A Novel Deep Multi-Instance Convolutional Neural Network for Disaster Classification From High-Resolution Remote Sensing Images

Chengfan Li, Zixuan Zhang, Lan Liu, Jung Yoon Kim, and Arun Kumar Sangaiah

***Abstract*—The fully supervised deep convolutional neural network (CNN) cannot detect the discriminant local information that is responsible for spatial transformations in high-resolution remote sensing images. To address the various types and missing labels of natural disasters, a new deep multi-instance convolutional neural network (DMCNN) model for disaster classification in high-resolution remote sensing image is presented in this article. Specifically, based on sample enhancement and atrous spatial pyramid pooling, we first extract and integrate the features via the CNN structure to obtain the instance feature of bags in the image. Besides, introducing a prototype learning layer with distance measure, the instance features extracted from pretrained CNN are mapped into a series of prototype instance features with bag-level. Subsequently, all instance features from prototype and bag take part in disaster detection and image classification. Finally, we conduct extensive experiments on xBD dataset and discussions from qualitative and quantitative aspects. Experimental results show that the proposed DMCNN model achieves better classification accuracy of natural disaster from high-resolution remote sensing images compared to traditional CNNs, and effectively improves the disaster classification performance with weakly supervised from high-resolution remote sensing images.**

Manuscript received 5 June 2023; revised 24 August 2023 and 16 October 2023; accepted 4 December 2023. Date of publication 11 December 2023; date of current version 2 January 2024. This work was supported in part by the Natural Science Foundation of Shanghai under Grant 22ZR1423200, in part by the Key Laboratory of National Geographic Census and Monitoring, Ministry of Natural Resources, Wuhan University under Grant 2022NGCM12, in part by the Shanghai Foundation for Development of Science and Technology under Grant 21142202400, and in part by the Key Laboratory for Digital Land and Resources of Jiangxi Province, East China University of Technology under Grant DLLJ202103. *(Corresponding authors: Lan Liu; Jung Yoon Kim.)*

Chengfan Li is with the School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China, and with the Key Laboratory of National Geographic Census and Monitoring, Ministry of Natural Resources, Wuhan University, Wuhan 430079, China, and also with the Key Laboratory for Digital Land and Resources of Jiangxi Province, East China University of Technology, Nanchang 330013, China (e-mail: lchf@shu.edu.cn).

Zixuan Zhang is with the School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China (e-mail: zixuanzhang@shu.edu.cn).

Lan Liu is with the School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China (e-mail: liulan@sues.edu.cn).

Jung Yoon Kim is with the College of Future Industry, Gachon University, Seongnam-Si, Gyeonggi-do 13120, South Korea (e-mail: kjyoon@gachon.ac.kr).

Arun Kumar Sangaiah is with the International Graduate School of Artificial Intelligence, National Yunlin University of Science and Technology, Douliu 64002, Taiwan, and also with the Department of Electrical and Computer Engineering, Lebanese American University, Byblos 11020, Lebano (e-mail: aksangaiah@ieee.org).

Digital Object Identifier 10.1109/JSTARS.2023.3340413

***Index Terms*—Deep learning, disaster classification, high-resolution remote sensing image, prototype representation.**

I. Introduction

DISASTERS, including natural disasters and man-made disasters, can cause harm and huge losses to human life and property safety [1], [2]. In general, natural disasters often cause more damage than man-made disasters because of the limited human disturbance [3], [4]. In terms of the differences of natural disasters, the type of natural disaster can be divided into geological-geomorphic disaster (e.g., earthquakes, volcanoes, debris flows), meteorological-hydrological disaster (e.g., floods, droughts, sandstorms, tsunamis), environmental pollution disaster (e.g., forest-steppe fires, garbage pollution), ecological disaster (e.g., biological disasters), etc.

As the performance improvement of satellite sensor, its advantages on wide coverage, low revisit period, and high spatial resolution are prominent, so it can acquire quickly and accurately detailed features of ground objects and be widely used in disaster monitoring. However, compared with traditional natural images, the complex types, uneven distribution, and multiscales of ground objects in high-resolution remote sensing images are very significant. To some extent, it increases the difficulty to extract feature of ground object from high-resolution remote sensing [5], [6]. Traditional classification method of remote sensing images is mainly implemented by using visual-based and pixel interpretation, which are usually used as an effective auxiliary means for remote sensing image processing and analysis at present [7]. The subsequent object-oriented method can make full use of ground object information (i.e., spectrum features, space, shape, and texture), and has a good classification effect [8]. However, the reasonable segmentation windows and extracted features are not easily acquired and restrict the accuracy of object-oriented method.

By simulating the working mechanism of human brain neurons, deep learning can learn and make decisions from the complex data. As one of the representative deep learning methods, convolutional neural networks (CNNs) can automatically learn and extract in-depth image features from large-scale data by convolution layer, and overcome the ambiguity and uncertainty of remote sensing classification [9], [10]. At present, deep learning has been widely used in the field of disaster monitoring

based on high-resolution remote sensing images, such as volcanic ash detection with moderate-resolution imaging spectroradiometer, surface deformation with synthetic aperture radar, and earthquake-damaged buildings with UAV images [11], [12], [13], However, on the one hand, the depth of these network models and structure types are numerous and different in the field of disaster monitoring, it affects the overall performance of deep learning models and detection accuracy of ground object from high-resolution remote sensing image [14], [15]. On the other hand, the ground object shows clearly small target characteristics with multiscale in the high-resolution remote sensing image, to some extent, the atrous spatial pyramid pooling (ASPP) is able to enlarge the receptive field of network and has good detection effect for the multiscale targets. In addition, the lack of sufficient sample data often results in poor generalization and overfitting problems in disaster classification from high-resolution remote sensing images [16].

Multi-instance learning (MIL) is first to infer and predict whether unknown molecules are suitable for pharmaceuticals via learning the characteristics of known molecules [17]. Different from the one-to-one relationship between samples and features in traditional supervised learning, the training objects in MIL are divided into bags and instances, as well as the bag feature is composed of instance features and shows one-to-many relationship between samples and features. In theory, the depth features, obtained from the MIL framework, can be used in weakly supervised disaster classification from high-resolution remote sensing images because of more focus on the image (bag) and image subblock (instance) features.

By combining the deep learning model with the specific disaster classification of high-resolution remote sensing images, a reasonable deep learning network structure can be constructed and then appropriate sample datasets are learned and trained, which can not only obtain higher disaster classification accuracy from high-resolution remote sensing image but also improve the generalization and robustness of the proposed network model [18], [19]. To this end, we design and propose a new deep multi-instance convolutional neural network (DMCNN) by combining the feature extraction with prototype learning to implement the disaster classification from high-resolution remote sensing image. In the DMCNN model, the input image is first transformed by ASPP to enlarge the receptive field of the convolution layer based on sample enhancement. Subsequently, we extract and integrate the features via the CNN structure to obtain the instance feature of bags in the image, and then the instance features extracted by pretrained CNN network are mapped into a series of prototype instance features with bag-level. Finally, all instance features of disasters from prototype and bag are integrated together to detect and classify from high-resolution remote sensing images.

Overall, the contribution of this article can be summarized as follows.

1) The prototype learning layer is introduced into deep MIL structure to realize the learning of the relationship between instance features and image features from the perspective of bag-level.
2) Through incorporating multi-instance classifiers into an integrated deep learning network, a new DMCNN disaster classification model for high-resolution remote sensing image is proposed in this article and further validated on publicly xBD dataset. Experiments show that it can improve disaster classification accuracy and computation performance of high-resolution remote sensing images under weak supervision.

The rest of this article is organized as follows. Section II reviews related work based on MIL, CNN, and disaster monitoring with remote sensing images. Section III describes the intuition behind the DMCNN model as well as its structure in detail. Section IV presents the detailed experiments containing datasets and environments. Section V shows the experimental results and analysis from qualitative and quantitative aspects. Section VI devotes the discussions for the experimental results. Finally, Section VII concludes the article.

II. Related Work

A. Multi-Instance Learning

To improve the accuracy of image classification, the MIL framework focused on image block and subblock information via extracting and representing multilevel semantic features, is introduced into image processing [20], [21]. However, the small target feature is easy to be interfered by the surrounding noises, and this method is difficult to accurately obtain the subtle features of local image details. To address this problem, a new bag-level image classification method based on spatial multi-instance sparse representation has gradually emerged [22]. In this method, sparse coding features of training samples can be obtained by clustering algorithm and further used for classification by suitable classifier, and it can significantly enhance key local features in the high-resolution remote sensing image.

The axial-parallel rectangle algorithm is originally used specifically to solve the multi-instance problem. Since then, the MIL gradually forms a series of classical algorithms [23], for example, multi-instance algorithms with support vector machine (SVM) and CNN, multi-instance multilabel, multi-instance active learning, multi-instance clustering, etc. In line with the different levels of measurement objects, the MIL can be divided into two types, bag-based measurement and instance-based measurement [24]. The former identifies different label information by means of the feature discrimination algorithm with bag-level, whereas the latter identifies the label in virtue of inference for each instance based on the assumption that there is an equal contribution to the label on instance and bag. In summary, the latter has more computing power and speed than that of the former. However, in these traditional classical methods, features are often acquired through manual design rather than deep learning, and its real computational efficiency, accuracy, and robustness are limited when dealing with large amounts of data [19].

B. Deep CNN

At first, the CNN can be traced back to the back propagation network in 1986 [25]; subsequently, a number of classical models have been appeared, for example, AlexNet, VGGNet, GoogleNet, fully convolutional networks, ParseNet, etc. [26], [27], [28], [29], [30]. In general, the more layers the network

has, the more complex features can be learned, and the more precise the detection effect will be. However, with the increase of the number of network layers, it leads to the decline of classification accuracy because of gradient disappearing. To overcome the problem, a type of network introduction of residual learning unit structure and residual network variants were proposed one after another, such as ResNet, ResNeSt, Res2Net, DenseNet, and capsule network (CapsNet) [31], [32], [33], [34]. These new emerged models have the advantages of less training parameters and training difficulty and higher classification accuracy.

Deep learning is fully supervised and often used to learn depth feature representations from sample datasets. In line with the different emphasis, it can be summarized into the generative depth Boltzmann machine model for learning data structure and the discriminant CNN model for learning two-dimensional spatial information in images. At present, a new network structure combining MIL framework and feature extraction of deep learning appears and has attracted wide attention [35], [36], such as CNN extraction depth feature and multi-instance classifier. However, in these methods, CNN training structure and multi-instance classifier are either two independent parts as well as carried out separately or a complete multi-instance deep network structure that focus on first class labels and then class label integration with bag-level. It cannot learn and obtain the relationship among samples. In addition, for disaster classification with weakly supervised and small samples from high-resolution remote sensing image, it is difficult to provide accurate class label information for each sample [19].

