


# Label-Driven Graph Convolutional Network for Multilabel Remote Sensing Image Classification

Boyi Ma , Falin Wu , *Member, IEEE*, Tianyang Hu , Loghman Fathollahi , Xiaohong Sui , Yushuang Liu ,  
and Byambakhuu Gantumur 

**Abstract**—Multilabel classification in remote sensing is very significant and plays an important role in extracting valuable information from satellite imagery. Ignoring the distinct information provided by labels in each image or transforming images into content-aware category representations without considering the inherent correlation of labels within the dataset can result in the establishment of improper relationships between images and labels, ultimately leading to a significant degradation in accuracy. To address this problem, this article proposes a label-driven graph convolutional network (LD-GCN) to excavate substantial information using the inherent correlation of labels from datasets and build a strong relationship between labels and images. The framework consists of two modules, i.e., the label recognition GCN (LRGCN) and the semantic enrichment module (SEM). The LRGCN module yields rich and valuable information from the inherent correlation of labels and builds a strong relationship between images and labels. The SEM further enriches the semantics obtained from LRGCN. Experiments conducted on UCM, AID, and DFC15 multilabel remote sensing datasets illustrate that LD-GCN outperforms the state-of-the-art methods on key evaluation metrics.

**Index Terms**—Graph convolutional network (GCN), label-driven GCN, multilabel image classification, remote sensing.

## I. INTRODUCTION

WITH the rapid advancement of technology, remote sensing images have entered an era characterized by high resolution and increasingly complex content. Classifying the satellite images by utilizing the ground object and spatial features

Manuscript received 6 September 2023; revised 6 November 2023 and 4 December 2023; accepted 14 December 2023. Date of publication 18 December 2023; date of current version 3 January 2024. This work was supported by Beihang University. (*Corresponding author: Falin Wu.*)

Boyi Ma, Falin Wu, and Tianyang Hu are with the SNARS Laboratory, School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China (e-mail: maboyi@buaa.edu.cn; falin.wu@buaa.edu.cn; huty\_11@buaa.edu.cn).

Loghman Fathollahi is with the SNARS Laboratory, School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China, and also with the Meteorological Department of West Azarbijan Province, Iran Meteorological Organization, Tehran 19395-4568, Iran (e-mail: fathollahi.loghman@gmail.com).

Xiaohong Sui is with the Institute of Remote Sensing Satellite, China Academy of Space Technology, Beijing 100048, China (e-mail: suixiaohong1118@163.com).

Yushuang Liu is with the Department of Navigation and Control, Beijing System Design Institute of Electro-mechanic Engineering, Beijing 100081, China (e-mail: yslu@buaa.edu.cn).

Byambakhuu Gantumur is with the Department of Geography, School of Arts and Sciences, National University of Mongolia, Ulaanbaatar 14200, Mongolia (e-mail: byambakhuu@num.edu.mn).

Digital Object Identifier 10.1109/JSTARS.2023.3344106

present in remote sensing images has emerged as a significant research challenge [1]. The single-label remote sensing image classification no longer satisfies the need to describe the rich information inside. Therefore, multilabel remote sensing image classification is becoming more and more important [2].

Due to advancements in deep learning algorithms, plenty of work has been done on multilabel classification based on deep neural network (DNN) [3], [4], [5], [6], [7]. Still, there is a lack of information inside the labels, which is very important in multilabel classification and needs to be addressed. Other works use recurrent neural network (RNN) or long-short term memory (LSTM) to process the information inside labels [8], [9], [10]. Nevertheless, such methods, only considering the adjacent label, omit the true relationship within different labels. As compared to the abovementioned methods, graph convolution network (GCN) can learn the relationship between nodes and has been improved by different methods [11], [12], which is more suitable for multilabel classification. Recently, abundant work has been done on GCN [13], [14], [15], [16], [17], [18], [19], [20], [21].

However, the works based on GCN have two main problems. Some of those works [14], [18], [19], [20], [21] only use GCN to extract labels' information. They build an improper relationship between images and labels. Moreover, the inherent correlation between labels that contains rich information are ignored by other works [15], [16], [17]. Specifically, inherent correlation is contained in the labels from remote sensing datasets. For example, category "tree" and "grass" always appear in the same image, while "car" and "sea" do not. Ignoring the important inherent correlation between labels will seriously weaken the rich information and an appropriate relationship between images and labels cannot be built by then.

To address those problems, this article proposes a label-driven graph convolutional network (LD-GCN) for multilabel remote sensing image classification. The framework of the LD-GCN consists of two main parts, which are the label recognition GCN (LRGCN) and the semantic enrichment module (SEM). The features of each image are first extracted by convolutional neural network (CNN). Rich information of labels is obtained by excavating their inherent correlation and a strong relationship between images and labels are built by the LRGCN module. Then, the relationship is further consolidated and the semantic information is acquired through the SEM. The overall framework is proposed to predict the results of each image's labels. The main contribution of our work can be summarized as follows.

- 1) The significant contribution is that this article proposes a novel label-driven graph convolutional network for multilabel remote sensing image classification, which builds a strong relationship between labels and images using inherent correlation of labels.
- 2) The proposed LRGCN module provides significant information from the inherent correlation of labels and establishes a strong relationship between images and label.
- 3) The proposed LD-GCN has a better performance than the state-of-the-art methods on key evaluation metrics. Specifically, LD-GCN achieves new records on the benchmark of mAPs of 97.96%, 83.49%, and 99.08% on three multilabel remote sensing datasets, i.e., UCM, AID, and DFC15, respectively.

## II. RELATED WORK

In recent years, the advancement of deep learning networks has accelerated the improvement of image classification. As multilabel classification tasks have become increasingly common, those methods can be categorized into DNN-based, relation-based, and GCN-based, due to their characteristics.

### A. Deep Neural Network-Based Methods for Multilabel Classification

With the development of CNN, many deep learning networks are employed in classification tasks. Boutell et al. [3] were the first scholar who proposed to use independent binary classifiers for each class when the labels were not mutually exclusive. Maxime et al. [4] employed the single-label classification on the ImageNet dataset and transferred the weights to multilabel classification datasets based on ground-truth bounding boxes, which highly limited the usage. In the remote sensing field, Wu et al. [7] proposed S-MAT that used a masked attention transformer to learn the contents of labels from images. Chen et al. [22] used a recurrent attention reinforcement learning framework to discover attention regions and the objects related to semantic and predict the final results. Gencer et al. [23] employed RNN to model spatial relationship, K-Branch CNN to extract image features, and eventually defined multiple attention scores for local descriptors.

With the emergence of vision transformer [24] and swin transformer [25] networks, much research has been conducted with transformer. Specifically, for processing remote sensing images, Tan et al. [26] used a transformer to extract semantic attentional regions from image features extracted by a deep CNN. Kaselimi et al. [27] utilized vision transformer to leverage the benefits of the self-attention mechanism and obviated any convolution operations.

### B. Relation-Based Methods for Multilabel Classification

To address the issue of DNN-based methods not uncovering the correlation of labels within each image, some studies have employed relation-based approaches to highlight dependencies between labels. Wang et al. [8] proposed a CNN-RNN framework for multilabel image classification tasks, which used RNN

after CNN to learn a joint image-label embedding and predicted the labels. Yeh et al. [28] derived a deep latent space and a label-correlation sensitive loss function to relate feature and label domain data. Hua et al. [29] extracted fine-grained semantic feature maps through an attention-based convolutional network and produced structured multiple object labels by the bidirectional long-short term memory (LSTM) network. Alshehri et al. [30] performed a network based on multiple loss functions to increase the similarity between the image with its corresponding labels. Liu et al. [25] captured the sequence information of the text and selected the valid features related to labels using bidirectional gated recurrent unit network (Bi-GRU). Despite considering the dependencies between labels, these works are unable to fully learn the complex relationship between images and labels.

