Journal of
Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

# Combined multiscale segmentation convolutional neural network for rapid damage mapping from postearthquake very high-resolution images

Hui Huang
Genyun Sun
Xuming Zhang
Yanling Hao
Aizhu Zhang
Jinchang Ren
Hongzhang Ma

SPIE.

Hui Huang, Genyun Sun, Xuming Zhang, Yanling Hao, Aizhu Zhang, Jinchang Ren, Hongzhang Ma, "Combined multiscale segmentation convolutional neural network for rapid damage mapping from postearthquake very high-resolution images," *J. Appl. Remote Sens.* **13**(2), 022007 (2019), doi: 10.1117/1.JRS.13.022007.

# Combined multiscale segmentation convolutional neural network for rapid damage mapping from postearthquake very high-resolution images

**Hui Huang,[a,b] Genyun Sun,[a,b,*] Xuming Zhang,[a,b] Yanling Hao,[a,b] Aizhu Zhang,[a,b] Jinchang Ren,[c] and Hongzhang Ma[d]**

[a]China University of Petroleum (East China), School of Geosciences, Qingdao, China
[b]Qingdao National Laboratory for Marine Science and Technology, Laboratory for Marine Resources, Qingdao, China
[c]University of Strathclyde, Department of Electronic and Electrical Engineering, United Kingdom
[d]China University of Petroleum (East China), College of Science, Qingdao, China

**Abstract.** Classifying land use from postearthquake very high-resolution (VHR) images is challenging due to the complexity of objects in Earth surface after an earthquake. Convolutional neural network (CNN) exhibits satisfied performance in differentiating complex postearthquake objects, thanks to its automatic extraction of high-level features and accurate identification of target geo-objects. Nevertheless, in view of the scale variance of natural objects, the fact that CNN suffers from the fixed receptive field, the reduced feature resolution, and the insufficient training sample has severely contributed to its limitation in the rapid damage mapping. Multiscale segmentation technique is considered as a promising solution as it can generate the homogenous regions and provide the boundary information. Therefore, we propose a combined multiscale segmentation convolutional neural network (CMSCNN) method for postearthquake VHR image classification. First, multiscale training samples are selected based on segments derived from the multiscale segmentation. Then, CNN is directly trained to classify the original image to further produce the preliminary classification maps. To enhance the localization accuracy, the output of CNN is further refined using multiscale segmentations from fine to coarse iteratively to obtain the multiscale classification maps. As a result, the combination strategy is able to capture objects and image context simultaneously. Experimental results show that the proposed CMSCNN method can reflect the multiscale information of complex scenes and obtain satisfied classification results for mapping postearthquake damage using VHR remote sensing images.  [DOI: 10.1117/1.JRS.13.022007]

**Keywords:** deep learning; convolutional neural network; multiscale segmentation; postearthquake image classification.

Paper 180828SS received Oct. 17, 2018; accepted for publication Dec. 6, 2018; published online Jan. 2, 2019.

## 1 Introduction

Rapid earthquake damage mapping provides rapid, accurate, and comprehensive knowledge about the conditions of damaged area, which is vital in the disaster assessment and mitigation.[1–3] For decades, remote sensing techniques play an essential role in investigating damage information caused by earthquakes due to its prompt availability after disaster and wide coverage.[4,5] Especially, the increasing availability of very high-resolution (VHR) imagery has significantly contributed its wide use in the damage mapping fields.

The VHR remote sensing imagery offers the opportunity to detect the fine details for damage mapping. However, the rich spatial information presented in postearthquake VHR remote sensing images requires more sophisticated processing techniques, leading to considerable problems

*Address all correspondence to Genyun Sun, E-mail: genyunsun@163.com

1931-3195/2019/$25.00 © 2019 SPIE

such as the increasing computation complexity.[6] That is mainly because the postearthquake VHR remote sensing images generally exhibit a complex combination of various damage types. These various damage types show scale variance in a variety of damage structures and spatial layouts, which makes it challenging to obtain the classification maps of postearthquake VHR remote sensing images.[7]

Over the past few decades, considerable efforts have been made to improve the classification accuracy of postearthquake VHR remote sensing images. Intensive studies have focused on this topic based on handcrafted features elaborated from both spectral and spatial domains. Generally, the spectral features are mainly based on the brightness of each band, which are usually regarded as the primary features for recognition of targets objects.[8] The application of the spectral features can be widely found in visual interpretation of damage objects, such as collapsed buildings and debris flow.[8] Semantic spectral features with physical meanings have also been used as they can strengthen the reflectance discrepancy of different objects on specific wavelengths, such as the normalized difference vegetation index[9–11] and the modified normalized difference water index.[12] Image texture is further explored to demonstrate that the combination of texture and spectral features can give supplementary information for efficient damage mapping, such as the gray-level co-occurrence matrix.[7,13,14] Comparing with spectral properties, VHR imagery has a much richer spatial, which improves discriminative ability. For instance, Vu[15] utilized morphological profiles to efficiently capture spatial information for rapid damage mapping in urban areas. Similarly, spatial filters (such as Canny filters and Gabor filters) were also proposed for the extraction of spatial features in the context of postearthquake VHR images.[2,16,17] However, these techniques rely on a prior fixed, albeit usually rich choice of a suitable data representation, which depends on the knowledge of the analyst and on the specificities of the image at hand. As a result, a few image classification methods in these studies can be practically used under time pressure.

Recently, deep learning,[18] one of the state-of-the-art techniques in the field of machine learning and visual recognition, is identified as the best way to extract discriminative and representative high-level features.[19] Deep learning can learn nonlinear spatial filters automatically and generalize a hierarchy of increasingly complex features.[20–22] A superiority of deep learning is that it learns features from the original data directly, showing great flexibility and capability than traditional classification methods.[18,20,22] Especially, convolutional neural network (CNN), constituted of stacked nonlinear adaptive layers, has been proved to be more efficient models in image processing. The entire system of CNN is trained end to end, from raw pixels to ultimate categories, thereby alleviating the requirement to manually design a suitable feature extractor.[23,24] This enables CNN to be widely utilized in hyperspectral image classification.[25–27] However, in spite of the remarkable achievement in the application of hyperspectral images, there still exist recurring limitations when applied in the feature extraction of high-resolution images.