C. Disaster Monitoring Based on Deep Learning for High-Resolution Remote Sensing Images

Using high-resolution remote sensing images for accurate disaster monitoring, the key constraints include high-quality remote sensing datasets and diverse types of natural disasters.

As is known to all, a large enough dataset is the data basis to ensure the accuracy of disaster classification from high-resolution remote sensing image. Based on the analysis of traditional target detection, semantic and instance segmentation algorithms, it is very important to explore precise disaster classification combined with the multiscale characteristics of ground objects in high-resolution remote sensing images with the specific type of disaster. So far, a series of remote sensing datasets has been initially formed for natural disasters in terms of geographical locations and data acquisition ways by sensors, for example, xBD [37] and AIDER [38]. Under the perfect conditions, it can detect and classify the multiple types of disaster information from different sources of remote sensing images to ensure the robustness and universality of the model. Even so, there is still a large number of exploratory work from different specific disaster scenes, such as the convolutional neural-deep belief network model [39], deep CNN model with manual mark samples [40], EmergencyNet model [41], flood disaster emergency information model [42], and postdisaster management model [43]. These works also confirm the effectiveness and feasibility of deep neural networks in disaster monitoring from high-resolution remote sensing image.

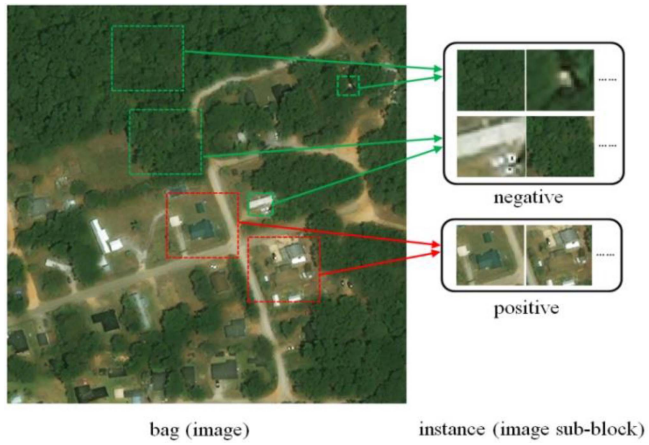

Fig. 1. Correspondence between MIL principle and disaster classification with high-resolution remote sensing image.

III. Methodology

A. Problem Formulation

A high-resolution remote sensing image can be regarded as a combination by a series of image subblocks with certain regular distribution. From the perspective of MIL, the high-resolution remote sensing image can be regarded as bag (image) and instance (image subblock) [44]. In the MIL framework, the positive instance is the image subblock marked as positive correlation with the disaster information in high-resolution remote sensing image, whereas the negative instance is the image subblock as negative correlation with the disaster information. However, the actual situation is that there is often a lot of noise information (nondisaster information) in the bag and instance. As shown in Fig. 1, there are many differences among the morphology, spatial distribution, color, and texture in the negative instances, which greatly decrease the accuracy and computational efficiency of the disaster classification model from high-resolution remote sensing image.

The proposed DMCNN model in this article is used for disaster classification from high-resolution remote sensing image based on the MIL framework. Given a training set T, each image $X \in T$ is converted to a set of instances; each instance is represented by a local descriptor x_i, which is mapped to a class label $y_i = h(x_i)$ by a classifier h. In standard hypothesis, the pooling function of MIL framework can be expressed as

$$f(y_i, \ldots, y_n) = \begin{cases} 1, & \text{if } \exists y_i = 1 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $i = 1, 2, \ldots, n$. It means that a negative image subblock includes only one negative instance, whereas the positive image subblock includes more than one positive instance. Since the class label is unknown in the training process, that is, class label y_i is potential variable, then the corresponding bag label is formed via formula (1).

In addition, in the training and testing phase, all instance features in the classified bag are first extracted by a deep CNN, and further took part in the construction of feature representation in the prototype learning layer. And then all the instance features

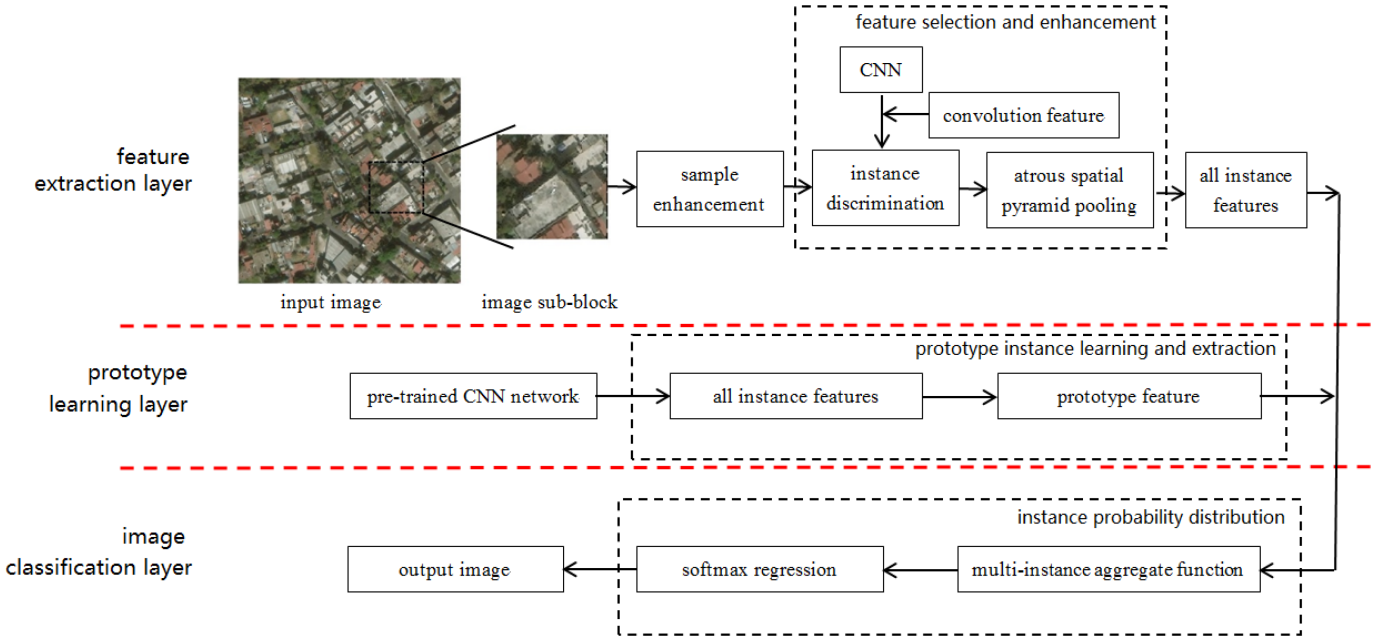

Fig. 2. Proposed DMCNN model framework.

in the prototype vector set obtained earlier are converted to bag-level features through a multi-instance aggregate function. Finally, on the basis of preprocessing and sample enhancement, we perform and achieve the instance feature extraction and disaster classification from the perspective of bag-level in high-resolution remote sensing image.

B. Network Overview and Structure

The proposed DMCNN model in this article can be regarded as a classification framework of high-resolution remote sensing image. As demonstrated in Fig. 2, there are three layers in the DMCNN network framework. The first layer is the feature extraction layer, based on sample enhancement and image preprocessing; we carry out the feature extraction and integration through CNN structure and ASPP to obtain the enhanced instance features in the image. The second layer is the prototype learning layer, which outputs the features of a series of prototype instances through the instance features extracted by a pretrained CNN network. The third layer is the image classification layer. All the instance features and prototype instance features in the image are input into the multi-instance aggregate function to obtain the instance probability distribution of different bags and then are classified by the softmax regression.

1) ASPP Definition: Compared with the standard convolution, ASPP enlarges the receptive field of the convolution kernel by increasing the expansion rate. Therefore, ASPP can be considered as a standard convolution kernel with a larger size, to some extent, and its definition of receptive field can be expressed as

$$F = k + (k-1)(r-1) \tag{2}$$

where F is the receptive field of ASPP, k is the size of the convolution kernel, and r is the expansion rate.

As shown in Fig. 3, by deleting the expanded convolution with $r = 24$, it can overcome the limitation that ASPP cannot extract the effective features from complex high-resolution remote sensing image because of the excessive expansion rate. Besides, image features are parallel extracted on a fixed feature map from the four atrous convolution with multiscale expansion rates in the process of downsampling [45]. Therefore, the advantage of ASPP for the disaster detection and classification from high-resolution remote sensing image is very obvious. It can not only keep the number of trainable parameters in each layer of network unchanged, greatly reducing the amount of data computation, but also retain the resolution of feature map, and extract more and more detailed information, especially in small local information.

The model of ASPP can be formulated as

$$y[i] = \sum_{k=1}^{K} x[i + r \cdot k]w[k] \tag{3}$$

where $y[i]$ is the output signal of the ith output feature map, x is the input signal, w is the convolution filter, r is the expansion rate of filter, and $r = 1$, 6, 12, and 18, respectively.

2) Feature Extraction: Feature extraction is very important for instance learning and feature selection in disaster classification from high-resolution remote sensing image. The model structure of feature extraction network is shown in Fig. 4.

1) The input image includes high-resolution remote sensing image (bag) and image subblock (instance).
2) To better extract disaster feature from high-resolution remote sensing image, the image preprocessing and sample enhancement are first used to increase the recognizability of disaster ground object. Specific enhancement methods include random up-down (left-right) flip and image

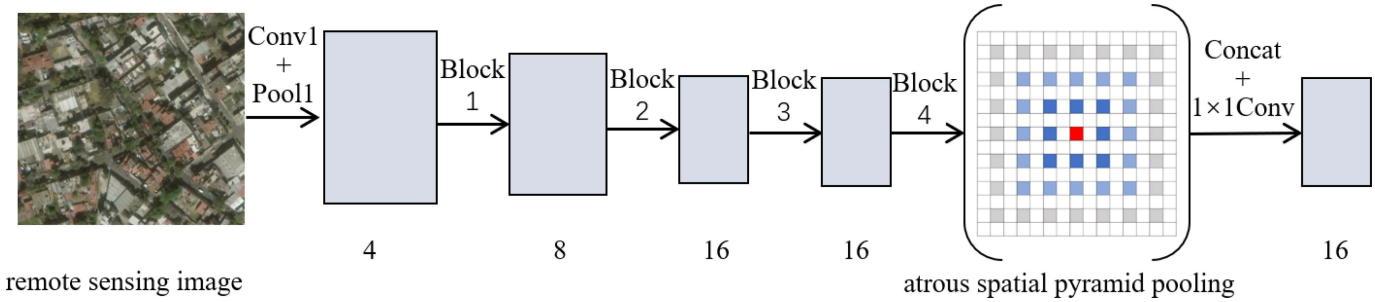

Fig. 3. ASPP structure. The red line, dark blue line, light blue line, and gray line represent the rate = 1, 6, 12, and 18, respectively.

