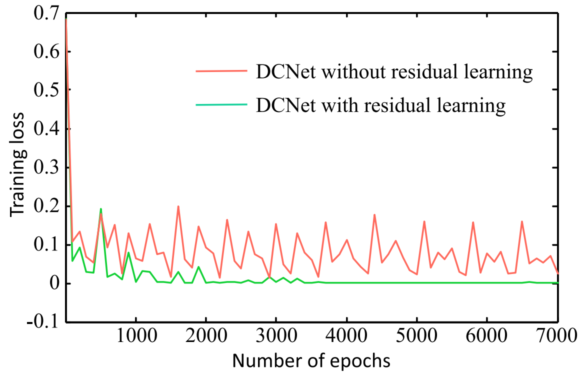


Fig. 16. Relationship between the training loss and number of epochs with/without residual learning on the Ottawa dataset.

problem and has higher error throughout the training process. It is demonstrated that the residual learning in the proposed DCNet can effectively alleviate the degradation problem, and therefore, improve the change detection performance.

E. Results on the Ottawa Dataset

In order to verify the effectiveness of the proposed DCNet, we compare our method with several closely related methods, including PCAKM [46], NR-ELM [20], RMG-FDA [10] GaborPCANet [47], LR-CNN [37], and DBN [28].

In PCAKM [46], the contextual information is taken into account by principal component analysis (PCA), and the extracted features are clustered by the k-means algorithm. NR-ELM [20] utilizes an ELM [48] as the classifier. GaborPCANet [47] is a simplified deep learning model, which is comprised of two PCA convolutional layer, binary hashing layer, and block-wise histogram generation layer. RMG-FDA [10] employs a classifier based on random multigraphs. LR-CNN [37] is a CNN with local spatial restrictions on the output layer. In DBN [28], a deep belief network is employed to complete the SAR change detection task.

Among the comparisons, LR-CNN [37] and DBN [28] attempt to accomplish the change detection task by deep learning. It should be noted that the LR-CNN [37] takes the polarimetric information into account. We modified the LR-CNN model to make it suitable for datasets used in this article. The result of PCAKM, NR-ELM, RMG-FDA, and GaborPCANet are implemented by using the authors' publicly available codes and default parameters.

Both visual and quantitative analyses are made in our experiments. For visual examination, the change maps generated by different methods are exhibited in figure form. For quantitative analysis, the change maps are exhibited in tabular form.

Fig. 17 shows the change detection results on the Ottawa dataset, and Table III lists the evaluation criteria. From Table III, we can observe that for PCAKM, RMG-FDA, and LR-CNN, many changed pixels are missed, and therefore, these methods suffer from high FN values. In addition, the FP values of PCAKM, GaborPCANet, NR-ELM, and DBN are relatively high. Compared with other methods, the proposed DCNet can provide more similar results to the ground-truth change map.

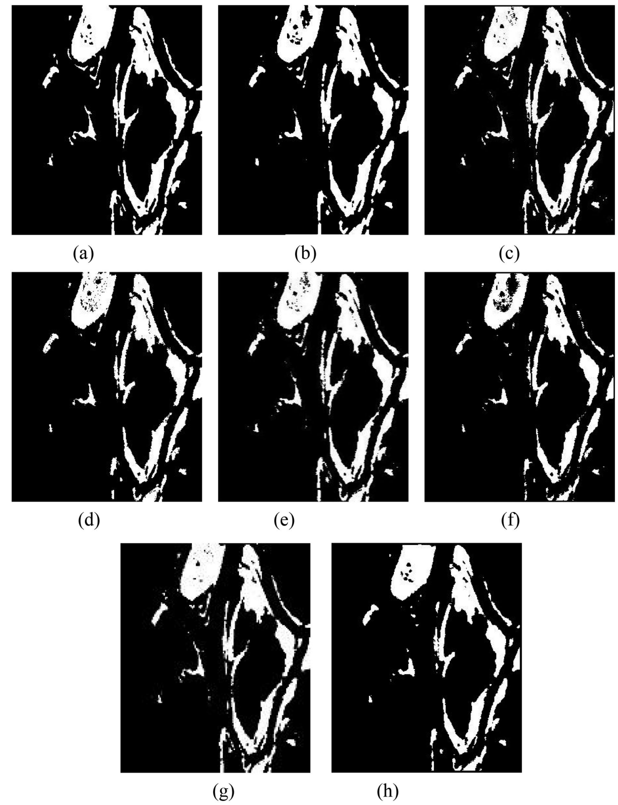


Fig. 17. Visualized results of different change detection methods on the Ottawa dataset. (a) Ground-truth image. (b) Result by PCAKM. (c) Result by NR-ELM. (d) Result by RMG-FDA. (e) Result by GaborPCANet. (f) Result by LR-CNN. (g) Result by DBN. (h) Result by the proposed DCNet.