### C. GCN-Based Methods for Multilabel Classification

With the widespread application of GCN, Lee et al. [13] first introduced the knowledge graph into the multilabel classification and accomplished the classification with zero samples. Chen et al. [14] proposed a multilabel classification model based on GCN (ML-GCN), which had a profound impact on multilabel classification tasks. They employed the word embedding of labels to represent nodes and learned about the interdependent relationship through GCN. And improvements on MLGCN had been done by some works [18], [19], [20], [21]. Ye et al. [16] supposed that ML-GCN trained the labels and images separately and used a static matrix led to the lack of generality of the model. Therefore, they proposed an attention-driven dynamic GCN (ADD-GCN) to dynamically generate a specific graph for each image. Following ADD-GCN, some improved networks have also been used to process the remote sensing images [15], [17]. Experiments show that GCN-based methods do perform better in multilabel image classification tasks.

Despite considering the dependencies between labels, these works are unable to fully learn the complex relationship between images and labels. Relationship-based methods can excavate information inside labels, however, the relationship between images and labels still cannot be learned. In order to handle these problems, the proposed framework inspired by the GCN-based methods, can utilize the inherent correlation of labels and capture the relationship between images and labels.

## III. METHODOLOGY

The overall architecture of the proposed framework is illustrated in Fig. 1. The framework is composed of two main module that are label recognition GCN (LRGCN), and semantic enrichment module (SEM). The details of LRGCN and SEM will be illustrated in Sections III-B and III-C, respectively. Moreover, the prediction layer and loss function will be presented in Section III-D.

### A. Preliminaries

The multilabel graph convolutional network (ML-GCN) [14], using GCN to learn the relationship of every label after word embedding, ignored the fact that images have their distinctive

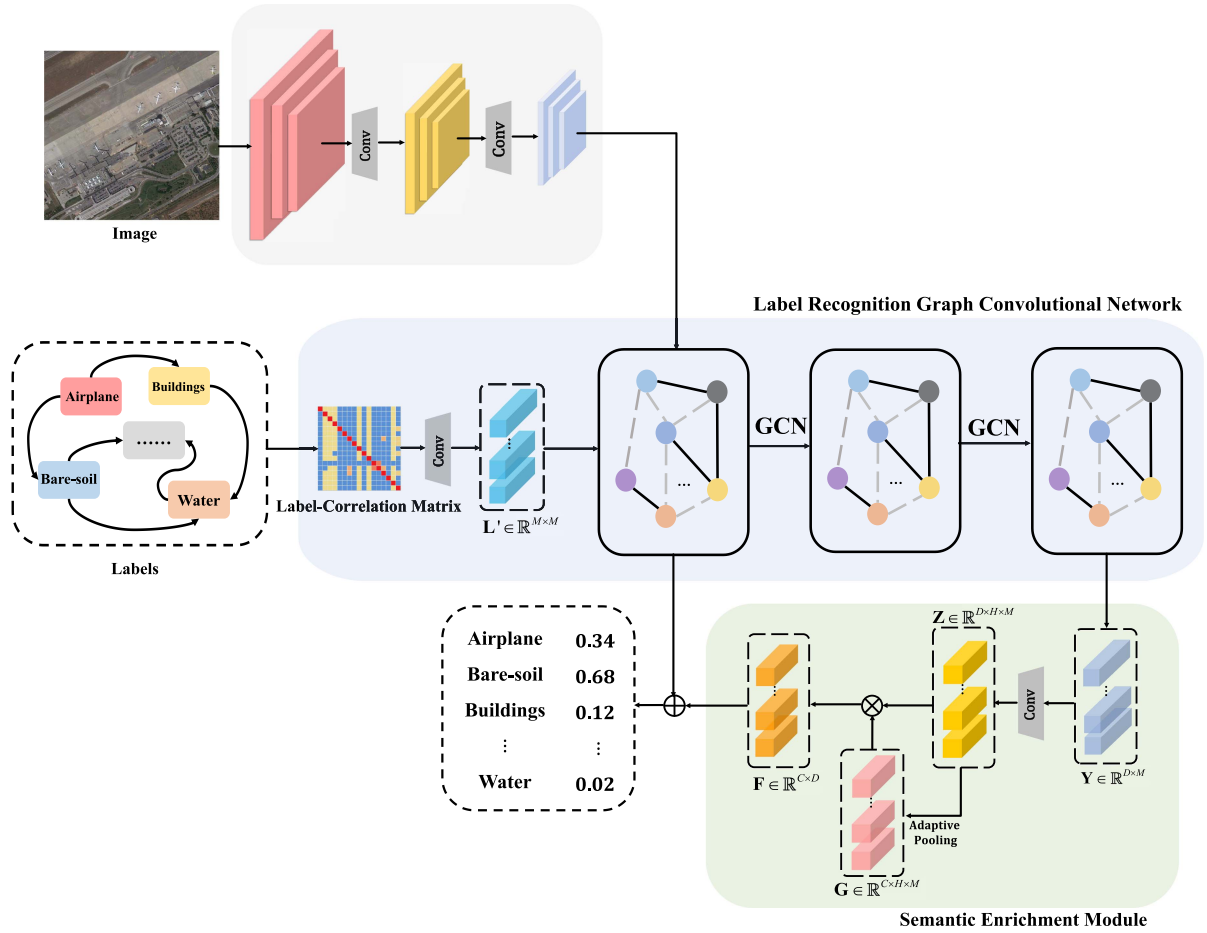


Fig. 1. Overall framework of the LD-GCN. The image features are extracted by CNN and sent to LRGCN. The inherent correlation of labels is sent to LRGCN to excavate rich information fully first. LRGCN builds a strong relationship between images and labels. SEM enriches the profound semantic further.

labels and the fusion of images and labels is very crucial in multilabel classification tasks and the rich information inside of images and related labels needs to be learned further. The attention-driven dynamic graph convolutional network (ADD-GCN) [16] obtained the content-aware category directly from image feature and the GCN is only used to learn the content out of images, which ignored the inherent correlation of labels from the dataset and did not build a relationship between images and labels.

After going through these methods, this article proposes a label-driven graph convolutional network which utilizes the inherent correlation of labels from dataset and extracts rich content fully while using GCN to learn the deep relationship between images and labels. The proposed framework can excavate the profound information inside of labels and build a strong relationship between labels and images.

### B. Label Recognition GCN (LRGCN)

Extracting features from images are essential for the relationship learning between images and labels. ResNet [31] as an image feature extraction has a clear advantage. On the basis of Bottleneck, which is composed of  $1 \times 1$  and  $3 \times 3$  convolutional

layers, the stage in ResNet also has batch normalization and a dropout module with residual, which can help the network build an ultra-deep structure. Therefore, ResNet with 3, 4, 23, and 3 layers, respectively, is chosen for the feature extraction stage. When the resolution of an input image is  $3 \times 448 \times 448$ , the output feature map from ResNet is  $2048 \times 14 \times 14$ . This network can collect global and local features in remote sensing images.

1) *Label Information Extracted:* While different images have distinctive labels, the labels in the dataset have corresponding correlation, which is inherent, especially in the remote sensing field. For remote sensing images, some labels are always presented in the same image as they have inherent correlation like “cars” and “pavement.” While other labels barely emerge in one image like “cars” and “sea.” The inherent correlation among the labels can be useful in multiclassification task due to their probability representation. Therefore, obtaining the inherent correlation of labels is very crucial.

To learn the inherent correlation of labels, a label-correlation matrix  $\mathbf{L} \in \mathbb{R}^{C \times C}$  is constructed, where  $C$  denotes the number of categories. The probability of the occurrence of label  $L_i$  when label  $L_j$  appears is demonstrated as the vector  $\mathbf{l}_{ij}$ . To obtain this probability, the appearing times of label  $L_i$  in the dataset

is calculated as  $N_i$ , and the occurrence times of  $L_i$  and  $L_j$  is calculated as  $Z_{ij}$ . Then, the vector  $\mathbf{l}_{ij}$  is obtained as follows:

$$\mathbf{l}_{ij} = P_{ij} = P(L_i|L_j) = Z_{ij}/N_i. \quad (1)$$

Through this formulation, the label-correlation matrix can be obtained.