One problem lies in the fixed receptive fields[20] of deep network, which make it unable to characterize the objects with varied size.[28] The fixed receptive field of deep CNN requires a fixed-size input. However, objects in remote sensing images often appear at various observation scales. The fixed-size input is unable to characterize the varied objects in images, which will lead to incomplete feature representation of the image contents. Consequently, the outputs of deep CNN suffer from speckles and noises. To tackle this problem, some recent studies implement the parallel CNN as multiscale CNN to extract different features and hence to produce complementary outputs. The parallel CNN combined the outputs of different individual CNN architectures with different-size inputs to achieve multiscale, convolutional representation of the natural objects.[20,29,30] However, this scheme still has many shortcomings. The parallel CNN method is equivalent to training multiple CNN structures in one algorithm, so both the number of samples required and the training time are significantly increased.

Another dilemma is that the improvement in representation capability of CNN has come partly at the price of reduced feature resolution.[31] This is mainly caused by the repeatedly max-pooling and downsampling in CNN.[32] The reduced resolution makes CNN insensitive to the object boundary, and thereby causes poor objects localization, even suffering a so called "salt-and-pepper" effect.[31,32] Previously, two ways are usually carried out to encounter the localization challenge. The first idea is to harness information from multiple layers in the convolutional network to conserve as much of the object boundaries information as possible.[33]

The second strategy is to seek a superpixel representation, essentially delegating the localization task to a low-level segmentation method.[34]

Labeling of samples is also a research challenge for CNN training.[35] This is mainly due to the fact that compared with traditional supervised methods, deep CNN-based methods usually need more training samples to overcome the overfitting problem.[36] The number of labeled datasets in practice is usually far from sufficient for a deep CNN architecture training. In view of scale variance of natural objects, the main important concerns for the samples labeling are related to the size and redundancy of the training set.[30] The size and quality of the training set have a direct impact on the execution time needed for training and on the final result of the classification.[30] The training set must, thus, be carefully chosen, avoiding redundancy patterns, but also ensuring a good representation of the considered classes.

To tackle these issues, multiscale segmentation approach is attractive.[37] It generates a set of segmentations of the same image at different scales, which allows that image objects are identified or extracted at several levels of segmentation detail.[38] The benefit is twofold. First, image segmentation has a potential to improve the localization accuracy. It preserves the abundant boundary information by partitioning a given image into a number of homogeneous regions.[39] Embedding the object boundary information of multiscale segmentations into the CNN can shape the real contour of geo-objects and filter out the spurious areas, thus potentially improving localization accuracy.[40] Furthermore, multiscale segmentation benefits for the sample selection. Image segmentation allows to extract both the objects and object contexts such as shape and texture based on the homogeneous regions in the segmentation results. In this way, samples can be directly selected from the multiscale segmentation results, where objects can be well-presented in their own observation scales.

In this paper, we presented a combined multiscale segmentation convolution neural network (CMSCNN) for rapid earthquake damage mapping. First, multiscale training samples database is constructed based on multiscale segmentation algorithm. Specifically, the mean shift (MS) algorithm is first conducted on the postearthquake VHR remote sensing images to derive regions at three different scales. Multiscale training samples for each class are directly selected from the three types of regions, and then resampled to the same size, which serve as CNN inputs. Then, CNN is trained based on the multiscale training samples database to exploit the complex features of damaged objects, which do benefit to generate the classification results with accurate identification of postearthquake geo-objects. Finally, to account for the boundary information of geo-objects with various sizes, an iteratively region-based max voting is conducted based on the multiscale segmentations derived from the first step, to generate the final multiscale classification maps. Several experiments were conducted in four postearthquake VHR images, and the results demonstrate that CMSCNN is effective for rapid high-resolution damage mapping and has substantial practical merit.

The rest of this article is structured as follows: Sec. 2 presents the proposed CMSCNN for postearthquake VHR image classification. Experimental results of the proposed method and the comparisons with other methods are reported in Sec. 3. Discussions of the experimental results are given in Sec. 4. Finally, some conclusions are drawn in Sec. 5.

## 2 Proposed Method

In this paper, a CMSCNN method is proposed. As shown in Fig. 1, the proposed CMSCNN includes the following three steps: (1) multiscale training samples database construction; (2) preliminary CNN classification; and (3) combine multiscale segmentations and CNN classification results.

### 2.1 *Conventional Convolutional Neural Network*

The complexity of high-resolution postearthquake images causes traditional classification methods to fail due to the limited representation power of a few mapping layers. Compared with conventional neural networks, CNN performed much better for robust automatic feature extraction and complex object recognition in high-resolution images.[41] This is because CNN is