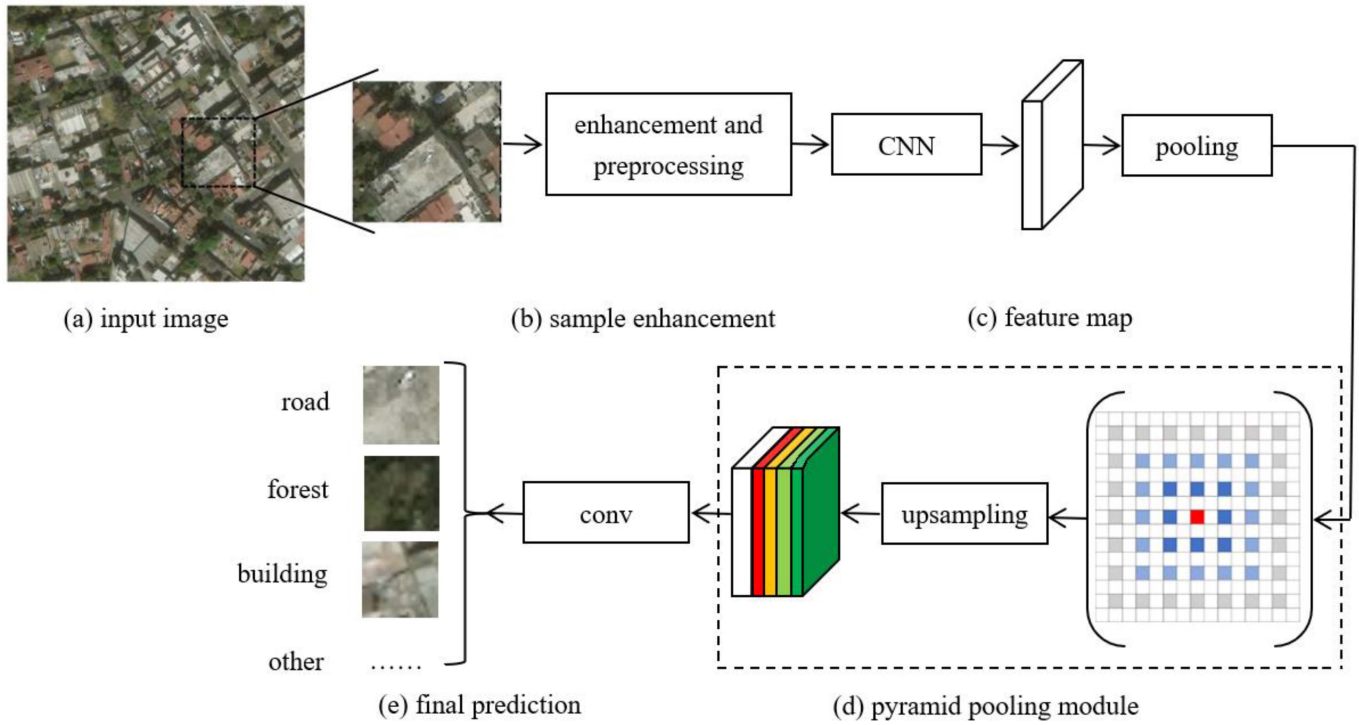

Fig. 4. Model structure of feature extraction network. (a) Input image. (b) Sample enhancement. (c) Feature map. (d) Pyramid pooling module. (e) Final prediction.

brightness adjustment by random factor, and the image preprocessing methods include image fusion, mosaic, cutting, etc. It significant increases the uniform distribution of samples in the entire image and decreases the image blur.

3) Multiscale feature maps are obtained by pooling operation with multiscale expansion rates, and then channel attention mechanism (CAM) in deep neural network is used to reduce the number of channels. Specifically, the feature map is obtained from ResNet through pretraining of the ImageNet dataset.
4) Adaptive pooling is first carried out in the ASPP module, and feature maps of corresponding sizes are obtained by four different pooling sizes. Meanwhile, we apply a 1×1 convolution operation on the feature map to reduce the number of channels. Then, we implement the upsample operation by using bilinear interpolation method to obtain the same size feature map. Finally, all generated maps are spliced on different channels, respectively.
5) The prediction results of disaster instance feature in high-resolution remote sensing image are obtained by suitable convolutions.

3) Prototype Learning: In the prototype learning layer of disaster classification from high-resolution remote sensing images, each prototype vector can be regarded as the distribution center of a certain category. The specific instances belonging to the same prototype vector have the minimum distance from the prototype vector, whereas the other categories' instance has a relatively large distance from the prototype vector. In other words, there are both high intracluster similarity and low intercluster similarity among different types of ground objects. Consequently, different bags eventually exhibit their own unique characteristics from different dimensions. In essence, we construct the feature descriptor with the distance measurement to

obtain the representation relation between the instance and the corresponding prototype vector.

To get prototype vectors from the dataset, let us suppose the following assumptions.

1) $H = [h_1, h_2, \ldots h_n] \in R^{d_h \times n}$ is the characteristic matrix corresponding to all instance features extracted by CNN network, d_h is the dimension of the instance feature outputted from CNN network, *n* and *p* are the number of instances and prototype vectors, respectively.
2) Prototype vectors represent the center points of spatial distribution of different types of ground objects, which can describe the relationship among all instances in the whole space.
3) The dataset contains *k* categories of ground objects, and theoretically, there are *k* central points that can be found in the space. For the distance among clusters to be as large as possible and the distance within clusters is as small as possible, we select the *p*/*k* samples closest to each center as prototype vectors.

4) Regression Classification: The CNN used in this article is a pretrained network on the ImageNet dataset; in specific, all convolution layers are activated by random weights and ReLU functions. First, the prototype instances are selected from the all instances obtained by transfer learning, and the minimum distance between the instances in the bag and the prototype instances is considered as the bag's characteristics. Then, a bag-level classifier is constructed and the characteristics of the bag are directly classified. Second, the CAM in the multi-instance aggregation function is used to fuse the instance information between the feature instances and prototype instances. Meanwhile, the attention is centralized at the key part of the disaster classification in terms of the weight of the instance features. Finally, the probability distribution of different types of disaster information in the CAM network can be defined as

$$y_c = \frac{\sum_{i,j} w_{i,j} y_{i,j,c}}{\sum_i w_{i,j}} \tag{4}$$

where $w_{i,j}$ is the attention weight value of the instance, which changes with the importance of instance features in the disaster classification from high-resolution remote sensing image.

C. Loss Function and Optimization

Given a multi-instance dataset $x = [x_1, x_2, \ldots, x_n]$, the *i*th bag is $x_i = [x_{i1}, x_{i2}, \ldots, x_{im}]$, $x_{i,j}$ is the *j*th instance in the bag x_i, y_i and $y_{i,j}$ are the bag label and hidden label of x_i, respectively. To better train the classifier network, we first normalize the parameters in the feature map by softmax function. Then, we convert the score of each category in the bag into conditional probabilities $p(c|\theta)$

$$p(c|\theta) = \frac{e^{y_c}}{\sum_k e^{y_k}} \tag{5}$$

where y_c is the output value of the *c*th parameter in the feature map obtained after upsampling, *k* is the number of the output parameters (i.e., the number of classified categories), and θ is the parameter of network, which can be obtained by the maximum-likelihood estimation.

The cross entropy loss function between the bag's category probability and its true label can be defined as

$$L(\theta) = -\sum_c y_c \log(p(c|\theta)) \tag{6}$$

where $y_c \in \{0, 1\}$ is the true label of the bag's category. In the training process, we input the feature vectors of the training instances into the network in the unit of bags to obtain the probability that each instance belongs to $\{0, 1\}$, where the maximum positive category probability is that the bag belongs to the positive category, and the loss is minimized by stochastic gradient descent method.

In view of the uneven distribution of positive and negative categories commonly existing in disaster classification from high-resolution remote sensing image, in this experiment, we adjust the distribution of positive and negative category bags by setting weights to find out the sensitivity of positive and negative category bags to classifiers. Let us assume that n_i is the number of negative samples, p_i is the number of positive samples, and w is the weight value. Then, the weight of positive and negative category bags can be defined as

$$w = \begin{cases} \frac{p_i}{n_i}, & y = 0 \\ 1, & y = 1 \end{cases}. \tag{7}$$

Finally, we use the Adam optimizer to process the proposed novel DMCNN model in this article. Specifically, the learning rate is 0.0001, and the model training is updated iteratively until convergence is achieved.

IV. Experiments

A. Datasets

The xBD dataset is an annotated public high-resolution remote sensing dataset and focuses on building damage assessment. It contains a total of 22 068 images with a spatial resolution of 1024 × 1024 and 19 types of natural disasters (earthquake, flood, wildfire, volcanic eruption, etc.). Specifically, the high-resolution remote sensing images in the dataset are collected from the WorldView-3 satellite with a 0.3 m resolution and cover a total of 45 361.79 km^2 [46].

The xBD dataset contains a train training set, test set, holdout set, and Tier3, respectively. Owing to the differences in satellite sensor imaging angles, lighting conditions, and disaster types, there are random irregular distribution of shape features of ground objects, large spatial complexity, and intraclass differences in the dataset. As mentioned above, the xBD dataset is very suitable for deep neural network learning and training for disaster scene with large scale.

By summarizing common natural disaster phenomena, we ultimately selected the six most common types of natural disasters from the xBD dataset, such as earthquake (building collapse), flood (flood and hydrological monitoring), volcanic eruption (debris flow), forest farm fire (forest and grassland wildfires), hurricane, and pinery-bush fire. The partial scheme of common hazard types in xBD dataset is shown in Fig. 5.

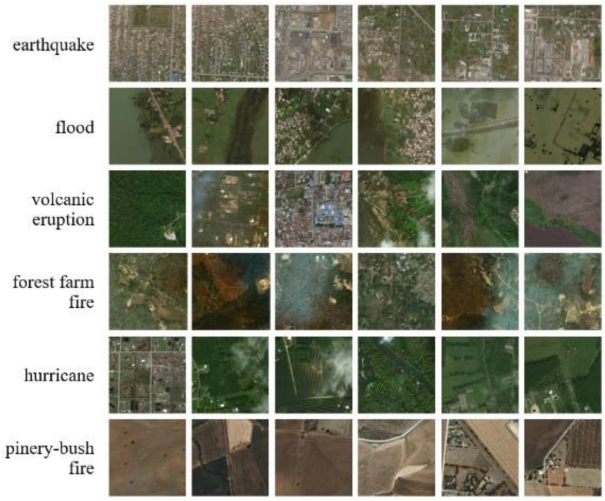

Fig. 5. Partial scheme of common hazard types in xBD dataset.