To fully acquire rich information out of labels, the labels-correlation matrix needs to be further excavated. To extract more profound content from the matrix and to excavates the latent connection between labels and images, the matrix  $\mathbf{L} \in \mathbb{R}^{C \times C}$  is transformed to  $\mathbf{L}' \in \mathbb{R}^{M \times M}$ , where  $M = H \times W$ ,  $H$ , and  $W$  are the height and width of the image, respectively. After this, the rich information of labels is fully extracted as matrix  $\mathbf{L}'$ , which can be processed with the features of images.

2) *GCN-Based Relationship Building*: The graph convolutional network (GCN) [32] uses  $\mathbf{H}^l \in \mathbb{R}^{n \times d}$  as nodes to denote the description of the input feature. The correlation of nodes is described as the adjacency matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ . GCN processes the input nodes  $\mathbf{H}^l \in \mathbb{R}^{n \times d}$  by propagating weight matrix  $\mathbf{W}_A \in \mathbb{R}^{d \times d}$  and writes the new input nodes as  $\mathbf{H}^{l+1}$ . The new  $\mathbf{H}^{l+1}$  will be learned from the nonlinear function

$$\mathbf{H}^{l+1} = \sigma(\hat{\mathbf{A}}\mathbf{H}^l\mathbf{W}_A) \quad (2)$$

where  $\sigma$  denotes the sigmoid function, and  $\hat{\mathbf{A}} \in \mathbb{R}^{n \times n}$  is a symmetrically normalized version of matrix  $\mathbf{A}$ , which can be calculated by function

$$\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}} \quad (3)$$

where  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ ,  $\mathbf{I}$  is an identity matrix, and  $\tilde{\mathbf{D}}$  is the degree matrix of  $\tilde{\mathbf{A}}$ .

As GCN can learn the relationship between the nodes using the adjacency matrix, the features of images and the features of labels extracted from the label-correlation matrix are sent to GCN. The features of images are seen as nodes of GCN to learn. The processed label-correlation matrix is used as the adjacency matrix for nodes. By using label-correlation as the link of features, the GCN can learn deep information related to the content of labels. With stacked layers of GCN to excavate the deep link between images and labels, the strong relationship between images and labels can be built.

As shown in Fig. 2, the features of images extracted from ResNet-101 is reshaped as  $\mathbf{E} \in \mathbb{R}^{M \times D}$ , where  $M$  is the multiplication of the height and width of images, and  $D$  denotes the dimension of images. The adjacency matrix of GCN is the extracted label-correlation matrix  $\mathbf{L}' \in \mathbb{R}^{M \times M}$ . Then the output  $\mathbf{Y} \in \mathbb{R}^{M \times D}$  of LDGCN module can be defined as

$$\mathbf{Y} = \sigma(\hat{\mathbf{L}}'\mathbf{E}\mathbf{W}_l) \quad (4)$$

in which  $\mathbf{W}_l$  is the weight matrix and  $\hat{\mathbf{L}}' = \tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{L}}'\tilde{\mathbf{D}}^{-\frac{1}{2}}$  ( $\tilde{\mathbf{L}}' = \mathbf{L}' + \mathbf{I}$ ,  $\mathbf{I}$  is an identity matrix) and  $\tilde{\mathbf{D}}$  is the degree matrix of  $\tilde{\mathbf{L}}'$ .

After LRGCN module, the rich information about labels is obtained and the relationship between images and labels is built.

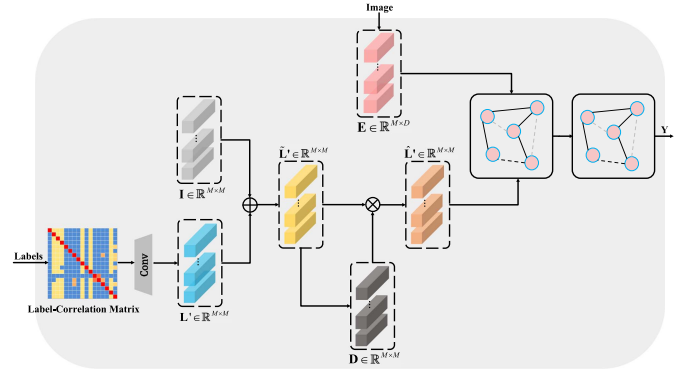


Fig. 2. Details of label recognition graph convolutional network.

### C. Semantic Enrichment Module (SEM)

Although the basic relationship between images and labels has been learned from the LRGCN module, there are still some limitations. Since the inherent correlation describes the relationship between all the labels in the dataset, the specific semantic of each image needs to be further extracted. In order to make full use of the information inside each distinctive image and further strengthen the connection between images and labels, the SEM module is proposed.

In order to find the rich representations of image, a kernel of  $3 \times 3$  convolutional operation and an activation function ReLu are utilized. To avoid overfitting, batch normalization is also used. The input  $\mathbf{Y} \in \mathbb{R}^{M \times D}$  is first reshaped as  $\mathbf{Y}' \in \mathbb{R}^{D \times H \times W}$  and the process can be described as

$$\mathbf{Z} = \text{ReLu}(\delta(\text{conv}(\mathbf{Y}')))) \quad (5)$$

where  $\text{conv}$  denotes a convolutional operation with the kernel of  $3 \times 3$  and  $\delta$  denotes the batch normalization. To find the link between the extracted matrix and the final category, the main focused feature of matrix  $\mathbf{Z} \in \mathbb{R}^{D \times H \times W}$  is first extracted by adaptive pooling as  $\mathbf{G} \in \mathbb{R}^{C \times H \times W}$  (where  $C$  denotes the number of categories). In order to retain the valuable information after the pooling, the  $\mathbf{G}$  is reshaped to  $\mathbf{G}' \in \mathbb{R}^{C \times M}$  (where  $M$  denotes the multiplication of height and width of images) and then multiplied with  $\mathbf{Z}' \in \mathbb{R}^{M \times D}$  which is reshaped from  $\mathbf{Z}$ . The formulation is written as

$$\mathbf{F} = \mathbf{G}' \times \mathbf{Z}' \quad (6)$$

where  $\mathbf{G}'$  is reshaped from  $\mathbf{G} = \text{LeakyReLu}(g(\mathbf{Z}))$ , in which  $g(\cdot)$  represents the average pooling operation and the LeakyReLu is the activation function. After this module, the enriched semantic information  $\mathbf{F} \in \mathbb{R}^{C \times D}$  is obtained and can be sent to the final prediction layer.

### D. Prediction Layer and Loss Function

To acquire the original features, which contain significant information from objects in remote sensing images, the dimension of feature  $\mathbf{E} \in \mathbb{R}^{M \times D}$  is reduced to  $\mathbf{E}' \in \mathbb{R}^{C \times D}$  and then is added by  $\mathbf{F}$ . The final matrix is sent into the binary classification

to predict a reliable result. The function can be demonstrated as

$$\text{output} = \text{Softmax}(\text{fc}(\mathbf{E}' + \mathbf{F})) \quad (7)$$

in which the  $\text{fc}(\cdot)$  is the fully connected layer and the  $\text{Softmax}$  is the activated function.

The loss function is applied to adjust the values of parameters to obtain an optimal model. Therefore, it is necessary to decrease the gap between predicted and true values. Therefore, the function of loss needs to make the gradient of the return proportional to the difference between the predicted and the true value. The formulation is illustrated as follows:

$$L = - \sum_{i=1}^C \left[ y_i \log \left( \frac{1}{1 + \exp(-o_i)} \right) + (1 - y_i) \log \left( 1 - \frac{1}{1 + \exp(-o_i)} \right) \right] \quad (8)$$

where  $y_i$  denotes the true labels of the image,  $o_i$  represents the predicted labels of the image,  $C$  is the number of overall labels of the image.

#### IV. EXPERIMENTS

In this section, the extensive experiments and analyses of LD-GCN are presented on three public datasets. First, the datasets are demonstrated, following the evaluation metrics and implementation details. Then, the comparison between LD-GCN with other state-of-the-art methods is conducted on three different datasets. Finally, the ablation experiments and the visualization analyses are presented.