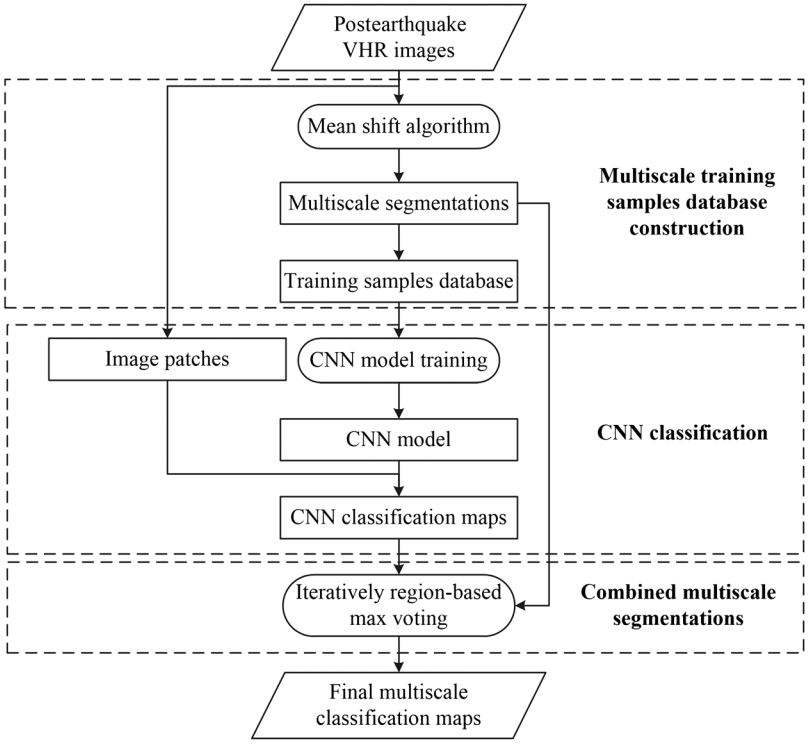


**Fig. 1** Overall workflow of the proposed CMSCNN framework.

characterized by the local receptive fields, the shared weights, and the spatial subsampling, which make it invariant to translation, scaling, skewing, and other forms of distortion.[42] A traditional CNN is a multilayer neural network that can model higher-level end-to-end features, hierarchically. Commonly, a typical CNN contains two types of major layers, named convolutional layer and subsampling layer that behave as feature detectors.

The convolutional layer offers a filter-like function to generate convoluted feature maps. At $l$'th convolution layer, the feature maps of $(L-1)$'th layer are first convolved with learnable filters $k$, and then put through the activation function $g(\cdot)$ to produce the $l$'th layer output feature map. The activation function in the neural network is commonly specified to be the sigmoid function $g(\cdot) = (1+e^{-x})^{-1}$. In general, the $l$'th convolution layer $C^l$ can be defined as

$$C^l = g(k^l h^{l-1} + b^l), \tag{1}$$

where $h^{l-1}$ refers to the hidden layer, with $h^0$ being raw input. $b^l$ is the bias term of $l$'th layer feature map. During the training of a convolutional layer, each filter $k$ slides over the entire image and produces feature maps. Unlike experience-guided spatial filter selection, convolutional layers can automatically learn and choose the best filter for the entire network.

The subsampling layer can generalize the features produced by previous layers, which will make features more robust and further reduce the computational complexity during the training progress. Through the subsampling operation, feature maps shrink but become more and more general and robust. Subsampling layers are defined as follows:

$$S^l = g[\text{down}(h^{l-1}) + b^l], \tag{2}$$

where $\text{down}(\cdot)$ represents a subsampling function. Typically, it will sum over each distinct $n$-by-$n$ block in the input map so that the output feature maps are $n$-times smaller than previous ones. Each output map is given its own additive bias parameter $b^l$, which is similar to convolutional layers. In this way, spatial-related features can be generated layer by layer and become more and more abstract and robust.

## 2.2 *Framework of Combined Multiscale Segmentation Convolutional Neural Network*

It is noticeable that deep features extracted by the traditional CNN are generally robust and effective for complex image pattern descriptions, especially for the case of high-resolution damaged scenes. However, the deep features are extracted from receptive fields with fixed sizes by traditional CNN.[19,28] Geo-objects in postearthquake images are often observed at various scales, which implicates the traditional CNN with fixed receptive fields unsuitable for postearthquake images classification. Moreover, due to the subsampling processes of CNN, deep spatial-related features with high-level abstractions naturally fail to detect the edges and contours of complex objects. Hence, the output classification result of CNN is commonly coarse-resolution that is characterized by accurate identification but poor delineation of geo-objects. Accordingly, to improve classification accuracy, the CMSCNN is designed due to its ability to learn multiscale deep features and to keep clear edges of objects in classification results.

### 2.2.1 *Multiscale training samples database construction*

To obtain deep feature representations, one priority is how to obtain enormous training samples for CNN. Generally, they are selected manually pixel by pixel. However, this method is time consuming and labor intensive, which is impractical for fast response to rapid earthquake damage mapping. In addition, due to the significant scale variance inherent in postearthquake objects, it is hard to select the accurate and homogeneous samples without any prior information. Multiscale segmentation algorithm can produce a set of homogeneous regions at different observation scales.[43] Therefore, in this paper, we utilize the regions produced by multiscale segmentation algorithm to generate training samples, which favors better generalization of the training samples.

Figure 2 shows the training samples by the use of multiscale segmentation algorithm. As shown in Fig. 2, the postearthquake VHR image is divided into a series of small regions at different scales using a segmentation algorithm. Then, for each class, the patches centered on the region are extracted as training samples for it. Moreover, considering the scale variance of different classes, the individual samples for each class are identified at the optimal level of segmentation details. In detail, samples of each class are first selected at three different sizes, and then all three types of samples are resampled to the smallest size. Finally, all training samples selected at their optimal scales form the training database.

The multiscale segmentation algorithm is significant for the construction of sample database. MS is a state-of-the-art segmentation algorithm with the advantages of simple parameter setting and no requirement for any prior knowledge.[2,44,45] Moreover, the ability of MS to maintain the saliency as well as the edge information has contributed to its wide applications in segmenting complex natural images, especially those without prior information. Therefore, we adopt MS here as the multiscale segmentation algorithm due to its outstanding ability of preserving the object boundaries.

### 2.2.2 *Convolutional neural network classification*

In this paper, we employed a two-layer CNN framework to produce the preliminary classification maps. The framework is shown in Fig. 3.

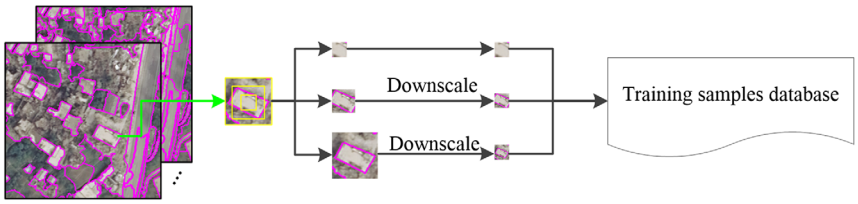


**Fig. 2** Multiscale training samples selection using multiscale segmentation algorithm.

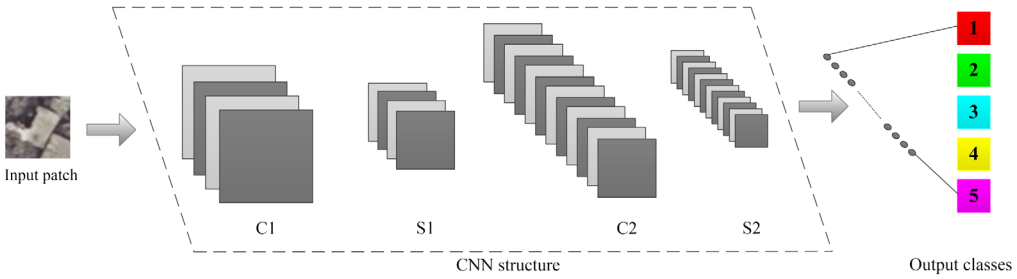


**Fig. 3** The general CNN framework. C1, S1, C2, and S2 represent the first and the second convolutional layer and subsampling layer, respectively.