TABLE III
CHANGE DETECTION RESULTS ON THE OTTAWA DATASET

Methods	FP	FN	OE	PCC(%)	KC(%)
PCAKM	955	1515	2470	97.57	90.73
NR-ELM	695	1076	1771	98.26	93.38
RMG-FDA	198	1883	2071	97.95	91.96
GaborPCANet	953	942	1895	98.13	92.99
LR-CNN	63	3747	3810	96.25	84.45
DBN	995	704	1699	98.33	93.76
Proposed DCNet	679	1051	1730	98.30	93.54

On this dataset, the DBN yields the best PCC value of 98.33%. The proposed DCNet achieves a PCC value of 98.30%, which is quite competitive to DBN on this dataset. It is evident that the proposed DCNet can exploit the nonlinear relations from the multitemporal data by channel weighting-based residual learning and feature fusion. The comparisons also demonstrate the effectiveness of the proposed method on the Ottawa dataset.

F. Results on the Sulzberger I Dataset

Fig. 18 illustrates the change detection results on the Sulzberger I dataset. The evaluation metrics are listed in

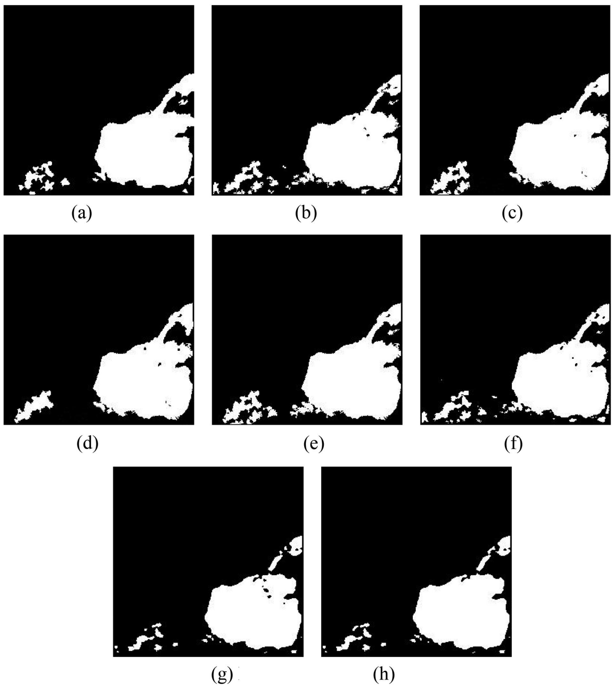


Fig. 18. Visualized results of different change detection methods on the Sulzberger I dataset. (a) Ground-truth image. (b) Result by PCAKM. (c) Result by NR-ELM. (d) Result by RMG-FDA. (e) Result by GaborPCANet. (f) Result by LR-CNN. (g) Result by DBN. (h) Result by the proposed DCNet.