##### A. Dataset

Plenty of public datasets are used in the multilabel remote sensing image classification field. Three representative datasets are chosen: UCM multilabel [33], AID multilabel [34], and DFC15 multilabel [29]. The details of these three datasets are illustrated below.

1) *UCM Multilabel Dataset*: UCM archive was introduced to the public in 2010 by Yang and Newsam [33]. The overall dataset was extracted from area images of the National Map of the U.S. Geological Survey. This dataset contains 2100 images for 21 categories. Not until Chaudhuri et al. [35] created a multilabel version of it, did the dataset called UCM multilabel applied to the multilabel image classification field. The UCM multilabel dataset has 17 different labels. Each image, with a size of  $3 \times 256 \times 256$  and a spatial resolution of 0.3 m, is marked with one or more labels. The examples of UCM multilabel dataset and their related labels are shown in Fig. 3(a).

2) *AID Multilabel Dataset*: The AID dataset was created by Xia et al. [34] from Wuhan University. The aerial images were collected from Google Earth imagery with sizes of  $3 \times 600 \times 600$  and spatial resolutions ranging from 0.5 to 8 m. In 2020, Hua et al. [36] selected some images from this dataset and created an AID multilabel dataset for multilabel classification. The AID multilabel dataset contains 3000 images which are marked by 17

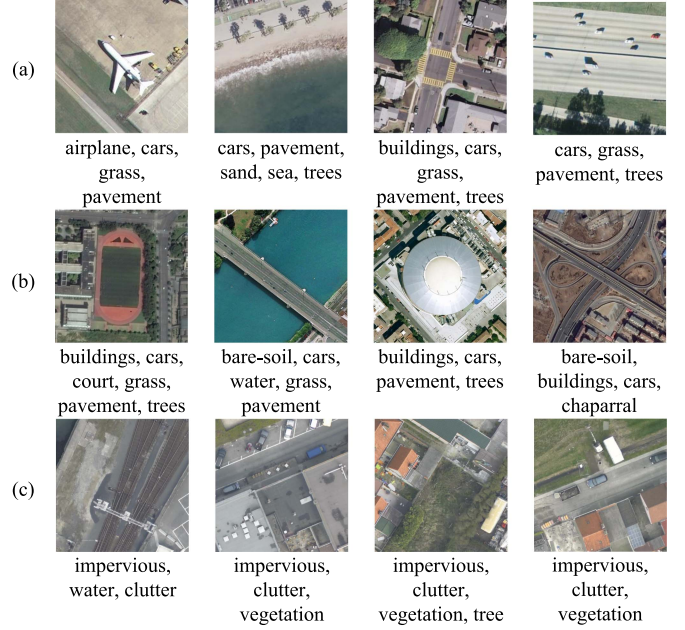


Fig. 3. Examples of the three multilabel datasets from different scenes.

different labels manually. The examples of the AID multilabel dataset and their related labels are shown in Fig. 3(b).

3) *DFC15 Multilabel Dataset*: The DFC15 multilabel dataset [29], containing 3342 images, is based on the original GRSS\_DFC\_2015 dataset, acquired over Zeebrugge with an airborne sensor. The original dataset is divided into eight categories. Each image from the DFC15 multilabel dataset is  $600 \times 600$  pixels with a spatial resolution of 5 cm and is marked with one or more labels. The examples of DFC15 multilabel dataset and their related labels are shown in Fig. 3(c).

##### B. Evaluation Metrics

To fully evaluate the performance of different methods, referring to some previous works [14], [16], seven evaluation metrics including the mean average precision ( $mAP$ ) and the average overall precision ( $OP$ ), recall ( $OR$ ), F1-scores ( $OF1$ ), and average per-class precision ( $CP$ ), recall ( $CR$ ), and F1-scores ( $CF1$ ) are used.

The formulas of the average overall precision, recall, and average per-class precision, and recall are defined as follows, respectively:

$$OP = \frac{\sum_i N_i^c}{\sum_i N_i^p} \quad (9)$$

$$OR = \frac{\sum_i N_i^c}{\sum_i N_i^g} \quad (10)$$

$$CP = \frac{1}{C} \sum_i \frac{N_i^c}{N_i^p} \quad (11)$$

$$CR = \frac{1}{C} \sum_i \frac{N_i^c}{N_i^g} \quad (12)$$

TABLE I  
COMPARISONS OF LD-GCN AND THE STATE-OF-THE-ART METHODS ON THE UCM, AID, AND DFC15 MULTILABEL DATASETS

Dataset	Methods	$OP$	$OR$	$OF1$	$CP$	$CR$	$CF1$	$mAP$
UCM	ResNet101 [31]	89.84	86.77	88.30	93.67	91.28	92.50	97.56
	ML-GCN [14]	74.48	67.89	71.00	69.34	54.56	61.02	68.89
	VAC [37]	79.72	90.70	84.86	88.92	88.91	88.12	94.39
	ADD-GCN [16]	84.48	83.72	84.10	89.30	84.07	86.58	93.67
	Zhu et al. [38]	91.75	91.65	90.62	92.96	92.60	92.66	—
	RBFNN [6]	88.37	87.75	88.05	92.40	87.79	90.02	—
	Huang et al. [39]	90.54	<b>92.98</b>	<b>91.74</b>	93.73	92.75	93.23	—
	ResNet50-SR-Net [26]	89.90	88.12	89.03	93.40	91.34	92.28	96.57
	MSGM [19]	83.86	85.48	84.54	89.98	85.07	87.46	—
	Our LD-GCN	<b>93.21</b>	88.39	90.73	<b>96.31</b>	<b>93.24</b>	<b>94.70</b>	<b>97.96</b>
AID	ResNet101 [31]	91.41	88.20	89.78	80.94	69.45	74.82	80.30
	ML-GCN [14]	84.10	83.36	83.72	65.86	52.66	58.36	63.24
	VAC [37]	86.22	91.23	88.64	75.16	73.73	73.40	80.57
	ADD-GCN [16]	92.15	89.01	90.55	82.75	71.38	76.62	81.82
	Zhu et al. [38]	89.72	88.41	87.49	80.89	74.08	76.50	—
	RBFNN [6]	88.52	86.56	87.52	78.78	64.16	70.70	—
	Huang et al. [39]	91.03	<b>91.88</b>	<b>91.44</b>	81.37	<b>74.60</b>	77.79	—
	ResNet50-SR-Net [26]	91.52	88.67	90.01	83.94	72.33	<b>77.89</b>	82.75
	Our LD-GCN	<b>92.81</b>	89.06	90.93	<b>85.14</b>	71.22	<b>77.49</b>	<b>83.49</b>
	DFC15	ResNet101 [31]	95.79	95.32	95.58	94.90	92.51	93.74
ML-GCN [14]		92.93	87.29	90.00	89.88	80.05	84.63	91.43
VAC [37]		94.26	96.25	95.24	91.43	95.07	93.10	98.19
ADD-GCN [16]		96.56	96.39	96.47	94.57	91.64	93.07	99.01
Huang et al. [39]		<b>96.68</b>	95.84	96.30	95.65	93.75	94.69	98.27
ResNet50-SR-Net [26]		96.16	96.09	96.12	94.70	94.28	94.48	98.76
MSGM [19]		94.61	92.71	93.65	91.42	90.70	91.06	—
Our LD-GCN		96.32	<b>97.09</b>	<b>96.74</b>	<b>95.43</b>	<b>95.50</b>	<b>95.54</b>	<b>99.08</b>

where  $i$  denotes the  $i$ th label,  $C$  denotes the number of labels in the dataset,  $N_i^c$  is the number of predicted images which is correctly predicted for the  $i$ th label,  $N_i^P$  is the number of images that are predicted for the  $i$ th label, and  $N_i^g$  is the number of ground truth images for the  $i$ th label.