As shown in Fig. 3, there are two convolutional layers followed by subsampling layers in the CNN. The CNN model is first trained using the multiscale training samples database. The parameters of this model empirically tuned as introduced in Sec. 3. Then, the trained CNN is used for images classification. Noteworthy, as for the inputs, for each pixel $p = (x, y)$ to be labeled, a patch of size $w \times w$ is extracted centered on $p$, which can take spatial information about the centered pixel under consideration.

### 2.2.3 *Combined multiscale segmentations*

CNN predicts the presence and rough positions of target objects, but it has poor delineation for object borders, e.g., neglected spatial consistency. As a consequence, nonsharp boundaries and spurious regions will be generated, which tend to amplify the damage mapping uncertainty. For instance, an intact building in the collapsed ruins may appear as a speckle in the classification map. Multiscale segmentation algorithm is considered as a promising process to address this problem. The regions and region contexts information are well-preserved in the multiscale segmentation results. There is a possibility to combine multiscale segmentations and the coarse classification of CNN to improve the localization accuracy of objects. It is supposed that coarse-resolution classification of CNN is important to differentiate between different objects while fine resolution segmentations is necessary for localization.

Therefore, the classification of CNN is combined with multiscale segmentations by max voting to reach a more detailed classification result. The scheme refines the classification from fine to coarse iteratively. The combination starts with the fine segmentation to produce the classified image, and then the classified image is regarded as the new classification to be combined with the following segmentation. The reason behind this is to avoid the situation that coarse segmentation tends to neglect the small objects and the fine segmentation inclines to generate the spurious regions in geo-object. The scheme consists of three steps. First, the classification is mapped to each segmentation to assign classification labels to each pixel accordingly. Then, instead of using single pixels, region-based segmentation is regarded as the classification unit. In the voting process, CMSCNN randomly selects a region and predicts its label by assuming that the region belongs to the label that accounts for the majority of labeled values from all pixels within that region, whereas the other labels contained in the region are inferred as either noise or misclassified object boundaries. So, the label $SC_r$ of region $r$ is assorted as

$$SC_r = \arg\ \max_{t=1}^{N} \sum_i \sum_j \text{sign}[f_{r(i,j)} = t], \tag{3}$$

where $f_{r(i,j)}$ is the label of the pixel $r(i, j)$ in region $r$ from the initial classification, and $(i, j)$ is the coordinates of the pixel $r(i, j)$. $N$ is the total number of expected classes. Finally, the combination is continued until all the segmentations are implemented.

In this way, the multiscale classification result confirms the accuracy of the CNN recognition, and refines different objects by the abundant multiscale information contained in segmentations. Thus, the CMSCNN learns the features of different levels, creating more robust classifiers.

## 3 Experiments

### 3.1 *Study Area and Data Description*

The experiments were performed on four postearthquake VHR images, including three subarea images and one urban image. Three test subarea images T1 to T3 (the spatial resolution is 0.67 m) were acquired three days after a violent Ms 8.0 earthquake struck in Wenchuan, China, on May 12, 2008, captured by RGB sensors mounted on aerial platforms. The earthquake was centered at ~30.98°N and 103.36°E. The focal depth of this earthquake was 14 km and the earthquake devastated a huge area in Wenchuan County. These test images cover a variety of damage objects, such as landslides, debris flow, and collapsed residential sites. Some of the study areas contain certain portions of collapsed residential buildings, whereas others are partly covered by the landslides or debris flow. That is to say, the mapping of the selected damaged areas is difficult and challenging.

To assess the accuracy of CMSCNN, we used the ground-truth images as the references in this study. The reference images of the four test images were manually interpreted in commercial software eCognition[46] by different experienced experts.

### 3.2 *Experiments Setups*

First of all, the segmentations were produced by MS to obtain the boundary information of objects at three different scales on these test images. Three scale parameters will affect the performance of the MS, named the window widths of color, spatial domain, and the minimum area size. For T1 in the first row, the window widths of color/spatial domain are set to 7/4, 8/6, and 9/4 pixels, respectively. As for T2 in the second row, the window widths of color/spatial domain are set to 7/6.5, 8/6.5, and 9/5.5 pixels, respectively. In terms of T3 in the third row, the window widths of color/spatial domain are set to 15/4, 20/4, and 24/5 pixels, respectively. The minimum area size is set to 20 pixels as default.[44] In addition, the training samples for each class are selected from the segmentations at three scales of $20 \times 20$, $22 \times 22$, $24 \times 24$, and then downscaling them with a size of $20 \times 20$ pixels. And the patch of input data is set as $5 \times 5$.The resultant segmentations were shown in Fig. 4, where the first to the fourth rows represented the test images T1 to T3, and the second to the fourth columns were the segmentations at three scales from fine to coarse, respectively.

From Fig. 4, we can find that all the interesting boundaries are clearly present in the images. When different threshold values were applied to the original test images [e.g., Fig. 4(a)], multiscale segmentation results at different scales could be obtained corresponding to a level of the boundary details [e.g., Figs. 4(b)–4(d)].