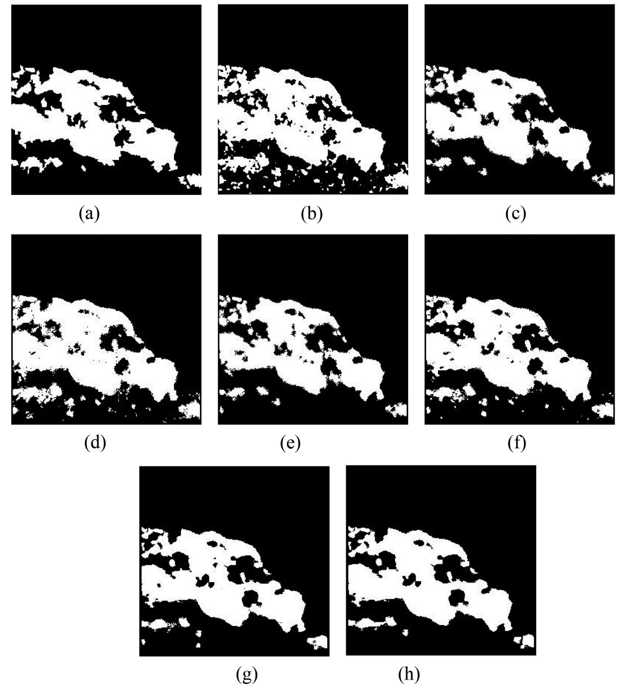


Fig. 19. Visualized results of different change detection methods on the Sulzberger II dataset. (a) Ground-truth image. (b) Result by PCAKM. (c) Result by NR-ELM. (d) Result by RMG-FDA. (e) Result by GaborPCANet. (f) Result by LR-CNN. (g) Result by DBN. (h) Result by the proposed DCNet.

TABLE IV
CHANGE DETECTION RESULTS ON THE SULZBERGER I DATASET

Methods	FP	FN	OE	PCC(%)	KC(%)
PCAKM	711	479	1190	98.18	93.90
NR-ELM	719	832	1551	97.63	91.95
RMG-FDA	626	1094	1720	97.38	90.97
GaborPCANet	895	703	1598	97.56	91.79
LR-CNN	957	119	1076	98.36	94.59
DBN	149	764	913	98.61	95.18
Proposed DCNet	103	681	784	98.80	95.87

TABLE V
CHANGE DETECTION RESULTS ON THE SULZBERGER II DATASET

Methods	FP	FN	OE	PCC(%)	KC(%)
PCAKM	2368	203	2571	96.08	90.15
NR-ELM	1490	1415	2905	95.57	88.43
RMG-FDA	456	1794	2250	96.57	90.78
GaborPCANet	1410	1437	2847	95.66	88.63
LR-CNN	1198	680	1878	97.13	92.58
DBN	632	1242	1874	97.14	92.43
Proposed DCNet	506	1183	1689	97.42	93.17

Table IV. From the visual comparison, it can be observed that PCAKM, LR-CNN, and GaborPCANet generate many noisy regions, and therefore, they suffer from high FP values. For NR-ELM and RMG-FDA, many changed pixels are missed, and therefore, the FN values of NR-ELM and RMG-FDA are relatively high. Moreover, it can be seen that deep learning-based methods (LR-CNN, DBN, and the proposed DCNet) can achieve better performance than shallow model methods. Compared with LR-CNN, the KC value of the proposed DCNet has increased by 1.28%. Compared with the DBN, the KC value of the proposed DCNet has increased by 0.69%. This further demonstrates that feature fusion and channel weight-based residual learning improves the change detection performance on the Sulzberger I dataset.

G. Results on the Sulzberger II Dataset

Fig. 19 presents the change detection results on the Sulzberger II dataset. The quantitative metrics of different methods are lists in Table V. The result of PCAKM is polluted with noise regions. Therefore, PCAKM suffers from a very high FP value. For NR-ELM, RMG-FDA, and GaborPCANet, many changed regions are missed, and therefore, the FN values of these methods are relatively high. It can be seen that deep learning-based methods (LR-CNN, DBN, and the proposed DCNet) perform better than classical shallow models. It should be noted that the proposed DCNet generates the best change map, which is very similar to the ground truth. Moreover, the proposed DCNet detects changed regions correctly with a clean background. It is demonstrated that the proposed DCNet can effectively suppress