For F1 metrics, the value is the harmonic mean of precision and recall and is described as follows:

$$OF1 = \frac{2 \times OP \times OR}{OP + OR} \quad (13)$$

$$CF1 = \frac{2 \times CP \times CR}{CP + CR}. \quad (14)$$

For mean average precision ( $mAP$ ), the formula is presented as

$$mAP = \frac{\sum_{i=1}^K AP_i}{K} \quad (15)$$

for  $K \in \{1, \dots, C\}$ , where  $K$  contains all the categories, and  $AP$  is defined as the average of precision of all categories. It is suggested that an image contains the label when the prediction result is greater than 0.5.

### C. Implementation Details

During training, to avoid overfitting, data augmentation is employed: resize the images to  $224 \times 224$  for the UCM multilabel and DFC15 multilabel datasets with random horizontal flip; crop and resize the images to  $512 \times 512$  for the AID multilabel dataset. For the proposed framework, ResNet is used as the backbone and the model is trained for 50 epochs with a batch size of 16. Stochastic gradient descent (SGD) with a momentum of 0.9 and weight decay of  $10^{-4}$  is chosen as the optimizer of the proposed model. The learning rate is set up to 0.05 initially for the whole framework and is reduced by 0.1 at 30 and 40 epoch,

TABLE II  
PARAMETERS AND FLOPS OF MODELS ON THE AID MULTILABEL DATASETS

Backbone	Params (M)	FLOPs (G)
ResNet101 [31]	42.54	252
ADD-GCN [16]	47.84	255
Huang et al. [39]	137.54	132
Our LD-GCN	99.28	334

respectively. The overall experiments are conducted based on a Pytorch platform with NVIDIA GeForce GTX 3090 GPU.

### D. Comparisons With State-of-the-Art Methods

To illustrate the performance of the proposed framework, LD-GCN is compared with other state-of-the-art methods on three different datasets, which are the UCM, AID, and DFC15 multilabel datasets. The results of the comparison are shown in Table I, and the best results are marked as bold in the table. ResNet101 [31] is the baseline. ML-GCN [18] is the framework that originally employed GCN in multilabel classification tasks. VAC [37] uses an attention mechanism to classify multilabel images. ADD-GCN [16] employs the framework with an attention mechanism and GCN. Zhu et al. [38] use dual-level semantics to guide multilabel classification. RBFNN [6] utilizes data augmentation technique to increase the size of the dataset. Huang et al. [39] use multiscale feature fusion with channel-spatial attention learning to achieve classification tasks. ResNet50-SR-Net [26] is a transformer-driven semantic relation inference network. MSGM [19] is a spatial pyramid convolutional network.

The comparison results show that LD-GCN is superior at most evaluation metrics on the UCM multilabel dataset. For the evaluation metrics like  $OP$  and  $CP$ , our method surpasses

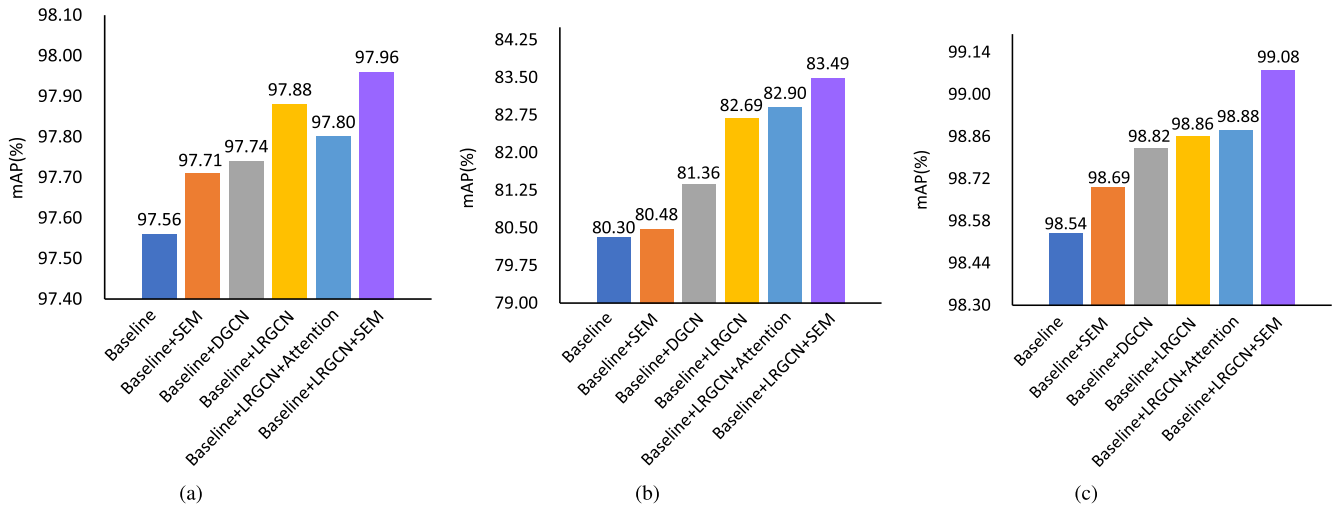


Fig. 4. Performance analysis of LRGCN and SEM modules on UCM, AID, and DFC15 datasets. (a) UCM. (b) AID. (c) DFC15.

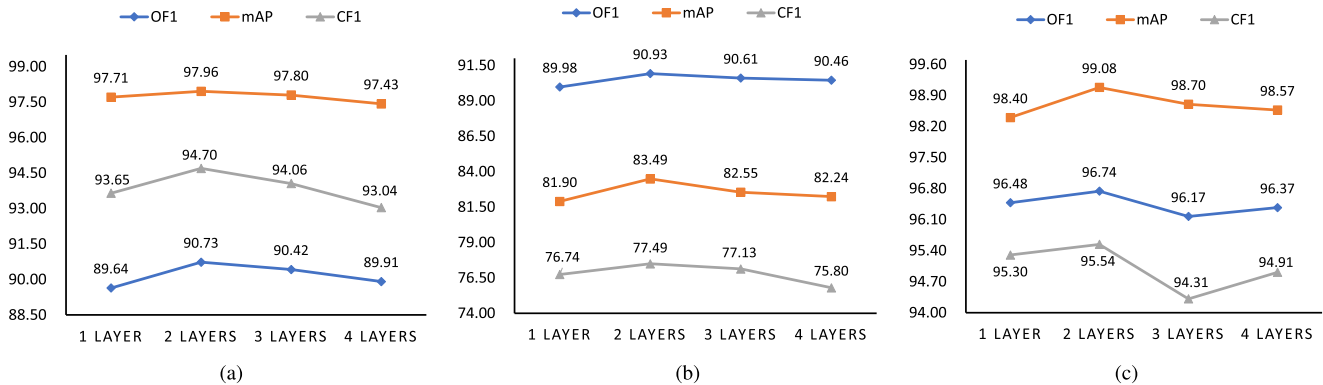


Fig. 5. Comparisons of the number of GCN layers on UCM, AID, and DFC15 multilabel datasets. Abscissa indicates the number of layers used and ordinate indicates the performance. The shape of square, rhombus, and triangle in the line chart illustrates  $mAP$ ,  $OF1$ , and  $CF1$ , respectively. (a) UCM. (b) AID. (c) DFC15.

the second highest scores by 1.46% and 2.61%, respectively. LD-GCN is only lower than the method proposed by Huang et al. [39] on  $OR$  and  $OF1$  scores while surpassing all the other evaluation metrics. In short, the comparison results demonstrate that our framework achieves superior performance on UCM multilabel dataset.

As for the AID multilabel dataset, LD-GCN has the best results at  $OP$ ,  $CP$ , and  $mAP$  scores, which shows the improvement over the state-of-the-art results by around 1%. In the meantime, LD-GCN achieves the second-highest results at  $OF1$  scores. Relatively, LD-GCN has lower results at recall scores, and thus, has lower performance at F1 scores. It can be inferred that because LD-GCN considers the correlation of labels, which is influenced by the occurrence times of two labels, and some labels hardly emerged in one image may repel each other, it is harder for LD-GCN to find out all the labels in one complex remote sensing image. Except for these factors, overall, the framework has relatively good results on the AID multilabel dataset.