The architecture of CMSCNN consists of two subsequent layers, as shown in Fig. 3. Each layer is composed of cascading structure with convolutional and polling stages. In terms of pooling stages, they are all set to be subsampling by a small factor of two to achieve a compromise between efficiency and accuracy. For the convolutional layers, we set the kernel size to 5, and set number of feature maps to 6 and 12 for two convolutional layers, respectively, by considering the size of training samples. Mini-batch strategy is adopted to update trainable parameters in the nets and the size of the training batch is set to 100 samples each. Learning rate is controlled as one and the number of training epochs is set to 200 to ensure the nets converges both quickly and accurately.

To verify the superiority of the proposed CMSCNN, two popular algorithms, including the support vector machine (SVM) and the conventional CNN, are adopted as the compared algorithms. SVM is advanced supervised kernel classification approach.[47] Specifically, the RBF kernel function was selected, and the parameter Gamma in kernel function was set to 0.015. To have a fair comparison, the classification results of SVM were also improved by the use of iteratively voting based on the same multiscale segmentations as CMSCNN. Therefore, we named the SVM here as combined multiscale segmentations SVM (CMSSVM). As for the conventional CNN, it is designed to demonstrate the superiority of the combined multiscale segmentations scheme of CMSCNN. Therefore, the training samples database and the parameters of CNN structure are all

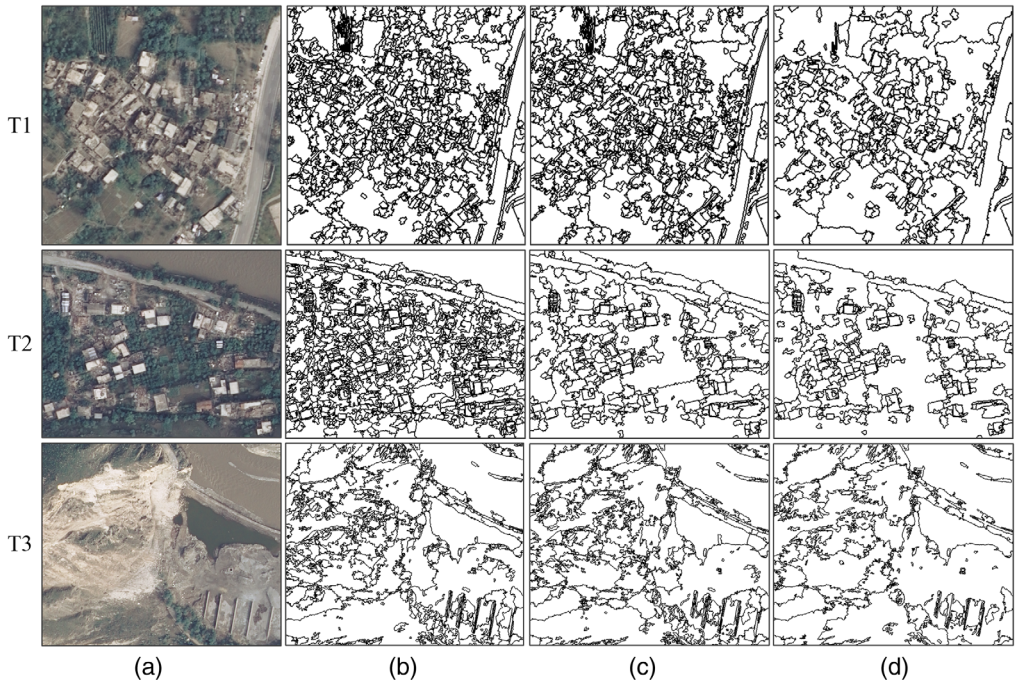


**Fig. 4** Three scale segmentation results of four test images. (a) The original postearthquake image: from top to down are T1 to T3. (b)–(d) The first to fourth rows are the three segmentations at different scales from fine to coarse for T1 to T3.

the same as CMSCNN. Moreover, the comparative evaluations of the three different algorithms are conducted qualitatively and quantitatively.

## 3.3 *Comparative Evaluations of Experimental Results*

### 3.3.1 *Qualitative evaluation*

The classification results are shown in Fig. 5, where Fig. 5(d) shows the ground-truth images of T1 to T3, Figs. 5(b) and 5(c) shows the classification results of T1 to T3 by conventional CNN and CMSSVM, respectively, and the results from CMSCNN are given in Fig. 5(c).

Generally, as we can see from Fig. 5, the classification results indicate the superiority of the proposed CMSCNN compared with CNN and CMSSVM. Specifically, CNN classification results showed a salt-and-pepper appearance throughout study areas [Fig. 5(a)], whereas the classification maps of CMSCNN and CMSSVM were much more homogeneous [Figs. 5(b) and 5(c)]. However, as shown in Fig. 5(b), CNN has poor object delineation for intact buildings and the roads, resulting in the nonsharp or speckled regions, as shown in objects labeled in ellipses of T3 study area in Fig. 5(b). This is probably because that the convolution and max-pooling processes in CNN ignored the subtle details of objects. In addition, the CNN detected many spurious objects especially those with varied sizes and similar spectral characteristics, such as the landslides and the grassland (shown in red ellipses corresponding to T3 site in Fig. 5). This is mainly due to the fixed extractor of CNN, which limited the recognition of objects at different scales.

In contrast to the results of CNN, as shown in Fig. 5(c), the classification results of CMSCNN are more consistent and with satisfied structures. It is worth noting that compared with CNN, the classifications by CMSCNN show more realistic object shapes with merit of clear boundaries, especially the collapsed buildings areas with varied sizes [in yellow ellipses of T1 and T2 areas in Fig. 5(c)] and the landslides [in red ellipses of T3 in Fig. 5(c)]. It indicates the effectiveness of the combination strategy of CMSCNN, which combines CNN with multiscale segmentations.