To efficiently allocate resources and coordinate various forces for rapid assistance, it is necessary to acquire the precise disaster location and loss assessment in postdisaster reconstruction. At present, the disaster relief efforts are still mainly done by hand, and the fully automatic detection and rescue for disaster scene is only a supplementary means. Correspondingly, xBD dataset contains a large number of images, labels, and annotation information including annotation polygons, annotation proportion, etc. Theoretically, the xBD dataset can be used to quickly acquire the geographic location of the disaster body and assess its damage, and provide important experimental data support and method verification for realistic postdisaster relief and reconstruction work.

B. Sample Enhancement and Preprocessing

In the actual disaster classification from high-resolution remote sensing image, as a special case that one certain type of data in the sample is far more than other types of data, it results in inaccurate discrimination for a few categories of ground objects. To some extent, some small samples in high-resolution remote sensing images often contain critical information for the disaster detection (i.e., disaster body and surrounding background), so it is necessary to first implement the sample enhancement and preprocess to overcome the sample imbalance in the disaster classification from high-resolution remote sensing image.

In the experiment, the detailed combination strategies include the following few aspects.

1) Flipping random with up-down and left-right.
2) Generating some subimages with a certain size by cutting random from original image.
3) Generating images in random sequence.
4) Adjusting the image brightness by random factors.

Through the above combination ways of sample enhancement and preprocessing, it can effectively enhance the quantity and diversity of sample data and generalization of the constructed DMCNN model, and further avoid the overfitting problems.

For the disaster classification from xBD dataset, we take the whole image as a bag and an image subblock as an instance. In specific, BING algorithm [47] is first used to generate a series of target boxes on the image, and then a certain number of target boxes are filtered and reserved on each image. Finally, the subimage in the reserved target box is zoomed and used as an instance of the disaster image.

C. Experimental Settings

In the experiment, the xBD dataset is used to verify the performance of the proposed DMCNN model, the training ratios are set to 80% and 20%, and the left for testing. To increase the stability of the proposed DMCNN model and obtain reliable results, we select 10 image instances in bag, then repeat each experiment 10 times, and then calculate the average and standard deviations as the measurement data. Since there are many types of ground objects and complex distribution in xBD dataset, we pretrain the ResNet50 network with ImageNet dataset to test the DMCNN model performance. Then, we use Adam optimizer with a batch size of 16 to adjust and optimize the network model. For ResNet50 models, the initial learning rate is set to 0.0001, and each model was trained for 20 rounds.

In addition to the dataset and network model, the specific experimental environment is as follows: compilation environment is Python 3.7, deep learning framework is Tensorflow2.0, operating system is Linux Ubuntu 20.01.4 LTS and Cuda-toolkit 8.0, hardware platform is Intel Xeon E5-2620 v4 CPU @ 3.80 GHz, Nvidia Quadro M4000 GPU graphics card (8 GB memory).

D. Evaluation Index

In actual, each evaluation index has its own concerns; the synthesis application of multiple indexes to evaluate can reflect the comprehensive performance. To evaluate the proposed DMCNN model from the aspect of computing performance and classification accuracy, in this experiment, we finally ascertain and select the following indexes including precision (P), Recall (R), F1 Score ($F1$), overall accuracy (OA), Kappa coefficient, and standard deviation.

P and R can be defined as

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

where TP is the number of true positive, FN is the number of false negative, and FP is the number of false positive that true category misjudged to other categories.

However, there is a contradiction between R and P. To obtain a higher P, it is necessary to reduce the threshold of the network model, which will miss some positive information and lead to the decline of R, and vice versa. $F1$ is used to balance the relationship between P and R, and it can be expressed as

$$F1 = \frac{2 \times P \times R}{P + R}. \tag{10}$$

TABLE I
CORRESPONDING RELATIONSHIP BETWEEN THE KAPPA COEFFICIENT AND CONSISTENCY

Grade	1	2	3	4	5
Kappa coefficient	0.0–0.20	0.21–0.40	0.41–0.60	0.61–0.80	0.81–1
Consistency	slight	fair	moderate	substantial	almost perfect

For the given multi-instance sample $x = [x_1, x_2, \ldots, x_n]$, OA can be defined as

$$OA = \frac{1}{N}\sum_{i=1}^{g} x_{ii} \tag{11}$$

where N is the total number of pixels of all true references, $x_{i,j}$ is the diagonal of the confusion matrix, and g is different types of ground objects.

Kappa coefficient can be defined as

$$Kappa = \frac{N\sum_{i=1}^{g} x_{ii} - \sum_{i=1}^{g}(x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^{g}(x_{i+} \times x_{+i})} \tag{12}$$

where $i+$ is matrix row element and $+i$ is matrix column element.

In general, the value range of the Kappa coefficient for the DMCNN model is from –1 to 1. To obtain the measuring accuracy of image classification, we compute the Kappa coefficient and consistency, respectively. Then, we can classify the corresponding relationship between the Kappa coefficient distribution and consistency into five grades including slight, fair, moderate, substantial, and almost perfect. Table I reports the corresponding relationship between the Kappa coefficient distribution and consistency.

Standard deviation, also known as quadratic mean deviation, can be defined as

$$S_x = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}}. \tag{13}$$

Standard deviation can accurately represent the discrete degree of a dataset and is often used to evaluate the model stability in the specific application.

V. RESULTS AND ANALYSIS

A. Qualitative Evaluation and Analysis

1) Feature Map Visualization: For a given input image, feature map visualization refers to the feature map output by each intermediate layer in the deep network model, that is, the output result of activation function. Correspondingly, for the heat map in the feature map visualization, the darker the color, the greater the contribution of the area's features to the image classification. In general, the feature map visualization consists of three parts: width, height, and depth (channel), and each channel corresponds to relatively independent features (two-dimensional image).

Feature map visualization can clearly show the decomposition process of the input image in the deep learning network model. It focuses more on the specific geographic location of the input image and highlights the importance of some key information for image classification. Fig. 6 illustrates the partial scheme of feature map visualization by the proposed DMCNN model in this article.

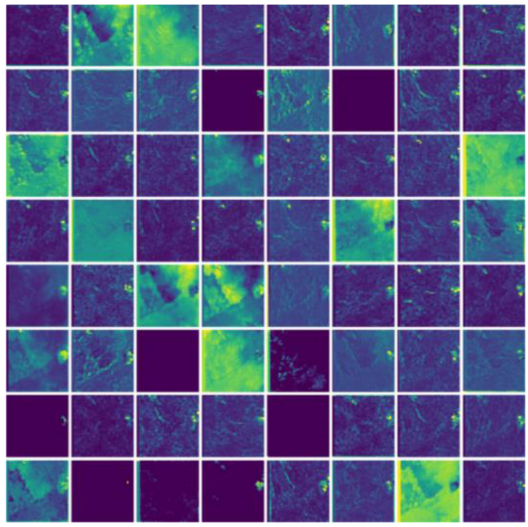

Fig. 6. Partial scheme of feature map visualization by the proposed DMCNN model.

As shown in Fig. 6, the heat map represents the different areas in brightness; the brighter the hue, the higher the activity and visualization. In the feature map, the obtained boundary of the ground object is relatively clear, whereas the distribution location and contour of the ground object are bright colors and its contrast with the surrounding object is obvious. It enlarges the discriminability between disaster area and noise background in the high-resolution remote sensing image and improves the classification accuracy of natural disaster.

In view of the consistency of classification results in the heat map with the distribution of real ground objects, in the experiment, taking the fire disaster as an example, we implement the test for the proposed DMCNN model from the following three indexes: TP, FN, and FP. Fig. 7 demonstrates the partial scheme of fire disaster samples in heat map obtained by the proposed DMCNN model.

1) As shown in Fig. 7(a) and (b), in the original input images, the box area with black line is the real fire disaster area, and the color and texture are more consistent with the surrounding ground objects. There are fewer interference factors and the real disaster area is identified accurately by the proposed DMCNN model in this article, and shows the red distribution area of TP fire disaster samples in the heat map.
2) As shown in Fig. 7(c) and (d), in the original input images, the box area with black line is buildings and some bare soil, which represents the FP samples in fire disaster classification. The buildings have regular shape and texture

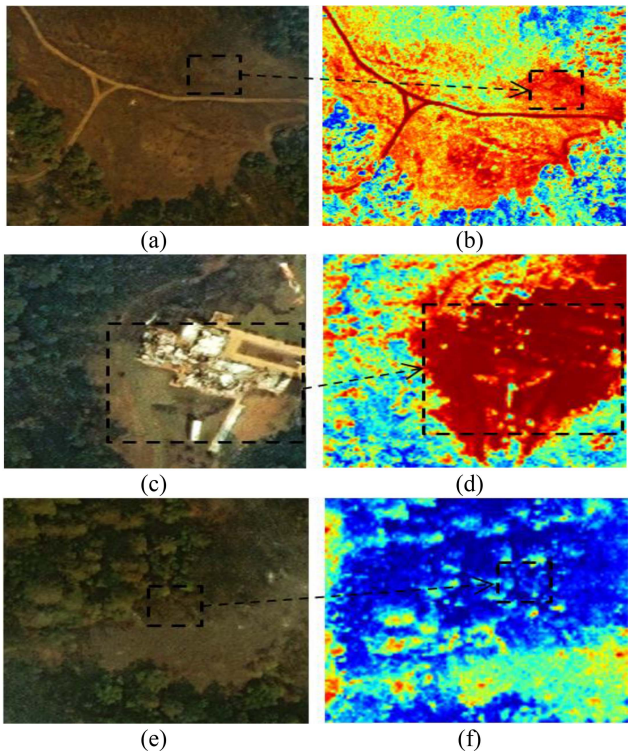

Fig. 7. Partial scheme of fire disaster samples in heat map obtained by the proposed DMCNN model. (a), (c), and (e) Input original images. (b), (d), and (f) Visualization maps of the corresponding input images. (a), (b) TP samples, some true fire objects are recognized as the fire disaster area by the DMCNN model. (c), (d) FP samples, the building objects are not the fire disaster, but the DMCNN model still recognized it as the fire disaster area. (e), (f) FN samples, some objects are the fire disaster, but the DMCNN model does not recognize it as the fire disaster area.

features, and the bare soil is similar to the true fire disaster in the color and texture features. However, the buildings and bare soil are not identified by the proposed DMCNN model, and these samples are wrongly classified as fire disaster area and show the red distribution area of TP fire disaster samples in the heat map.