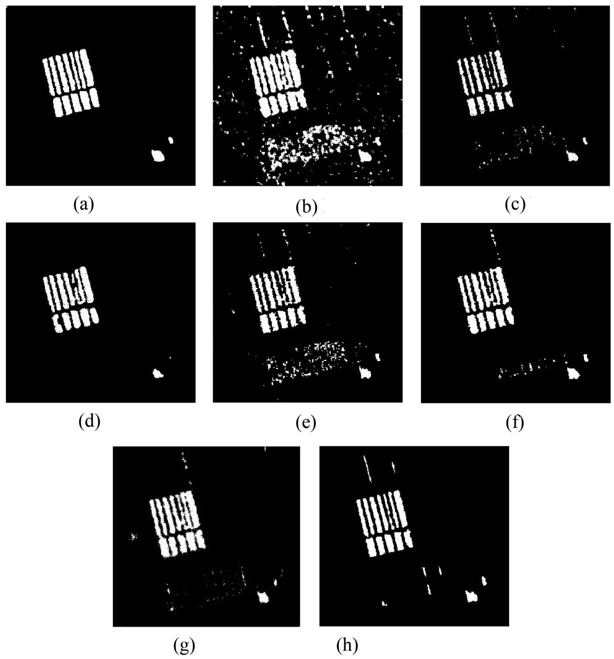


Fig. 20. Visualized results of different change detection methods on the Farmland dataset. (a) Ground-truth image. (b) Result by PCAKM. (c) Result by NR-ELM. (d) Result by RMG-FDA. (e) Result by GaborPCANet. (f) Result by LR-CNN. (g) Result by DBN. (h) Result by the proposed DCNet.

TABLE VI
CHANGE DETECTION RESULTS ON THE FARMLAND DATASET

Methods	FP	FN	OE	PCC(%)	KC(%)
PCAKM	5158	155	5273	94.08	63.29
NR-ELM	256	1794	2050	97.70	76.05
RMG-FDA	169	1614	1783	98.00	79.37
GaborPCANet	2942	493	3435	96.14	71.55
LR-CNN	1118	423	1541	98.27	85.36
DBN	561	668	1229	98.62	87.49
Proposed DCNet	493	658	1151	98.71	88.33

the speckle noise in SAR images. The comparison shows that the proposed DCNet is powerful in discriminative feature extraction and is indeed effective on the Sulzberger II dataset.

H. Results on the Farmland Dataset

Fig. 20 illustrates the change detection results by different methods on the Farmland dataset. The evaluation metrics of different methods are listed in Table VI. The Farmland dataset is seriously interfered by different characteristics of speckle noise. Specifically, the image captured in 2008 is a single-look image, while the image captured in 2009 is a four-look image. The speckle noise in the image captured in 2008 is much greater than the image captured in 2009. Therefore, it is rather challenging to identify changed regions accurately. PCAKM and GaborPCANet do not perform well, and the FP values of both methods are rather high. We can observe that there are many noisy regions in their change maps. For RMG-FDA, many changed regions are missed, and the FN value of RMG-FDA is relatively high. The LR-CNN, DBN, and the proposed DCNet have better performance by employing deep models in feature extraction. The proposed DCNet exhibits the best performance, since channel-weighting-based residual learning and multilevel feature fusion are applied.

Based on the aforementioned experiments on the four real SAR datasets, the proposed DCNet has superior performance over shallow classification models. In addition, by employing channel weighting-based residual learning, the proposed DCNet has better performance than other deep learning-based methods in most cases. Moreover, the feature fusion strategy in the proposed DCNet exploits the complementary information among different feature layers, which further improves the change detection performance. Therefore, it is concluded that the proposed DCNet is a powerful and useful tool for the multitemporal SAR image change detection.

IV. CONCLUSION

Deep learning-based models have been recently discussed to solve the problem of SAR image change detection. Deep feature extraction has presented promising performance. However, with the increase of network depth, the exploding gradient problem often occurs. Besides, deep networks tend to produce a lot of redundant features that are very similar. These redundant features also affect change detection performance.

In this article, a deep learning-based method, DCNet, is proposed to solve the aforementioned problems. In the DCNet, residual learning introduced to solve the exploding gradient problem. Besides, a fusion mechanism is employed to combine the outputs of different hierarchical layers, and then, the exploding gradients problem can be further alleviated. Moreover, a simple yet effective channel weighting-based model is designed to solve the feature redundancy problem. Average pooling and max pooling are used to aggregate channel-wise information. After that, meaningful channel features are emphasized and unnecessary ones are suppressed. Therefore, the similarity problem in feature maps can be alleviated, and the change detection performance is improved. Compared with other closely related works, the proposed DCNet has presented a superior performance in terms of visual comparison and quantitative metrics.

ACKNOWLEDGMENT

The authors would like to thank Prof. T. Celik for sharing the source code of PCAKM. The authors would also like to thank the Associate Editor and anonymous reviewers for the very insightful comments and suggestions, which have significantly improved the quality of this article.

REFERENCES

[1] 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.

[2] D. Burnner, G. Lemonie, and L. Bruzzone, “Earthquake damage assessment of buildings using VHR optical and SAR imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.