A bunch of methods using the DFC15 multilabel dataset are listed for comparison, which are ML-GCN, VAC, ADD-GCN, Huang et al. [39], ResNet50-SR-Net, MSGM, and baseline ResNet101. The LD-GCN has a significant performance on this dataset with only 0.36% lower than ADD-GCN on  $OP$  scores. The performance of the recall is also good on this dataset, which demonstrates that this dataset has fewer labels for every image and the proposed framework provides the best results.

As shown in comparison, LD-GCN achieves the best results at most evaluation metrics on UCM and DFC15 multilabel datasets, which justifies that the proposed framework performs an obvious superiority. In the meantime, LD-GCN reaches the best results on  $OP$ ,  $CP$ , and  $mAP$  scores and the second-highest results on  $OF1$  scores. Therefore, the proposed framework shows the good performance on the AID multilabel dataset. In summary, the experiments demonstrate the superiority of the proposed framework—LD-GCN.

To demonstrate the computational space complexity and the time complexity, the parameters and floating-point operations

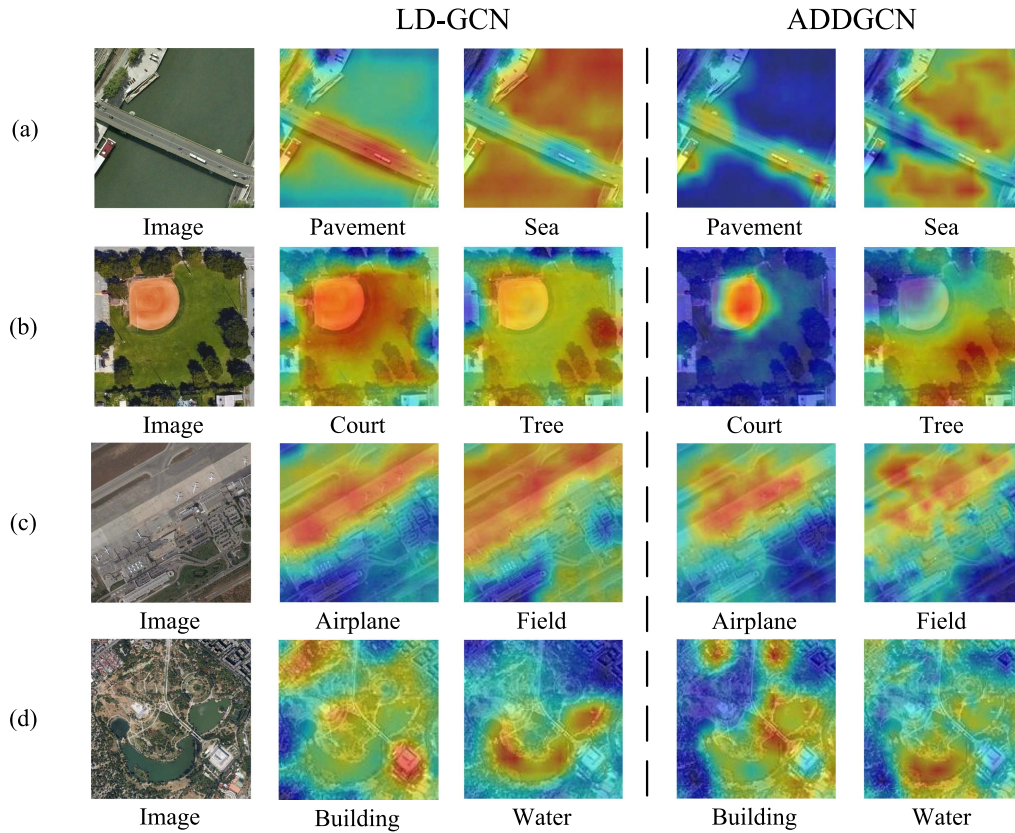


Fig. 6. Sample visualization of training process of LD-GCN compared with ADDGCN on AID multilabel dataset. The visualization results of LD-GCN are shown on the left and the compared ADD-GCN are shown on the right.

(FLOPs) of the proposed model is compared with some models, which are the baseline ResNet101, ADDGCN, and the model proposed by Huang et al. [39].

As can be seen from Table II, while the parameters and FLOPs of LD-GCN are both higher than ResNet101 and ADDGCN, the performance of the proposed LD-GCN is far better than that of these two models. Compared with the model proposed by Huang et al. [39], the LD-GCN has a higher computational time complexity with a lower parameter. Despite the higher FLOPs, the proposed LD-GCN shows the superiority at many evaluation metrics on three different datasets compared with Huang et al. [39].

### E. Ablation Experiments

The ablation experiments are carried out to further evaluate the effect of the proposed LD-GCN. Experiments are conducted on the effects of LRGCN and SEM modules and the number of GCN layers.

1) *Effects of LRGCN and SEM Modules*: To find out the contribution of each module in LD-GCN, and to illustrate the better performance of our module, the experiments are also conducted on the baseline. Then, LRGCN module and the SEM module are added to the framework with the DGCN module and an attention module for comparison. To investigate the effect more thoroughly, experiments are made on UCM, AID, and DFC15 multilabel datasets and the results are conveyed

by the most representative evaluation metric— $mAP$ . All the comparison results for the three datasets are shown in Fig. 4.

From the comparison results, when the SEM module is added to the framework, the accuracy improves by 0.15%, 0.18%, and 0.15% on three datasets, respectively. It demonstrates that enriching the semantic information does improve the expression of features. When the compared module DGCN is added to the framework, the accuracy increases by about 1% on the AID dataset while rising by only around 0.2% on the UCM and DFC15 datasets. While adding the proposed LRGCN to the framework, as shown in the histogram, the performance has improved a lot from baseline on three datasets. LRGCN also outperformed DGCN greatly on UCM and AID multilabel datasets, which can be inferred that the proposed module can demonstrate complex images with multiple labels better. Applying the inherent correlation of labels into the framework and using GCN to build the relationship between images and labels does increase the accuracy significantly. To demonstrate the effectiveness of SEM on the network, the SEM is compared with an attention module, which has been used in many multilabel classification methods. The framework with an attention module can slightly improve the baseline on the AID and DFC15 datasets. However, it can damage the accuracy on UCM dataset. It illustrates that an attention module can cause disorder in the representations extracted from LRGCN. The framework with SEM is higher than the framework with the attention module by 0.16%, 0.59%, and 0.20% on the UCM, AID, and DFC15 datasets, respectively.



In conclusion, with the module LRGCN and SEM, the network can achieve the best results on three multilabel datasets.

2) *Effect of the Number of GCN Layers*: Different numbers of stacked layers are used in the LRGCN module for comparison, as shown in Fig. 5. From the line chart, as for  $mAP$  scores on three datasets, it can be inferred that when using one GCN layer, the performance is not the best because, through only one layer, the network cannot find the right relationship between the images and labels, which leads to confused judgment. The accuracy when stacking more layers is not as good as stacking two layers. This is because when using many GCN layers, the model takes a large amount of data and becomes overfitting which badly affects precision. As for  $OF1$  and  $CF1$  scores, although stacking four layers is better than three layers on DFC15 multilabel dataset, it shows that using two GCN layers has the best F1 scores on three datasets. In summary, the comparison with different layers on three datasets demonstrates that when using two layers of GCN, the network can achieve the best performance.

### F. Visualization

In this section, the visualization analysis of training of the proposed LD-GCN is conducted. The visualization of ADDGCN is carried out for comparison. The AID multilabel dataset has more complex remote sensing images compared to the other datasets. Therefore, four representative images are chosen from AID multilabel dataset. As shown in Fig. 6, each row presents the original image, activation maps with a related label on LD-GCN, and activation maps with the same label on ADDGCN.