3) As shown in Fig. 7(e) and (f), in the original input images, the box area with black line is true fire disaster area, which represents the FN samples in fire disaster classification. However, the proposed DMCNN model in this article cannot classify the distribution area of fire disaster, and shows the blue distribution area of FN fire disaster samples in the heat map.

As can be seen in Fig. 7, although the proposed DMCNN model achieves good fire classification effect, especially when the spatial geometry structure and color of surrounding ground objects are more uniform, the classification effect is not ideal for some objects with regular spatial geometry shape. Moreover, there are false positive samples and false negative and other misjudgments because of the complex geographical background.

2) Classification Visual Effect: At present, the disaster classification from high-resolution remote sensing image is mainly completed through human-computer interaction. In the actual operation, these traditional methods demand user to have some programming ability and previous experience, and the interaction form is relatively complex. Since color difference and distribution of surrounding ground objects in high-resolution remote sensing image are crucial to disaster classification, to ensure the classification accuracy and model generalization, in the experiment, we take earthquake disaster as an example to carry out qualitative evaluation from the aspects of color, surrounding background, geometry shape of ground objects.

Fig. 8 illustrates the classification results of the single type of earthquake disaster in high-resolution remote sensing image. The red line frame is the hand-drawn result by visual interpretation, and the purple and yellow areas are the classification results of earthquake disasters with different disaster degrees obtained by the proposed DMCNN model in this article. In the first column, the color distribution of earthquake disaster areas and surrounding ground objects is uniform, and the contours of the hand-drawn disaster area are basically consistent with those of the true earthquake disaster, as shown in Fig. 8(a). In the second and third columns, it clearly shows the classification results of earthquake disaster with complex color distribution and large contrast area, and complex distribution of surrounding ground objects, as shown in Fig. 8(b) and (c). In summary, there are complex surrounding backgrounds distributed around the earthquake disaster area, for example, buildings, shadows, vegetation, bare soil, etc. On the one hand, the changes of color and texture in complex surrounding environment increase the difficulty to accurately classify earthquake disaster from high-resolution remote sensing image. On the other hand, the difference of damage degree for buildings and other physical facilities caused by earthquake disaster contains the following: complete damage, severe damage, and minor damage. It raises the difficulty of detection and classification of earthquake disaster from high-resolution remote sensing image, and will be discussed in detail later.

In addition to the single earthquake disaster, we use the proposed DMCNN model to classify the multidisasters (e.g., flood, pinery-bushfire, volcano eruption, fire, and hurricane) in terms of actual distribution of different types of disasters. The classification results of different types of disasters are shown in Fig. 9.

As demonstrated in Fig. 9, for different types of disaster areas with various shape distributions, although the disaster area boundaries drawn by visual interpretation are hand-drawn and have diverse irregular shapes, the hand-drawn boundaries are basically consistent with the classification results of the proposed DMCNN model. To some extent, this also once again verifies the accuracy and effectiveness of the DMCNN model in this article.

B. Quantitative Evaluation and Analysis

1) Comparison of the Prototype Number: To obtain the appropriate number of prototypes in the DMCNN model, we first use the verification method in the literature [19] and select several representative prototype numbers to test the relationship between the prototype numbers and classification performance on the xBD dataset, as shown in Table II and Fig. 10.

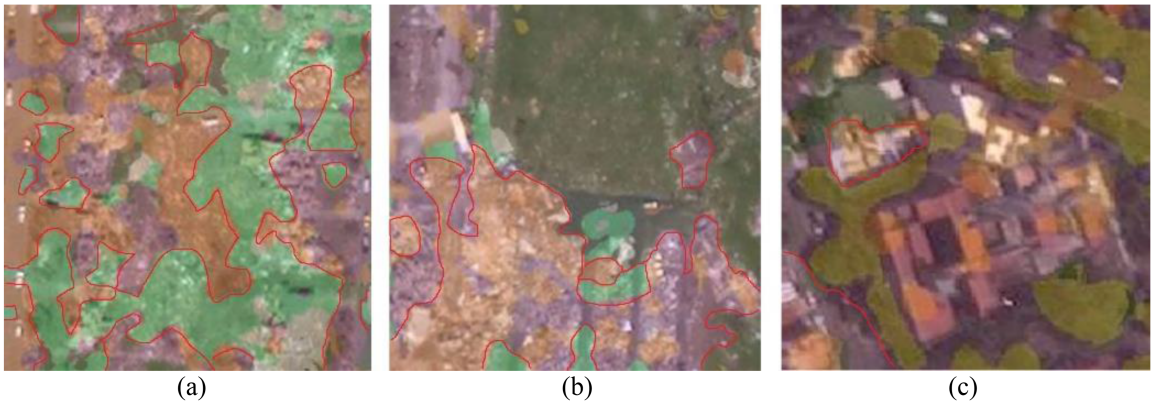

Fig. 8. Classification results of the earthquake disaster with different surrounding background of ground objects. (a) Uniform color area. (b) Complex color distribution and large contrast area. (c) Complex distribution of surrounding ground objects. The red line frame is the hand-drawn result by visual interpretation, and the purple and yellow areas are the classification results of earthquake disasters obtained by the proposed DMCNN model.

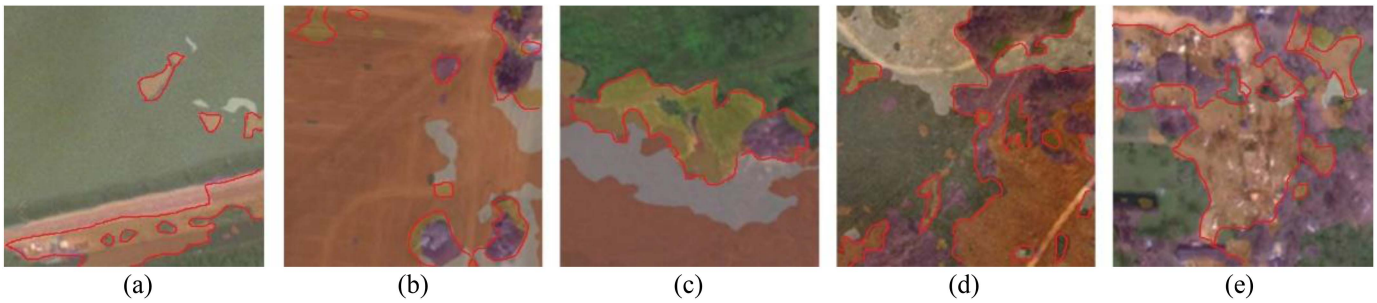

Fig. 9. Classification results of different types of disasters. (a) Flood. (b) Pinery-bushfire. (c) Fire. (d) Volcano. (e) Hurricane. The red line frame is the hand-drawn result by visual interpretation, and the purple and yellow areas are the classification results of different types of disasters obtained by the proposed DMCNN model.

TABLE II
RELATIONSHIP BETWEEN THE PROTOTYPE NUMBERS AND CLASSIFICATION PERFORMANCE

Number	OA	P	R	F1
40	0.9800	0.9800	0.9800	0.9800
80	0.9900	0.9910	0.9900	0.9900
120	0.9850	0.9850	0.9850	0.9900
160	0.9800	0.9800	0.9800	0.9800
200	0.9750	0.9760	0.9750	0.9750

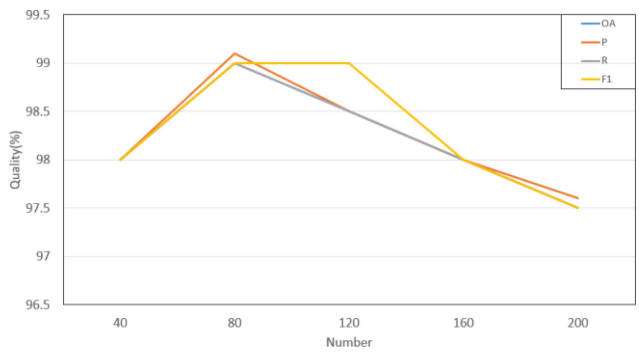

Fig. 10. Relationship between the prototype numbers and classification performance.

Table II reports the results of the relationship between the prototype numbers and classification performance. We also deduce that the variation of the number of prototypes has an obvious influence on the disaster classification from xBD dataset. With the increase of the number of prototypes, the performance of disaster classification gradually increased. When the number of prototypes reached 80, the performance of image classification achieves the highest at $OA = 0.9900$ and $P = 0.9910$ as well as at $R = 0.9900$ and $F1 = 0.9900$, respectively. This means that the proposed DMCNN model achieves the optimal disaster classification effect from high-resolution remote sensing image at this point of prototype numbers. Subsequently, with the continuous increase of the number of prototypes, the performance of disaster image classification begins to gradually and rapidly decline. Therefore, we select the number of prototypes to be 80 to implement the subsequent DMCNN model testing and verification to ensure the optimal performance of the network. This also proves the necessity and rationality of introducing a prototype learning layer into the proposed MDCNN model structure from another point of view.

2) Comparison of Disaster Classification With Single Method: To evaluate the performance of the single DMCNN model on the disaster classification from xBD dataset, in the experiment, we use the multi-index (e.g., confusion matrix, P, R,

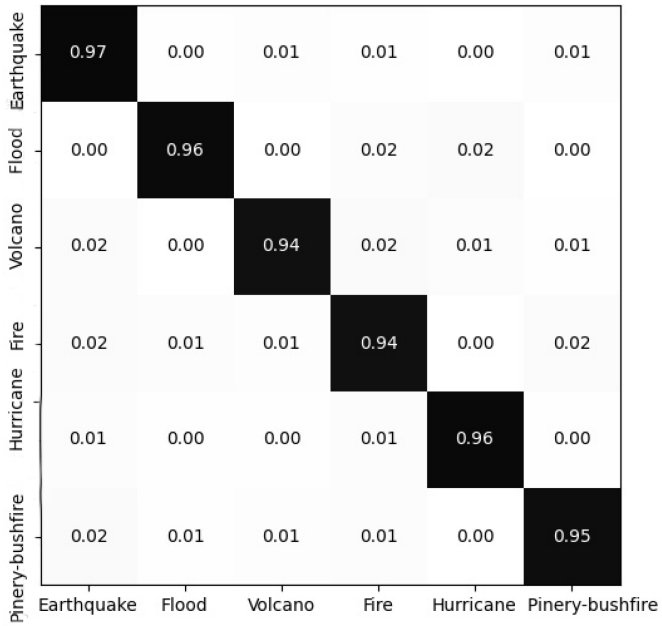

Fig. 11. Confusion matrix of DMCNN model on xBD dataset.

and $F1$) to compare and analyze the classification results on xBD dataset. For a better comparison, Figs. 11 and 12 clearly show the confusion matrix and comparison of classification performance of the proposed DMCNN model on xBD dataset, respectively.