[3] S. Quan, B. Xiong, D. Xiang, L. Zhao, S. Zhang, and G. Kuang, “Eigenvalue-based urban area extraction using polarimetric SAR data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 2, pp. 458–471, Feb. 2018.

[4] M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone, “Supervised change detection in VHR images using contextual information and support vector machines,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 20, pp. 77–85, Feb. 2013.

[5] S. Wang, S. Yang, and L. Jiao, “Saliency-guided change detection for SAR imagery using a semi-supervised Laplacian SVM,” *Remote Sens. Lett.*, vol. 7, no. 11, pp. 1043–1052, Nov. 2016.

[6] L. Bruzzone and D. F. Prieto, “Automatic analysis of the difference image for unsupervised change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1170–1182, Mar. 2000.

[7] Z. Yetgin, “Unsupervised change detection of satellite images using local gradual descent,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1919–1929, May 2012.

[8] B. Hou *et al.*, “Unsupervised change detection in SAR image based on Gauss-log ratio image fusion and compressed projection,” *IEEE J. Sel. Topic Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3297–3317, Aug. 2014.

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

[10] F. Gao, X. Wang, J. Dong, S. Wang, “Synthetic aperture radar image change detection based on frequency-domain analysis and random multigraphs,” *J. Appl. Remote Sens.*, vol. 12, no. 1, 2018, Art. no. 016010.

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

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

[13] L. Jia, M. Li, P. Zhang, Y. Wu, and H. Zhu, “SAR image change detection based on multiple kernel k-means clustering with local-neighborhood information,” *IEEE Geosci Remote Sens. Lett.*, vol. 13, no. 6, pp. 856–860, Jun. 2016.

[14] Y. Bazi, F. Melgani, and H. D. Al-Sharari, “Unsupervised change detection in multispectral remotely sensed imagery with level set methods,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3178–3187, Aug. 2010.

[15] X. Lu, Y. Yuan, and X. Zheng, “Joint dictionary learning for multispectral change detection,” *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 884–897, Apr. 2017.

[16] C. Wu, B. Du, X. Cui, and L. Zhang, “A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion,” *Remote Sens. Environ.*, vol. 199, pp. 241–255, Sep. 2017.

[17] F. Li, F. Fang, and G. Zhang, “Unsupervised change detection in SAR images using curvelet and L1-norm based soft segmentation,” *Int. J. Remote, Sens.*, vol. 37, no. 4, pp. 3232–3254, 2016.

[18] T. G. Michael, B. Marchand, J. D. Tucker, T. M. Marston, D. D. Sternlicht, and M. R. Azimi-Sadjadi, “Image-based automated change detection for synthetic aperture sonar by multistage coregistration and canonical correlation analysis,” *IEEE J. Ocean Eng.*, vol. 41, no. 3, pp. 592–612, Mar. 2016.

[19] F. Gao, X. Liu, J. Dong, G. Zhong, and M. Jian, “Change detection in SAR images based on deep semi-NMF and SVD networks,” *Remote Sens.*, vol. 9, no. 5, May 2017, Art. no. 435.

[20] F. Gao, J. Dong, B. Li, Q. Xu, and C. Xie, “Change detection from synthetic aperture radar images based on neighborhood-based ratio and extreme learning machine,” *J. Appl. Remote Sens.*, vol. 10, no. 4, Dec. 2016, Art. no. 046019.

[21] R. Wang, J. Zhang, J. Chen, L. Jiao, and M. Wang, “Imbalanced learning-based automatic SAR images change detection by morphologically supervised PCA-Net,” *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 554–558, Apr. 2019.

[22] E. Cambria and B. White, “Jumping NLP curves: A review of natural language processing research,” *IEEE Comput. Intell. Mag.*, vol. 9, no. 2, pp. 48–57, Feb. 2014.

[23] G. Hinton *et al.*, “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Jun. 2012.

[24] N. Liu and J. Han, “A deep spatial contextual long-term recurrent convolutional network for saliency detection,” *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3264–3274, Mar. 2018.

[25] V. John, K. Yoneda, Z. Liu, and S. Mita, “Saliency map generation by the convolutional neural network for real-time traffic light detection using template matching,” *IEEE Trans. Comput. Imag.*, vol. 1, no. 3, pp. 159–173, Sep. 2015.

[26] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[27] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

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

[29] A. Ghosh, B. N. Subudhi, and L. Bruzzone, “Integration of Gibbs Markov random field and hopfield-type neural networks for unsupervised change detection in remotely sensed multitemporal images,” *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3087–3096, Aug. 2013.