As shown in Fig. 5(b), compared with the proposed CMSCNN, we can find that CMSSVM generally shows a similar appearance, where most object boundaries are correctly extracted, as it

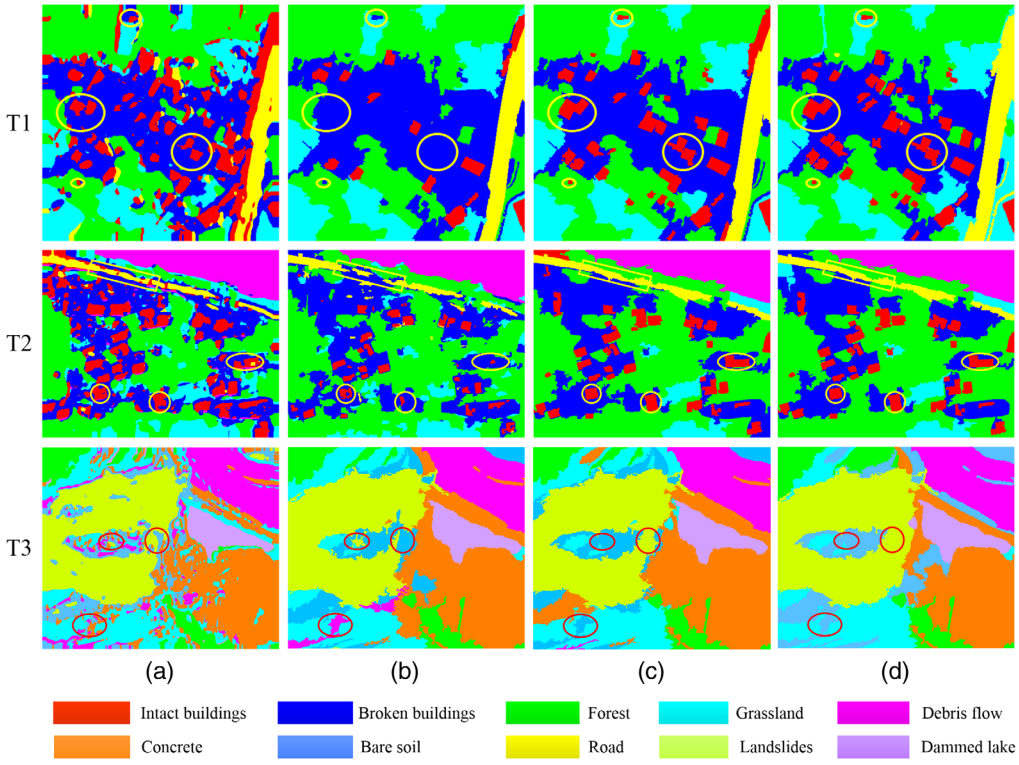


**Fig. 5** The classification results of four test images by (a) CNN, (b) CMSSVM, (c) CMSCNN, and (d) the ground truths. From top to down are the results of T1 to T3 study areas.

adopted the same scheme of combining multiscale segmentations as CNN. However, there are some misclassifications existing. For example, some isolated, thin, and elongated objects, such as roads, are confused with their surroundings, as shown in yellow rectangles of T2 site corresponding to Fig. 5(c). In addition, CMSSVM ignored some objects, such as the intact buildings in the yellow ellipses of T1 and T2 sites in Fig. 5(c). Further, CMSSVM also misclassified some damaged areas into the intact buildings. This discovery also indicates that, compared with traditional machine learning algorithms, the CNN can adopt more satisfied classification results for complex geo-objects.

### 3.3.2 *Quantitative evaluation*

For quantitative assessment, the confusion matrixes are adopted to evaluate the performance of CMSCNN quantitatively. Table 1 summarizes the classification accuracy by three different methods for test images T1 to T3. For the accuracy of each class, the best obtained results were shown in boldface, and PA represents the product accuracy, UA represents the user accuracy, OA represents the overall accuracy, and $K$ represents the Kappa coefficient.

As we can see from Table 1, CMSCNN has significantly improved the classification accuracy for all four test areas. For example, compared with CNN and CMSSVM, CMSCNN has increased the overall accuracy (OA) by 20.04% and 16.23%, as well as increased the kappa coefficient by 0.27 and 0.21, respectively. These quantitative indicators can also demonstrate the above findings. For example, from the classification maps, we can find that CMSSVM seriously confused the intact buildings into the broken building areas in T1, resulting in its small user accuracy of broken buildings in Table 1. In terms of T2 site, CMSCNN still performed best by the fact that OA was 96.88% and the Kappa coefficient was 0.96, significantly higher than those of CNN (70.60% and 0.60) and taking substantial advantage over those of CMSSVM (83.19% and 0.76). With respect to T3 shown in Table 1, CMSCNN again obtained the best results, which were slightly higher than CMSSVM but significantly outperformed CNN.

**Table 1** The classification accuracy of T1 to T3 images using CNN, CMSSVM, and CMSCNN.

| Method | | CNN | | | | CMSSVM | | | | CMSCNN | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Test images | /% | PA | UA | OA | K | PA | UA | OA | K | PA | UA | OA | K |
| **T1** | Forest | 75.81 | 92.82 | 71.62 | 0.62 | **94.07** | **94.61** | 87.85 | 0.83 | 85.57 | 84.65 | **92.56** | **0.9** |
| | Broken buildings | 74.83 | 66.9 | | | **99.99** | 74.86 | | | 96.74 | **94.93** | | |
| | Road | 65.86 | 69.42 | | | 82.54 | **100** | | | **99.99** | 87.41 | | |
| | Grassland | 61.82 | 80.76 | | | 80.81 | **100** | | | **82.5** | 96.91 | | |
| | Intact buildings | 70.74 | 35.04 | | | 40.35 | 82.02 | | | **80.11** | **100** | | |
| **T2** | Intact buildings | 79.83 | 46.37 | 70.6 | 0.6 | 45.41 | **93.41** | 83.19 | 0.76 | **90.89** | 92.63 | **96.88** | **0.96** |
| | Forest | 63.14 | 96.71 | | | 87.83 | 92.53 | | | **95.94** | **99.98** | | |
| | Broken building | 69.69 | 61.08 | | | 86.97 | 71.06 | | | **97.93** | **90.83** | | |
| | Road | 67.68 | 80.41 | | | 69.79 | 79 | | | **100** | **100** | | |
| | Grassland | 83.96 | 25.97 | | | 58.6 | 45.47 | | | **88.27** | **99.96** | | |
| | Debris flow | 79.83 | 46.37 | | | 96.23 | **100** | | | **100** | **100** | | |
| **T3** | Grassland | 72.43 | 54.43 | 74.37 | 0.69 | 82.64 | 79.97 | 87.42 | 0.84 | **95.31** | **80.08** | **92.49** | **0.91** |
| | Forest | 60.43 | 87.49 | | | **74.6** | **100** | | | **74.6** | **100** | | |
| | Debris flow | 88.77 | 95.88 | | | **100** | 87.01 | | | **100** | **97.9** | | |
| | Dammed lake | 79.92 | 39.75 | | | **96.61** | **100** | | | **96.61** | **100** | | |
| | Bare soil | 62.65 | 43.53 | | | 66.08 | 70.23 | | | **97.15** | **92.15** | | |
| | Landslides | 83.23 | 97.14 | | | 89.68 | 96.88 | | | **91.58** | **100** | | |
| | Concrete | 67.42 | 92.34 | | | **93.92** | **83.13** | | | 85.68 | 80.21 | | |