As shown in Fig. 11, despite the appearance of the misclassification of disaster area in flood, volcano, and fire disaster due to the complex surrounding environment, the proposed DMCNN model can well recognize and classify the above six disaster types from xBD dataset. In terms of the results of the above analysis, in specific, the color and texture of different disaster areas in some images are similar and indistinguishable. And what is more, there are numerous types of ground objects around the disaster areas and complex distribution in practice. All these factors increase the difficulty and complexity of disaster classification from xBD dataset.

Fig. 12 illustrates the detailed classification performance comparison of a single DMCNN model on the xBD dataset. For index P, the classification effect of earthquake and hurricane obtained by the proposed DMCNN model reaches 0.97 and 0.96, respectively, followed by pinery-bushfire and volcano with 0.95 and 0.94, respectively, and the flood and fire disasters only reach 0.92 and 0.91. This is also consistent with the changing trend of index $F1$. As given in index P, the index R of earthquake and pinery-bushfire reaches 0.97 and 0.96, respectively, followed by volcano and hurricane with both 0.95, and the flood and fire disasters are both 0.92. From the above analysis of the P, R, and $F1$ indexes, the best disaster classification effect is achieved for earthquakes, hurricanes, and pinery-bushfire, whereas the effect for volcanoes, floods, and fire is poor. It is also consistent with the confusion matrix in Fig. 11.

3) Comparison of Disaster Classification With Multiple Methods: To compare the disaster classification effect of multiple methods from xBD dataset, in this experiment, we use five deep learning methods: SVM, CNN, MIL, MI-CNN, and CNN-SVM, to test and evaluate the model performance. Table III reports the comparison results of disaster classification by the multiple methods on xBD dataset.

In terms of the results of classification comparison in Table III, the OA and P of the traditional SVM method based on manual feature extraction reach 0.8896 and 0.8889, respectively, and the classification effect on high-resolution remote sensing image is significantly lower than that of the other five deep learning methods. Meanwhile, on the whole, the classification effect of the above deep learning methods all reach more than 0.9, which shows that the deep learning method with depth feature extraction has great advantages for disaster detection and classification from complex high-resolution remote sensing images. However, there are some differences on xBD dataset shown in the following.

1) In the above five types of deep learning methods, the multi-instance deep learning methods (e.g., MIL, MI-CNN, and the proposed DMCNN) have higher indexes OA and P on xBD dataset. Correspondingly, the indexes R and $F1$ of the multi-instance methods reach 0.9536 and 0.9516, 0.9459 and 0.9534, 0.9521 and 0.9543, respectively. In summary, compared to the SVM method and other deep learning methods, the proposed DMCNN model in this article has the best classification performance and achieves the optimal effect in the four evaluation indexes.
2) Although the proposed DMCNN model has a better classification effect compared with the traditional CNN + classification method, it has no significant advantage in the performance of disaster classification. According to the analysis, the possible reasons caused by are as follows: the classification effect of deep learning method on a single dataset tends to be saturated, some commonalities at the bottom features between natural images and high-resolution remote sensing images. This also verifies that the method of sharing natural image datasets by training CNN model is still effective and feasible.

Table IV reports the $F1$ comparison results of different methods on xBD dataset.

As demonstrated in Table IV, in summary, the proposed method in this article achieves better results than the first two methods in Table IV, and achieves the same effect compared to the third method. The first two methods in Table IV are implemented combining specific detection with recognition, and the processing method is more flexible and faster. The third method and our proposed method both adopted BING algorithm for ground object detection. There was a large amount of noise on the retained image blocks, and the test is implemented in a weakly supervised environment. In summary, this verifies the effectiveness of our proposed method in resisting the interference of instance noise.

In addition to the comparison with indexes OA, P, R, and $F1$, we apply six indexes: number of layers, OA, Kappa coefficient, standard deviation, parameter size, and training time, to test and evaluate the overall performances for the multiple deep learning frameworks. We take the volcano disaster scene in xBD dataset as example, the classification results by multiple learning models

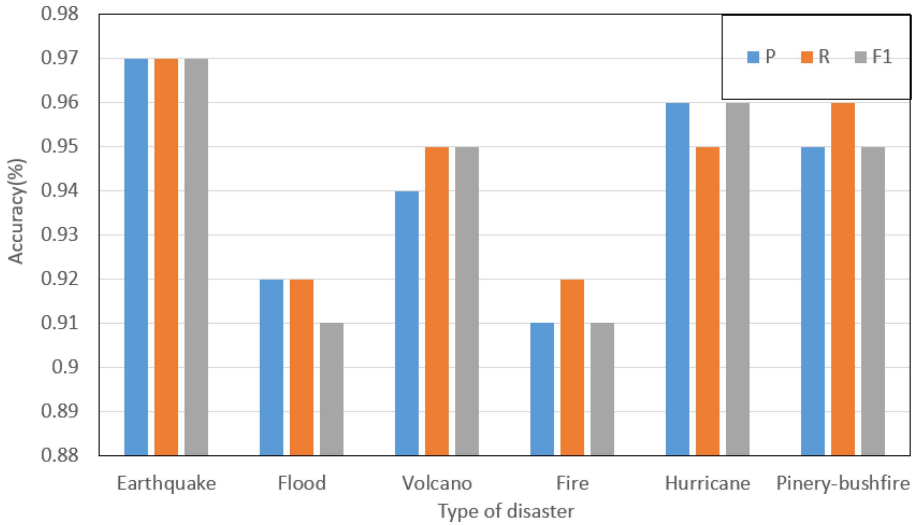

Fig. 12. Classification performance comparison of DMCNN model on xBD dataset.

TABLE III
COMPARISON OF DISASTER CLASSIFICATION BY THE MULTIPLE METHODS ON XBD DATASET

	OA	P	R	F1
SVM	0.8889	0.8896	0.8889	0.8888
CNN	0.9156	0.9171	0.9156	0.9154
MIL	0.9536	0.9537	0.9536	0.9516
MI-CNN	0.9553	0.9556	0.9459	0.9534
CNN-SVM	0.9200	0.9204	0.9200	0.9198
The proposed method	0.9561	0.9561	0.9521	0.9543

TABLE IV
COMPARISON RESULT OF $F1$ WITH DIFFERENT METHODS ON XBD DATASET

	Earthquake	Flood	Volcano	Fire	Hurricane	Pinery-bushfire
Wang and Hu [9]	0.9481	0.8976	0.9186	0.9165	0.8993	0.9165
Li *et al*. [44]	0.9475	0.9234	0.9206	0.9415	0.9073	0.9427
He *et al*. [19]	0.9562	0.9124	0.9422	0.9588	0.9362	0.9582
The proposed method	0.9342	0.9227	0.9625	0.9630	0.8894	0.9487

TABLE V
COMPARISON OF CLASSIFICATION PERFORMANCE BY MULTIPLE DEEP LEARNING MODELS

	Number of layers	OA	Kappa coefficient	Standard deviation	Parameter size (M)	Training time (s)
CNN	8	0.9071	0.8968	0.2903	11.17	13.87
MIL	18	0.9156	0.9062	0.2780	11.17	16.15
MI-CNN	18	0.9200	0.9111	0.2713	11.25	20.74
CNN-SVM	18	0.9536	0.9484	0.2103	11.68	34.55
The proposed method	50	0.9561	0.9512	0.2049	11.69	46.23

are shown in Fig. 13, and some details of the comparison of classification performance by multiple deep learning models have been summarized in Table V.

As shown in Fig. 13, the above five deep learning methods mentioned in Table V almost can accurately classify disasters from high-resolution remote sensing images, and the overall visual effect of the classified images is good. In these deep-learning-based methods, the CNN and MIL are considered as a benchmark [see Fig. 13(a) and (b)]. Compared to the former two methods, the classification performance of MI-CNN and CNN-SVM methods has been improved, and the visual effect of corresponding classified images is well [see Fig. 13(c) and (d)]. Meanwhile, the proposed DMCNN model in this article has the best classified image quality [see Fig. 13(e)]. However, there are many fragmented patches in the above-classified images. According to analysis, the Worldview image with 0.3 m

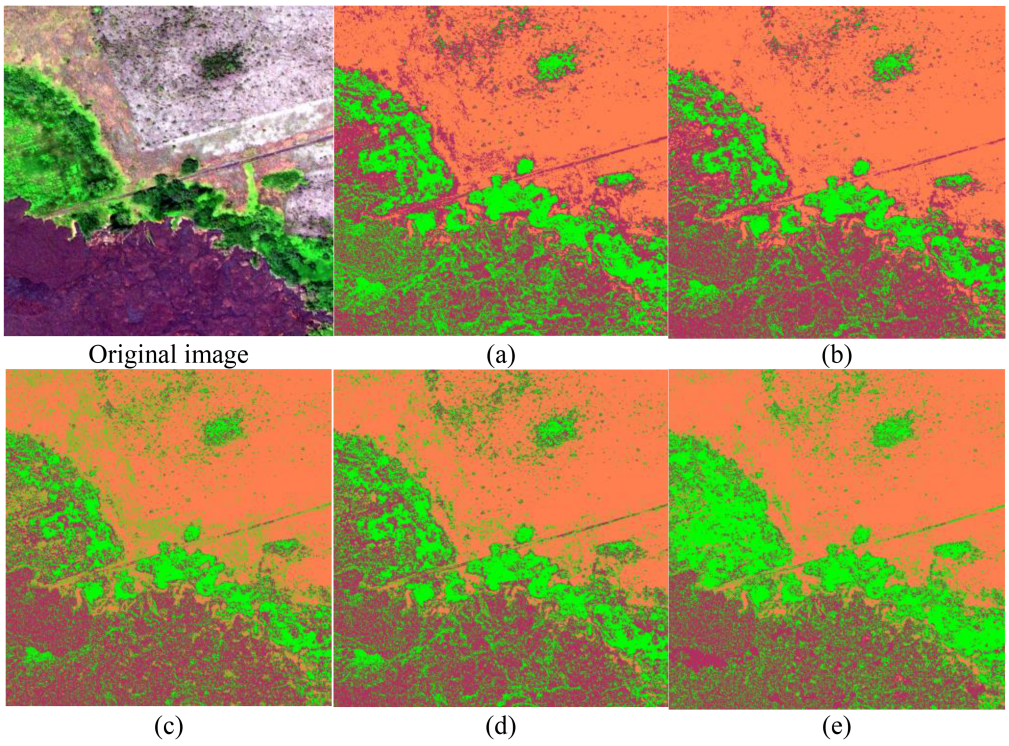

Fig. 13. Classification results of volcano disaster with multiple deep learning models. (a) CNN. (b) MIL. (c) MI-CNN. (d) CNN-SVM. (e) Proposed method.

spatial resolution in the xBD dataset is a very high-resolution remote sensing image; the size, shape, and distribution of ground objects are smaller and complicated compared to traditional remote sensing images. It remarkably increases the difficulty at a geometric rate to accurately detect and classify the disaster ground object from high-resolution remote sensing image.