In Fig. 6(a), the labels in this image are “pavement” and “sea.” These two labels have their own activated areas, which are clear and noninterference because these two labels have little correlation. While the label “pavement” has a key-activated area on “cars,” which is the known label in the dataset. This shows that the label focused by the proposed LD-GCN contains the correlation between labels. In Fig. 6(b), the labels of the image are “court” and “trees.” Those two labels are presented at the same time in most occasion, therefore, the activated maps have an overlapped area. In Fig. 6(c) and (d), the images are more complex and the predicting procedure is more complicated. For example, the label “airplane” in Fig. 6(c) has a bigger activated area than the airplane itself, because the semantic information around the plane is also considered when predicting the true label. So as the label “field.” Compared with ADDGCN, LD-GCN always has more activated area for one label and the activated areas are interference which means that the LD-GCN considers the inherent correlation of labels while predicting the final results.

The more complex an image is, the more activated areas will overlap. It can be seen from the visualization, some of the labels have complicated correlations with other labels, which also illustrates the importance of the inherent correlation between labels.

## V. CONCLUSION

In this article, an LD-GCN is proposed for multilabel remote sensing image classification. The inherent correlation of labels

in images is very crucial for multilabel image classification and the content inside of labels needs to excavate further. Therefore, the label-correlation matrix containing the inherent correlation of labels is learned and sent into the proposed module LRGCN to obtain the significant information inside the labels. Moreover, after extracting profound information from labels, a strong relationship between images and labels is built by LRGCN. Extensive experiments on UCM, AID, and DFC15 multilabel datasets demonstrate the superiority of the proposed LD-GCN model. The  $mAP$ s of LD-GCN on three datasets achieved 97.96%, 83.49%, and 99.08%, which are the new records.

In our future research, we plan to delve deeper into exploring more effective network backbones to further augment the robustness of multilabel classification in remote sensing applications. In addition to exploring network architectures, another crucial aspect of our future work will involve investigating strategies for incorporating domain-specific knowledge and data augmentation techniques which can potentially enhance the robustness and generalizability of the classification system.

## ACKNOWLEDGMENT

The authors would like to thank S. Sarwar for improving the English writing of this article and the researchers who kindly shared their codes. The authors would also like to thank the open multilabel remote sensing datasets—UCM, AID, and DFC15 used in this article.

## REFERENCES

- [1] S. Dutta and M. Das, “Remote sensing scene classification under scarcity of labelled samples—a survey of the state-of-the-arts,” *Comput. Geosciences*, vol. 171, 2023, Art. no. 105295, doi: [10.1016/j.cageo.2022.105295](https://doi.org/10.1016/j.cageo.2022.105295).
- [2] M. Stojimchev, D. Koccev, and S. Dzeroski, “Deep network architectures as feature extractors for multi-label classification of remote sensing images,” *Remote Sens.*, vol. 15, no. 2, 2023, Art. no. 538, doi: [10.3390/rs15020538](https://doi.org/10.3390/rs15020538).
- [3] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, “Learning multi-label scene classification,” *Pattern Recognit.*, vol. 37, no. 9, pp. 1757–1771, 2004, doi: [10.1016/j.patcog.2004.03.009](https://doi.org/10.1016/j.patcog.2004.03.009).
- [4] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, “Learning and transferring mid-level image representations using convolutional neural networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1717–1724, doi: [10.1109/CVPR.2014.222](https://doi.org/10.1109/CVPR.2014.222).
- [5] Y. Wei et al., “HCP: A flexible CNN framework for multi-label image classification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1901–1907, Sep. 2016, doi: [10.1109/TPAMI.2015.2491929](https://doi.org/10.1109/TPAMI.2015.2491929).
- [6] R. Stivaktakis, G. Tsagkatakis, and P. Tsakalides, “Deep learning for multilabel land cover scene categorization using data augmentation,” *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1031–1035, Jul. 2019, doi: [10.1109/LGRS.2019.2893306](https://doi.org/10.1109/LGRS.2019.2893306).
- [7] H. Wu, C. Xu, and H. Liu, “S-MAT: Semantic-driven masked attention transformer for multi-label aerial image classification,” *Sensors*, vol. 22, no. 14, 2022, Art. no. 5433, doi: [10.3390/s22145433](https://doi.org/10.3390/s22145433).
- [8] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, “CNN-RNN: A unified framework for multi-label image classification,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2285–2294, doi: [10.1109/CVPR.2016.251](https://doi.org/10.1109/CVPR.2016.251).
- [9] J. Zhang, Q. Wu, C. Shen, J. Zhang, and J. Lu, “Multilabel image classification with regional latent semantic dependencies,” *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2801–2813, Oct. 2018, doi: [10.1109/tmm.2018.2812605](https://doi.org/10.1109/tmm.2018.2812605).
- [10] F. Zhu, H. Li, W. Ouyang, N. Yu, and X. Wang, “Learning spatial regularization with image-level supervisions for multi-label image classification,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2027–2036, doi: [10.1109/CVPR.2017.219](https://doi.org/10.1109/CVPR.2017.219).