## 4 Discussions

In this paper, we proposed a CMSCNN algorithm for the classification of postearthquake VHR remote sensing images. Due to the rich spatial information contained in high-resolution images and the complexity of the various damaged geo-objects, feature extraction is one of the biggest challenges for the analysis of postearthquake images. Traditional image classification methods require numerous image features to be empirically designed and depend on the knowledge of the analysts, which are time-consuming and often fail to achieve accurate interpretation of image. Recent machine learning methods are too limited to recognize the complex damage objects due to their shallow structures. As a result, few image classification methods in these studies can be practically used effectively and efficiently in postearthquake images. In this paper, the CNN, a well-known deep learning method, was chosen for automatic feature learning of postearthquake VHR images. With the hierarchical structure of the CNN, image features at higher levels can be automatically extracted. Moreover, CNN has shown satisfied robustness and accuracy in detecting complex targets. As the abstraction level increased, the extracted deep features demonstrated strong invariance in terms of semantic content. However, the method often fails to consider the scale variety of geo-objects due to its receptive fields with fixed sizes and fails to capture boundary information of objects due to the subsampling processes.

To this end, we combine the multiscale segmentations with preliminary CNN classification results for both efficient multiscale training sample selection and better extractions of targets' boundary. As demonstrated in our experiments, the combination of image objects and deep features is quite effective. For one thing, it alleviates people from the time-consuming process of training sample selection and allows choosing more optimal training samples for each target class at various scales. For another, the combination of multiscale segmentations and deep learning method provides accurate targets' localizations as well as identifications in the multiscale classified images. In addition, the final classification is capable to capture various targets due to the consideration of multiscale information.

## 5 Conclusions

Rapid damage mapping has always been a fundamental but challenging issue in the field of damage assessment and emergency rescue. More accurate and efficient classification methods for the postearthquake high-resolution images are required. This paper presents a CMSCNN approach to combine deep CNN with multiscale segmentations, and demonstrated its usefulness in rapid damage mapping in postearthquake VHR images. The results showed that the methodology is able to accommodate the rapid damage mapping. This is contributed to the following three processes: (1) the selection of multiscale training samples, which is efficient and beneficial to the rapid formation of the training sample database of the damages encountered; (2) the simple structure (two layers) and a few training iterations (200 times) contributed to intense the workflow for the practical application; (3) the combination of multiscale segmentations and preliminary CNN classification results, which significantly improved the accuracy in both localization and classification in comparison to the conventional CNN and the CMSSVM technique. The quantitative and qualitative evaluations also validated the fact that such a scheme renders CMSCNN simple, practical, and appropriate for rapid high-resolution damage mapping.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (41801275 and 41471353), the Shandong Provincial Natural Science Foundation, China (ZR2018BD007), the Fundamental Research Funds for the Central Universities (18CX05030A and 18CX02179A).

## References

1. L. Dong and J. Shan, "A comprehensive review of earthquake-induced building damage detection with remote sensing techniques," *ISPRS J. Photogramm. Remote Sens.* **84**, 85–99 (2013).

2. G. Sun et al., "Dynamic post-earthquake image segmentation with an adaptive spectral-spatial descriptor," *Remote Sens.* **9**, 899 (2017).
3. Y. Aimaiti, F. Yamazaki, and W. Liu, "Multi-sensor InSAR analysis of progressive land subsidence over the Coastal City of Urayasu, Japan," *Remote Sens.* **10**(8), 1304 (2018).
4. G. Cheng et al., "Effective and efficient midlevel visual elements-oriented land-use classification using VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.* **53**, 4238–4249 (2015).
5. C. Geiß et al., "Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques," *ISPRS J. Photogramm. Remote Sens.* **104**, 175–188 (2015).
6. T. Blaschke et al., "Geographic object-based image analysis—towards a new paradigm," *ISPRS J. Photogramm. Remote Sens.* **87**, 180–191 (2014).
7. A. Vetrivel et al., "Towards automated satellite image segmentation and classification for assessing disaster damage using data-specific features with incremental learning," in *Proc. GEOBIA* (2016).
8. F. Yamazaki et al., "Earthquake damage detection using high-resolution satellite images," in *IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Vol. 2284, pp. 2280–2283 (2004).
9. Z. Ren and A. Lin, "Co-seismic landslides induced by the 2008 Wenchuan magnitude 8.0 earthquake, as revealed by ALOS PRISM and AVNIR2 imagery data," *Int. J. Remote Sens.* **31**, 3479–3493 (2010).
10. F. Nex et al., "Automated processing of high resolution airborne images for earthquake damage assessment," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XL-1**, 315–321 (2014).
11. K. K. Singh and A. Singh, "Detection of 2011 Sikkim earthquake-induced landslides using neuro-fuzzy classifier and digital elevation model," *Nat. Hazards* **83**, 1027–1044 (2016).
12. N. Tamkuan and M. Nagai, "Fusion of multi-temporal interferometric coherence and optical image data for the 2016 Kumamoto earthquake damage assessment," *ISPRS Int. J. Geo-Inf.* **6**, 188 (2017).
13. V. Walter, "Object-based classification of remote sensing data for change detection," *ISPRS J. Photogramm. Remote Sens.* **58**, 225–238 (2004).
14. L. Gong et al., "Earthquake-induced building damage detection with post-event sub-meter VHR Terrasar-x staring spotlight imagery," *Remote Sens.* **8**, 887 (2016).
15. T. T. Vu, "Rapid disaster damage estimation," *ISPRS—Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **xxxix-b8**, 65–69 (2012).
16. F. Samadzadegan, H. Rastiveisi, and W. G. Viii, "Automatic detection and classification of damaged buildings, using high resolution satellite imagery and vector data," in *The International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences* Vol. **37**, pp. 415–420 (2008).
17. G. T. Kaya, O. K. Ersoy, and M. E. Kamaşak, "Spectral and spatial classification of earthquake images by support vector selection and adaptation," in *Int. Conf. Soft Computing and Pattern Recognition*, pp. 194–197 (2010).
18. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**, 436–444 (2015).
19. H. Liu et al., "Single image super-resolution using multi-scale deep encoder-decoder with phase congruency edge map guidance," *Inf. Sci.* **473**, 44–58 (2019).
20. W. Zhao and S. Du, "Learning multiscale and deep representations for classifying remotely sensed imagery," *ISPRS J. Photogramm. Remote Sens.* **113**, 155–165 (2016).
21. F. Shen et al., "Unsupervised deep hashing with similarity-adaptive and discrete optimization," *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(12), 3034–3044 (2018).
22. W. Zheng, "Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis," *IEEE Trans. Cognit. Dev. Syst.* **9**(3), 281–290 (2017).
23. J. Han et al., "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Trans. Geosci. Remote Sens.* **53**, 3325–3337 (2015).
24. O. Walter et al., "Autonomous learning of representations," *KI—Künstliche Intell.* **29**, 339–351 (2015).