As demonstrated in Table V, we observe that the following holds.

1) From the vertical, the classification accuracy of several deep learning models increases successively, whereas the corresponding number of network layers deepens from 8 layers to 18 layers and then to 50 layers. Theoretically, deeper network layers can extract more complex and abstract features and has a better classification effect from remote sensing images.
2) For the index OA, the performance of MIL, MI-CNN, and DMCNN models is better than the benchmark CNN and MIL methods. In these three combined deep network models, the index OA of the proposed DMCNN model in this article reaches 0.9561, which achieves 3.61% and 0.25% improvements over the MI-CNN and CNN-SVM, respectively. For the index Kappa coefficient, the overall trends are similar to those on the index OA. The proposed DMCNN model achieves the optimal effect, and its Kappa coefficient reaches 0.9512.
3) For the index standard deviation, the smallest standard deviation is observed on our DMCNN model with 0.2049. That is, the DMCNN model can obtain more stable precision and accuracy of disaster classification from high-resolution remote sensing images.
4) In terms of the parameter size and training time, the parameter size and the corresponding training time will obviously increase along with the network deepens. Compared with the benchmark CNN, MIL, and other combination models, our proposed DMCNN model takes more time to train when the parameter size is not increased significantly.

Through the previous experiments, although the proposed DMCNN model in this article has a large parameter size and a long training time, it can obtain higher precision and accuracy of disaster classification from high-resolution remote sensing image, and has better stability and generalization ability of classification results.

4) Comparison of Disaster Classification With Different Damage Degrees: In this article, the proposed DMCNN model can classify high-resolution remote sensing image accurately and achieves good effect of disaster classification and calculation performance from xBD dataset. To test and evaluate the classification effect of the proposed DMCNN model on a single type of natural disaster, we take the collapsed buildings caused by earthquake disasters as an example and divide collapsed buildings into four damage degrees: undamaged, moderately damaged, severely damaged, and completely damaged. Fig. 14 demonstrates the collapsed buildings and the corresponding classification results caused by earthquake disaster with different damage degree, respectively. Table VI reports the details of the classification accuracy of collapsed buildings caused by earthquake disaster with different damage degrees.

Table VI illustrates the details of the classification performance of collapsed buildings by proposed DMCNN for the different damage degrees. The classification effect of the proposed

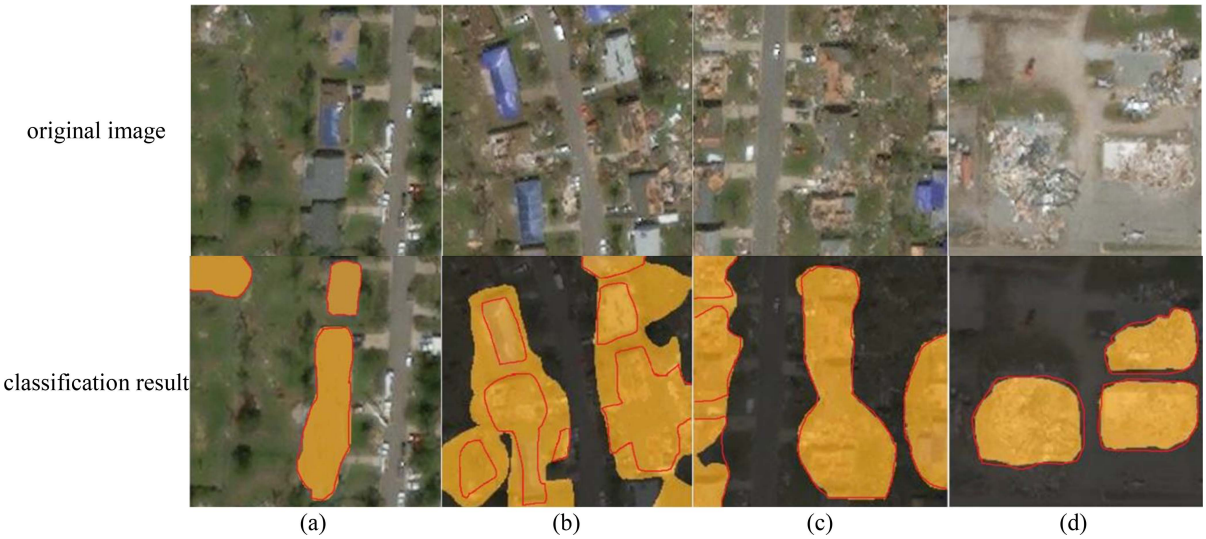

Fig. 14. Partial collapsed buildings (upper row) and classification results (lower row) caused by earthquake disaster with different damage degrees. (a) Undamaged. (b) Moderately damaged. (c) Severely damaged. (d) Completely damaged. The red line frame is the hand-drawn result by visual interpretation.

TABLE VI
CLASSIFICATION ACCURACY OF COLLAPSED BUILDINGS WITH DIFFERENT DAMAGE DEGREES

	Undamaged	Moderately damaged	Severely damaged	Completely damaged
P	0.9630	0.8990	0.9180	0.9710
R	0.9558	0.9026	0.9245	0.9691
F1	0.9488	0.9062	0.9310	0.9671
OA			0.9553	
Kappa coefficient			0.9503	

DMCNN in this article varies with the different damage degrees of collapsed buildings caused by earthquake disaster. In general, the proposed DMCNN can accurately classify undamaged buildings [Fig. 14(a)] and completely damaged buildings [Fig. 14(d)] in earthquake disasters, and the classification precision reaches 0.9630 and 0.9710, respectively. On the contrary, moderately damaged buildings [Fig. 14(b)] and severely damaged buildings [Fig. 14(c)] tend to have lower precision with only 0.8990 and 0.9180, respectively. This is also consistent with the hand-drawn result by visual interpretation for collapsed buildings in earthquake disasters. As shown in Fig. 14, the disaster area boundaries drawn by hand-drawn undamaged area and completely damaged area are more consistent with the classification results, whereas the boundaries drawn by hand-drawn in moderately damaged area and severely damaged areas have significant errors with the classification results. According to the analysis, there are some reasons shown in the following.

1) The spatial distribution structure of undamaged and completely damaged buildings varies greatly in the high-resolution remote sensing image. Meanwhile, the spatial distribution structure of moderately damaged and severely damaged buildings tends to be the same, and the texture is more uniform. For a better understanding and representation of the disaster feature in high-resolution remote sensing image, our proposed DMCNN model in this article focuses on instance features of collapsed buildings caused by earthquake and pays little attention to the overall structure of the whole image.
2) In practice, the extraction feature information of buildings is limited due to their geometric shape, small size, and spatial distribution in high-resolution remote sensing image. Correspondingly, via introducing the ARPP and prototype learning layer into our proposed DMCNN model in this article, we just solve the problem of weak feature extraction in small samples and further improve the classification performance of collapsed buildings in earthquake disasters. The OA and Kappa coefficient of the proposed DMCNN model reach 0.9553 and 0.9503, respectively, which indicates that the proposed DMCNN still achieves a good classification effect of collapsed buildings in earthquake disasters.

VI. DISCUSSIONS

In this article, we analyzed and discussed the experimental results from the following two aspects: qualitative evaluation and quantitative evaluation.

1) For qualitative evaluation and analysis, the feature map visualization and classification visual effect are selected for interpreting the classification effect of the proposed

DMCNN model. When the spatial geometry structure and color of surrounding ground objects are more uniform, the classification effect is not ideal for some objects with regular spatial geometry shape. In addition to the spatial distribution and ground object's own structure, there are some false positive samples and false negative and other misjudgments in the classified image.

2) For quantitative evaluation and analysis, the prototype number, single method, multiple methods, and different damage degrees are also selected for testing and evaluating the proposed DMCNN model from high-resolution remote sensing image. The number of prototypes is crucial to the classification effect of the whole DMCNN model. In the real application scene, how to set a reasonable number of prototypes that match the application scene is still a difficult problem. Compared with the traditional methods based on manual design, although the feature extraction method based on deep learning can largely improve the efficiency of feature extraction and the extents of its application, the model complexity, training difficulty, and calculation amount will be greatly increased. How to achieve the optimal balance between the two needs to be further discussed.
3) For a single type of natural disaster, owing to its varying degrees of damage and multiple influencing factors such as surrounding environment and geological conditions, the proposed DMCNN model in this article only performs better in classifying the two extreme cases with undamaged or completely damaged, whereas it is less effective in moderately damaged and severely damaged cases. However, these cases exactly are the focus that needs to be paid attention to in disaster loss assessment.
4) In our proposed DMCNN model, based on the enlargement of the network receptive field by ASPP module, the features are first extracted and integrated by CNN structure to obtain instance features of bags in image. And then instance features extracted by pretrained CNN are mapped to a series of prototype instance features. Finally, all mapped instance features from bags and prototype instance features are detected and classified in the classifier. Although the proposed model has reached good effects, it is just a preliminary experimental framework and there are still many details that need to be accurately explored in the future.

VII. Conclusion

In this article, to address the deficiencies (e.g., various scenes, complex distribution features, and missing sample class label) of traditional disaster classification from high-resolution remote sensing image, we proposed a novel DMCNN network model for disaster classification on xBD dataset. The key of DMCNN is to construct the prototype representation and receptive field of the convolution kernel through the prototype learning and ARPP and enhance the disaster classification effect via using discriminative instance features with bag-level. Experimental results show that the DMCNN model achieves better disaster classification effect from high-resolution remote sensing image. Compared with traditional CNNs, our DMCNN improves the computational performance of weakly supervised disaster classification from high-resolution remote sensing image.

Although the DMCNN model can promote the accuracy and performance of the disaster classification from high-resolution remote sensing image, there is still some misclassification in the classified image. In the future, on the one hand, in view of the actual situation (e.g., random distribution and small target) of disaster area in high-resolution remote sensing image, we consider designing a more powerful network model with stronger detection and classification capability to overcome these problems. On the other hand, as the important factors in our DMCNN network, the appropriate number of prototypes and prototype vector representation with various measures are crucial to the accuracy and performance of the classification model, we consider refining and verifying the mechanism of reasonable parameter numbers and prototype representation so as to further optimize and improve the DMCNN performance. Furthermore, experiments on other high-resolution remote sensing images and multiple disaster scene classifications are also planned to explore and test the applicability and feasibility of the DMCNN model.

Acknowledgment

The authors would like to thank the editor and reviewers for their many comments and suggestions, which are of great value for improving the quality of this work.