- [11] G. Zhou, J. Xu, W. Chen, X. Li, J. Li, and L. Wang, "Deep feature enhancement method for land cover with irregular and sparse spatial distribution features: A case study on open-pit mining," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4401220, doi: [10.1109/TGRS.2023.3241331](https://doi.org/10.1109/TGRS.2023.3241331).
- [12] H. Duan, Y. Zhu, X. Liang, Z. Zhu, and P. Liu, "Multi-feature fused collaborative attention network for sequential recommendation with semantic-enriched contrastive learning," *Inf. Process. Manage.*, vol. 60, no. 5, 2023, Art. no. 103416, doi: [10.1016/j.ipm.2023.103416](https://doi.org/10.1016/j.ipm.2023.103416).
- [13] C.-W. Lee, W. Fang, C.-K. Yeh, and Y.-C. F. Wang, "Multi-label zero-shot learning with structured knowledge graphs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1576–1585, doi: [10.1109/CVPR.2018.00170](https://doi.org/10.1109/CVPR.2018.00170).
- [14] Z. M. Chen, X. S. Wei, P. Wang, and Y. Guo, "Multi-label image recognition with graph convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5172–5181, doi: [10.1109/CVPR.2019.05932](https://doi.org/10.1109/CVPR.2019.05932).
- [15] N. Khan, U. Chaudhuri, B. Banerjee, and S. Chaudhuri, "Graph convolutional network for multi-label VHR remote sensing scene recognition," *Neurocomputing*, vol. 357, pp. 36–46, 2019, doi: [10.1016/j.neucom.2019.05.024](https://doi.org/10.1016/j.neucom.2019.05.024).
- [16] J. Ye, J. He, X. Peng, W. Wu, and Y. Qiao, "Attention-driven dynamic graph convolutional network for multi-label image recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 649–665, doi: [10.1007/978-3-030-58589-1\\_39](https://doi.org/10.1007/978-3-030-58589-1_39).
- [17] Y. Diao, J. Chen, and Y. Qian, "Multi-label remote sensing image classification with deformable convolutions and graph neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 521–524, doi: [10.1109/IGARSS39084.2020.9324530](https://doi.org/10.1109/IGARSS39084.2020.9324530).
- [18] M. H. Yang, H. Liu, L. Gao, Y. R. Qian, and Z. Q. Xiao, "DCA-GCN: A dual-branching channel attention and graph convolution network for multi-label remote sensing image classification," *J. Appl. Remote Sens.*, vol. 15, no. 4, 2021, Art. no. 044519, doi: [10.1117/1.Jrs.15.044519](https://doi.org/10.1117/1.Jrs.15.044519).
- [19] D. Lin, Z. Chen, and B. Dong, "Semantic understandings for aerial images via multigrained feature grouping," *Sci. Program.*, vol. 2022, 2022, Art. no. 1822539, doi: [10.1155/2022/1822539](https://doi.org/10.1155/2022/1822539).
- [20] D. D. Sun, L. L. Ma, Z. L. Ding, and B. Luo, "An attention-driven multi-label image classification with semantic embedding and graph convolutional networks," *Cogn. Computation*, vol. 15, pp. 1308–1319, 2022, doi: [10.1007/s12559-021-09977-9](https://doi.org/10.1007/s12559-021-09977-9).
- [21] P. Li, P. Chen, and D. Z. Zhang, "Cross-modal feature representation learning and label graph mining in a residual multi-attentional CNN-LSTM network for multi-label aerial scene classification," *Remote Sens.*, vol. 14, no. 10, May 2022, Art. no. 2424, doi: [10.3390/rs14102424](https://doi.org/10.3390/rs14102424).
- [22] T. Chen, Z. Wang, G. Li, and L. Lin, "Recurrent attentional reinforcement learning for multi-label image recognition," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 6730–6737, doi: [10.1609/aaai.v32i1.12281](https://doi.org/10.1609/aaai.v32i1.12281).
- [23] G. Sumbul and B. Demir, "A novel multi-attention driven system for multi-label remote sensing image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5726–5729, doi: [10.1109/IGARSS.2019.8898188](https://doi.org/10.1109/IGARSS.2019.8898188).
- [24] Y. Xu et al., "Transformers in computational visual media: A survey," *Comput. Vis. Media*, vol. 8, no. 1, pp. 33–62, Mar. 2022, doi: [10.1007/s41095-021-0247-3](https://doi.org/10.1007/s41095-021-0247-3).
- [25] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 9992–10002, doi: [10.1109/ICCV48922.2021.00986](https://doi.org/10.1109/ICCV48922.2021.00986).
- [26] X. W. Tan, Z. F. Xiao, J. J. Zhu, Q. Wan, K. Wang, and D. R. Li, "Transformer-driven semantic relation inference for multilabel classification of high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1884–1901, 2022, doi: [10.1109/Jstars.2022.3145042](https://doi.org/10.1109/Jstars.2022.3145042).
- [27] M. Kaselimi, A. Voulodimos, I. Daskalopoulos, N. Doulamis, and A. Doulamis, "A vision transformer model for convolution-free multilabel classification of satellite imagery in deforestation monitoring," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 7, pp. 3299–3307, Jul. 2023, doi: [10.1109/TNNLS.2022.3144791](https://doi.org/10.1109/TNNLS.2022.3144791).
- [28] C.-K. Yeh, W.-C. Wu, W.-J. Ko, and Y.-C. F. Wang, "Learning deep latent spaces for multi-label classification," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 2838–2844, doi: [10.1609/aaai.v31i1.10769](https://doi.org/10.1609/aaai.v31i1.10769).
- [29] Y. Hua, L. Mou, and X. X. Zhu, "Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional LSTM network for multi-label aerial image classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 149, pp. 188–199, 2019, doi: [10.1016/j.isprsjprs.2019.01.015](https://doi.org/10.1016/j.isprsjprs.2019.01.015).
- [30] A. Alshehri, Y. Bazi, N. Ammour, and N. Alajlan, "Multi-label classification of remote sensing imagery with deep neural networks," in *Proc. Mediterranean Middle-East Geosci. Remote Sens. Symp.*, 2020, pp. 97–100, doi: [10.1109/M2GARSS47143.2020.9105203](https://doi.org/10.1109/M2GARSS47143.2020.9105203).
- [31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [32] F. Gama et al., "Convolutional graph neural networks," in *Proc. Conf. Rec. 53rd Asilomar Conf. Signal, Syst. Comput.*, 2019, pp. 452–456, doi: [10.1109/IEEECONF44664.2019.9048767](https://doi.org/10.1109/IEEECONF44664.2019.9048767).
- [33] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279, doi: [10.1145/1869790.1869829](https://doi.org/10.1145/1869790.1869829).
- [34] G.-S. Xia et al., "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017, doi: [10.1109/tgrs.2017.2685945](https://doi.org/10.1109/tgrs.2017.2685945).
- [35] B. Chaudhuri, B. Demir, S. Chaudhuri, and L. Bruzzone, "Multilabel remote sensing image retrieval using a semisupervised graph-theoretic method," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1144–1158, Feb. 2018, doi: [10.1109/tgrs.2017.2760909](https://doi.org/10.1109/tgrs.2017.2760909).
- [36] Y. Hua, L. Mou, and X. X. Zhu, "Relation network for multilabel aerial image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4558–4572, Jul. 2020, doi: [10.1109/TGRS.2019.2963364](https://doi.org/10.1109/TGRS.2019.2963364).
- [37] H. Guo, K. Zheng, X. Fan, H. Yu, and S. Wang, "Visual attention consistency under image transforms for multi-label image classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 729–739, doi: [10.1109/CVPR.2019.00082](https://doi.org/10.1109/CVPR.2019.00082).
- [38] P. Zhu et al., "Deep learning for multilabel remote sensing image annotation with dual-level semantic concepts," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4047–4060, Jun. 2020, doi: [10.1109/tgrs.2019.2960466](https://doi.org/10.1109/tgrs.2019.2960466).
- [39] R. Huang, F. C. Zheng, and W. Huang, "Multilabel remote sensing image annotation with multiscale attention and label correlation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6951–6961, 2021, doi: [10.1109/Jstars.2021.3091134](https://doi.org/10.1109/Jstars.2021.3091134).



**Boyi Ma** received the B.S. degree in measurement and control technology and instrumentation from the School of Mechanical Engineering, Sichuan University, Chengdu, China, in 2021. She is currently working toward the M.S. degree in feature fusion for remote sensing multi-label image classification based on deep learning from the SNARS Laboratory, School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, China.

Her research interests include deep learning and image processing.



**Falin Wu** (Member, IEEE) received the Ph.D. degree in marine information system engineering from the Tokyo University of Marine Science and Technology, Tokyo, Japan, in 2004.

He is an Associate Professor with the SNARS Laboratory, School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, China. He is also currently working with the Regional Centre for Space Science and Technology Education in Asia and the Pacific (China) (affiliated to the United Nations) (RCSSTEAP). His research interests

include GNSS positioning and navigation, remote sensing applications, multi-sensor integration, GNSS atmospheric physics, meteorology, and space geodesy.



**Tianyang Hu** received the B.S. degree in measurement and control technology and instruments from Jilin University, Changchun, China, in 2021. She is currently working toward the M.S. degree in ship detection and recognition in synthetic aperture radar imagery using deep learning method with the SNARS Laboratory, School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, China.

Her research interests include deep learning and synthetic aperture radar (SAR) object detection.



**Loghman Fathollahi** received the master's degree in space technology application from Beihang University, Beijing, China, in 2018.

He is currently an Engineer with the Meteorological Department of West Azarbijan Province, Iran Meteorological Organization (IRIMO), Tehran, Iran, and also a Research Assistant with the SNARS Laboratory, School of Instrumentation and Optoelectronic Engineering, Beihang University. His research interests include GNSS, remote sensing, atmospheric science, machine learning, deep learning, and experimental

studies on the utilization of satellites in order to test the fundamental roles of nature.



**Yushuang Liu** received the and master's degree in control theory and control engineering from the Xi'an University of Technology, Xi'an, China, in 2011, and the Ph.D. degree in navigation guidance and control from Beihang University, Beijing, China, in 2016.

He is a Senior Engineer with the Beijing System Design Institute of Electro-mechanic Engineering, Beijing, China. His research interests include maneuvering target tracking, multisource information fusion, UAV route planning, and multiagent reinforcement learning.



**Xiaohong Sui** received the Ph.D. degree in spacecraft design from the China Academy of Space and Technology, Beijing, China, in 2017.

She is currently an Engineer with the Institute of Remote Sensing Satellite, China Academy of Space Technology, Beijing, China. Her research interests include ocean altimetry satellite system design, data application of ocean gravity inversion and bathymetry prediction, and remote sensing image processing.



**Byambakhuu Gantumur** received the Ph.D. degree in space technology applications from Beihang University, Beijing, China, in 2020.

He is currently an Associate Professor with the Laboratory of Remote Sensing and Geographic Information Systems (GEO-iLAB), Graduate School and the Department of Geography, School of Arts and Sciences, National University of Mongolia, Ulan Bator, Mongolia. His research interests include geography with urban remote sensing, and wildfire natural disaster.