25. P. Jia et al., "Convolutional neural network based classification for hyperspectral data," in *IEEE Int. Geoscience and Remote Sensing Symp.*, pp. 5075–5078 (2016).
26. K. Makantasis et al., "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, pp. 4959–4962 (2015).
27. S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing* **219**, 88–98 (2017).
28. K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.* **61**, 539–556 (2017).
29. J. Li, R. Zhang, and Y. Li, "Multiscale convolutional neural network for the detection of built-up areas in high-resolution SAR images," in *IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, pp. 910–913 (2016).
30. C. Farabet et al., "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1915–1929 (2013).
31. S. Gidaris and N. Komodakis, "Object detection via a multi-region and semantic segmentation-aware CNN model," in *IEEE Int. Conf. on Computer Vision*, pp. 1134–1142 (2015).
32. L. C. Chen et al., "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 834–848 (2018).
33. E. Li et al., "Integrating multilayer features of convolutional neural networks for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.* **55**, 5653–5665 (2017).
34. S. He et al., "SuperCNN: a superpixelwise convolutional neural network for salient object detection," *Int. J. Comput. Vision* **115**, 330–344 (2015).
35. G. Wang et al., "Aggregating rich hierarchical features for scene classification in remote sensing imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10**, 4104–4115 (2017).
36. A. Vetrivel et al., "Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning," *ISPRS J. Photogramm. Remote Sens.* **140**, 45–59 (2018).
37. R. Trias-Sanz, G. Stamon, and J. Louchet, "Using colour, texture, and hierarchial segmentation for high-resolution remote sensing," *ISPRS J. Photogramm. Remote Sens.* **63**, 156–168 (2008).
38. X. Zhang et al., "Hybrid region merging method for segmentation of high-resolution remote sensing images," *ISPRS J. Photogramm. Remote Sens.* **98**, 19–28 (2014).
39. H. C. Shih and E. R. Liu, "New quartile-based region merging algorithm for unsupervised image segmentation using color-alone feature," *Inf. Sci.* **342**, 24–36 (2016).
40. G. J. Hay et al., "An automated object-based approach for the multiscale image segmentation of forest scenes," *Int. J. Appl. Earth Obs. Geoinf.* **7**, 339–359 (2005).
41. J. Gu et al., "Recent advances in convolutional neural networks," *Pattern Recognit.* **77**, 354–377 (2018).
42. Y. Liu et al., "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion* **36**, 191–207 (2017).
43. N. Audebert, B. L. Saux, and S. Lefèvre, "How useful is region-based classification of remote sensing images in a deep learning framework?" in *IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, pp. 5091–5094 (2016).
44. D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 603–619 (2002).
45. T. Su et al., "Image segmentation using mean shift for extracting croplands from high-resolution remote sensing imagery," *Remote Sens. Lett.* **6**, 952–961 (2015).
46. U. C. Benz et al., "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *ISPRS J. Photogramm. Remote Sens.* **58**, 239–258 (2004).
47. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**, 273–297 (1995).

**Hui Huang** is a master's student at the School of Geosciences, China University of Petroleum (East China). She received her BS degree at China University of Petroleum (East China) in 2017.

Her research interests include high-resolution remote sensing image processing and deep learning.

**Genyun Sun** is currently an associate professor with China University of Petroleum, Qingdao, China. His research interests include remote sensing image processing, intelligent optimization algorithm and machine learning.

**Xuming Zhang** is a master's student at the School of Geosciences, China University of Petroleum (East China). She received her BS degree at China University of Petroleum (East China) in 2018. Her research interests include remote sensing image processing and deep learning.

**Yanling Hao** is a master's student at the School of Geosciences, China University of Petroleum (East China). She received her BS degree at China University of Petroleum (East China) in 2015. Her research interests include remote sensing image processing and deep learning.

**Aizhu Zhang** is currently working toward the PhD with the School of Geosciences, China University of Petroleum (East China). Her research interests include remote sensing image processing, pattern recognition, and intelligence optimization algorithms.

**Jinchang Ren** is currently the deputy director, Strathclyde Hyperspectral Imageing Centre for Signal and Image Processing (CeSIP), Department of Electronic and Electronic and Electrical Engineering University of Strathclyde. His research interests include intelligent information processing, visual computing and multimedia signal processing.

**Hongzhang Ma** is currently an associate professor with the College of Science, China University of Petroleum, Qingdao, China. His research interests include microwave remote sensing of soil moisture and application technology of multisource remote sensing data.