[30] H. Zhang, M. Gong, P. Zhang, L. Su, and J. Shi, “Feature-level change detection using deep representation and feature change analysis for multispectral imagery,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1666–1670, Nov. 2017.

[31] P. Planinsic and D. Gleich, “Temporal change detection in SAR images using log cumulants and stacked autoencoder,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 297–301, Feb. 2018.

[32] M. Gong, T. Zhan, P. Zhang, and Q. Miao, “Superpixel-based difference representation learning for change detection in multispectral remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2658–2673, May 2017.

[33] B. Hou, Y. Wang, and Q. Liu, “Change detection based on deep features and low rank,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2418–2422, Dec. 2017.

[34] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, “Change detection based on deep siamese convolutional network for optical aerial images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.

[35] T. Zhan, M. Gong, J. Liu, and P. Zhang, “Iterative feature mapping network for detecting multiple changes in multi-source remote sensing images,” *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 38–51, Dec. 2018.

[36] L. Su and X. Cao, “Fuzzy autoencoder for multiple change detection in remote sensing images,” *J. Appl. Remote Sens.*, vol. 12, no. 3, 2018, Art. no. 035014.

[37] F. Liu, L. Jiao, X. Tang, S. Yang, W. Ma, and B. Hou, “Local restricted convolutional neural network for change detection in polarimetric SAR images,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 1–16, Mar. 2019.

[38] H. Lin, Z. Shi, and Z. Zou, “Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1665–1669, Oct. 2016.

[39] Q. Liu, R. Hang, H. Song, and Z. Li, “Learning multiscale deep features for high-resolution satellite image scene classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 117–126, Jan. 2018.

[40] Y. Sun, X. Wang, and X. Tang, “Deep convolutional network cascade for facial point detection,” in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3476–3483.

[41] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[42] J. Hu and L. Shen, “Squeeze-and-excitation networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[43] W. Zhao, Y. Wang, X. Chen, Y. Tang, Q. Liu, “Learning deep feature fusion for group images classification,” in *Proc. Chin. Conf. Comput. Vis.*, Oct. 2017, pp. 566–576.

[44] S.Chaib, H. Liu, Y. Gu, and H. Yao, “Deep feature fusion for VHR remote sensing scene classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4775–4784, Aug. 2017.

[45] W. Song, S. Li, L. Fang, and T. Lu, “Hyperspectral image classification with deep feature fusion network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3173–3184, Jun. 2018.

[46] T. Celik, “Unsupervised change detection in satellite images using principal component analysis and k-means clustering,” *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Apr. 2009.

[47] F. Gao, J. Dong, B. Li, and Q. Xu, "Automatic change detection in synthetic aperture radar images based on PCANet," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1792–1796, Dec. 2016.

[48] G. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Feb. 2012.

Yunhao Gao received the B. Sc. degree in electronic information engineering from the Ocean University of China, Qingdao, China, in 2017, where he is currently working toward the M.Sc. degree in computer science and applied remote sensing with the School of Information Science and Technology.

His current research interests include computer vision and remote sensing image processing.

Feng Gao received the B.Sc. degree in software engineering from Chongqing University, Chongqing, China, in 2008, and the Ph. D. degree in computer science and technology from Beihang University, Beijing, China, in 2015.

He is currently an Associate Professor with the School of Information Science and Engineering, Ocean University of China, Qingdao, China. His research interests include remote sensing image analysis, pattern recognition, and machine learning.

Junyu Dong received the B.Sc. and M.Sc. degrees from the Ocean University of China, Qingdao, China, in 1993 and 1999, respectively, and the Ph.D. degree in image processing from Heriot-Watt University, Edinburgh, U. K., in 2003.

He is currently a Professor and Vice Dean with the School of Information Science and Engineering, Ocean University of China. His research interests include visual information analysis and understanding, machine learning, and underwater image processing.

Shengke Wang received the B.Sc. degree in computer science and technology from Jinan University, Jinan, China, in 2000, and the Ph. D. degree in computer science and technology from the South China University of Technology, Guangzhou, China, in 2005.

He is currently an Associate Professor with the School of Information Science and Engineering, Ocean University of China, Qingdao, China. His research interests include computer vision and pattern recognition.