References

[1] H. Wang et al., "Multi-source remote sensing intelligent characterization technique-based disaster regions detection in high-altitude mountain forest areas," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jun. 2022, Art. no. 3512905.
[2] A. S. Louw et al., "The role of remote sensing during a global disaster: COVID-19 pandemic as case study," *Remote Sens. Appl., Soc. Environ.*, vol. 22, 2022, Art. no. 100789.
[3] P. Ye, "Remote sensing approaches for meteorological disaster monitoring: Recent achievements and new challenges," *Int. J. Environ. Res. Public Health*, vol. 19, no. 6, 2022, Art. no. 3701.
[4] L. Zhou et al., "SWDet: Anchor-based object detector for solid waste detection in aerial images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 306–320, Nov. 2022.
[5] X. D. Yao, Q. Guo, and A. Li, "Light-weight cloud detection network for optical remote sensing images with attention-based DeeplabV3+ architecture," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3617.
[6] L. H. Ye, L. Wang, W. W. Zhang, Y. G. Li, and Z. K. Wang, "Deep metric learning method for high resolution sensing image scene classification," *Acta Geodaetica et Cartographica Sinica*, vol. 48, no. 6, pp. 698–707, 2019.
[7] J. C. Luo, T. J. Wu, and L. G. Xia, "The theory and calculation of spatial-spectral cognition of remote sensing," *J. Geo-Inf. Sci.*, vol. 18, no. 5, pp. 578–589, 2016.
[8] D. P. Ming, J. C. Luo, Z. F. Shen, M. Wang, and H. Sheng, "Research on information extraction and target recognition from high resolution remote sensing image," *Sci. Surv. Mapping*, vol. 30, no. 3, pp. 18–20, 2005.
[9] G. M. Wang and N. P. Hu, "Semantic segmentation of natural disaster remote sensing image based on deep learning," *Comput. Syst. Appl.*, vol. 32, no. 2, pp. 322–328, 2023.
[10] Q. Liu, J. F. Xie, J. Z. Chen, R. Y. Sun, and Z. F. Zhao, "Building scene recognition based on convolutional neural network using high resolution remote sensing image," *Bull. Surv. Mapping*, vol. s1, pp. 124–128, 2021.
[11] L. Liu, C. Li, Y. Lei, J. Yin, and J. Zhao, "Volcanic ash cloud detection from MODIS image based on CPIWS method," *Acta Geophys.*, vol. 65, no. 1, pp. 151–163, 2017.

[12] G. Liu, "Monitoring dynamics of Hailuogou glacier and the secondary landslide disasters based on combination of satellite SAR and ground-based SAR," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 44, no. 7, pp. 980–995, 2019.
[13] C. Zhou, H. Wang, X. Chen, and Y. Yang, "Detection of earthquake-damaged buildings via UAV high-resolution remote sensing images," *Nat. Remote Sens. Bull.*, pp. 1–16, 2022, doi: 10.11834/jrs.20221569.
[14] G. Menghani, "Efficient deep learning: A survey on making deep learning models smaller, faster, and better," *ACM Comput. Surv.*, vol. 55, no. 12, 2023, Art. no. 259, doi: 10.1145/3578938.
[15] J. Y. Xuan, J. Lu, Z. Yan, and G. Q. Zhang, "Bayesian deep reinforcement learning via deep kernel learning," *Int. J. Comput. Intell. Syst.*, vol. 12, no. 1, pp. 164–171, 2019.
[16] J. Y. Gong, Y. Xu, X. Y. Hu, L. C. Jiang, and M. Zhang, "Status analysis and research of sample database for intelligent interpretation of remote sensing image," *Acta Geodaetica et Cartographica Sinica*, vol. 50, no. 8, pp. 1013–1022, 2021.
[17] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Perez, "Solving the multiple instance problem with axis-parallel rectangles," *Artif. Intell.*, vol. 89, no. 1/2, pp. 31–71, 1997.
[18] Y. Li, H. K. Zhang, X. Z. Xue, Y. A. Jiang, and Q. Shen, "Deep learning for remote sensing image classification: A survey," *WIREs Data Mining Knowl. Discov.*, vol. 8, no. 6, 2019, Art. no. e1264.
[19] K. L. He, Y. H. Shi, Y. Gao, J. Huo, D. Wang, and Y. Zhang, "A prototype learning based on multi-instance convolutional neural network," *Chin. J. Comput.*, vol. 40, no. 6, pp. 1265–1274, 2017.
[20] X. S. Wei, H. J. Ye, X. Mu, J. Wu, C. Shen, and Z. H. Zhou, "Multiple instance learning with emerging novel class," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 5, pp. 2109–2120, May 2021.
[21] Y. X. Chen and J. Z. Wang, "Image categorization by learning and reasoning with regions," *J. Mach. Learn. Res.*, vol. 5, pp. 913–919, 2004.
[22] X. C. Wei, S. L. Xu, P. An, and J. Yang, "Multi-instance learning with incremental classes," *J. Comput. Res. Develop.*, vol. 59, no. 8, pp. 1723–1730, 2022.
[23] Y. Shi, Y. Gao, Y. Yang, Y. Zhang, and D. Wang, "Multimodal sparse representation-based classification for lung needle biopsy images," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2675–2685, Oct. 2013.
[24] J. Amores, "Multiple instance classification: Review, taxonomy and comparative study," *Artif. Intell.*, vol. 201, no. 1, pp. 81–105, 2013.
[25] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.
[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. IEEE Int. Conf. Neural Inf. Process. Syst.*, Doha, Qatar, 2012, pp. 1097–1105.
[27] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 1–9.
[28] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 3431–3440.
[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 770–778.
[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, SanDiego, USA, 2015, pp. 1–14.
[31] H. Zhang et al., "ResNeSt: Split-attention networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2022, pp. 2735–2745.
[32] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2Net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2021.
[33] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
[34] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 3859–3869.
[35] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, and E. I. C. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Florence, Italy, 2014, pp. 1626–1630.
[36] J. Wu, Y. Yu, C. Huang, and Y. Kai, "Deep multiple instance learning for image classification and auto-annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 3460–3469.
[37] C. Kyrkou and T. Theocharides, "Deep-learning-based aerial image classification for emergency response applications using unmanned aerial vehicles," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, 2019, pp. 517–525.
[38] Y. Long et al., "On creating benchmark dataset for aerial image interpretation: Reviews, guidances, and million-AID," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4205–4230, Apr. 2021.
[39] L. J. Long, F. He, and H. J. Liu, "The use of remote sensing satellite using deep learning in emergency monitoring of high-level landslides disaster in Jinsha River," *J. Supercomput.*, vol. 78, no. 9, pp. 11974–11920, 2022.
[40] L. H. Wan, S. H. Li, Y. Chen, Z. He, and Y. L. Shi, "Application of deep learning in land use classification for soil erosion using remote sensing," *Front. Earth Sci.*, vol. 10, 2022, Art. no. 849531.
[41] C. Kyrkou and T. Theocharides, "EmergencyNet: Efficient aerial image classification for drone-based emergency monitoring using atrous convolutional feature fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1687–1699, Mar. 2020.
[42] H. G. Sui, B. F. Zhao, C. Xu, M. T. Zhou, Z. T. Du, and J. Y. Liu, "Rapid extraction of flood disaster emergency information with multi-modal sequence remote sensing images," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 46, no. 10, pp. 1441–1449, 2021.
[43] C. Liu, S. M. E. Sepasgozar, Q. Zhang, and L. L. Ge, "A novel attention-based deep learning method for post-disaster building damage classification," *Expert Syst. Appl.*, vol. 202, 2022, Art. no. 117268.
[44] Z. Li, K. Xu, J. Xie, Q. Bi, and K. Qin, "Deep multiple instance convolutional neural networks for learning robust scene representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3685–3702, May 2020.
[45] X. H. Lian, Y. W. Pang, J. G. Han, and J. Pan, "Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation," *Pattern Recognit.*, vol. 110, 2021, Art. no. 107622.
[46] R. Gupta et al., "xBD: A dataset for assessing building damage from satellite imagery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop*, Long Beach, CA, USA, 2019, pp. 10–17.
[47] M. M. Cheng, Z. Zhang, W. Y. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, 2014, pp. 3286–3293.

Chengfan Li received the Ph.D. degree in computer science from Shanghai University, Shanghai, China, in 2012.

Since 2012, he has been an Associate Professor with the School of Computer Engineering and Science, Shanghai University. His research interests include disaster monitoring, remote sensing image processing, and computer application.

Zixuan Zhang received the B.E. degree in software engineering from Tianjin University of Finance and Economics, Tianjin, China, in 2021. She is currently working toward the M.S. degree in computer science and technology with the School of Computer Engineering and Science, Shanghai University, Shanghai, China.

Her research interests include object detection, deep learning, and remote sensing.

Lan Liu received the M.S. degree in english language and literature from Southwest University, Chongqing, China, in 2009, and the Ph.D. degree in computer application from Shanghai University, Shanghai, China, in 2018.

She is currently a Lecturer with the School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China. Her major is computer application. Her research interests include clustering classification, spatial image processing, deep learning, and laboratory management and information construction.

Jung Yoon Kim received the Ph.D. degree in game engineering from the Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, Seoul, South Korea, in 2013.

He is an Associate Professor with the Graduate School of Game, Gachon University, Seongnam, South Korea, and the Department of Game Media, College of Future Industry. He is also the Center Director of the Start-Up Education Center, Gachon University. From 2015 to 2018, he was the Vice President of Korea Game Developer Associations. Since 2016, he has been the Editor-in-Chief for the Korea Computer Game Association. His research interests include computer gaming, AI, virtual reality technology, and interactive technology.

Arun Kumar Sangaiah received the Ph.D. degree in computer science and engineering from the School of Computer Science and Engineering, VIT University, Vellore, India, in 2007 and 2014.

He is currently a Full Professor with the National Yunlin University of Science and Technology, Douliu, Taiwan. He has authored/coauthored more than 300 research articles in refereed journals (IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE, IEEE SYSTEMS JOURNAL, IEEE SENSORS JOURNAL, IEEE INTERNET OF THINGS JOURNAL, *ACM Transactions on Sensor Networks*), 11 edited books, as well as 1 patents (held and filed) and 3 projects, among one funded by Ministry of IT of India and few international projects (Chinese Academy of Sciences (CAS), Guangdong Research Fund, Australian Research Council) cost worth of 500 000 USD.

Dr. Sangaiah was the recipient of many awards, Clarivate Highly Cited Researcher, Yushan Young Scholar Fellowship, Top 2% Scientist, PIFI-CAS Fellowship, Top-10 Outstanding Researcher, CSI Significant Contributor, etc. He is also Editor-in-Chief and Associate Editor of various reputed ISI journals. He is a Visiting Scientist (2018–2019) with CAS, China, and Visiting Researcher (2019–2020) of Université Paris-Est, France.