

A hierarchical geometry-to-semantic fusion GNN framework for earth surface anomalies detection

Boan Chen^{1(✉)}, Aohan Hu¹, Mengjie Xie¹, Zhi Gao^{1(✉)}, Xuhui Zhao¹, and Han Yi²

¹ School of Remote Sensing and Information Engineering,

Wuhan University, Wuhan 430079, China

{cbchen1997,whu_ahah,mengjie_xie,zhaoxuhui}@whu.edu.cn
gaozhinus@gmail.com

² School of Computing, National University of Singapore,

Singapore 117417, Singapore

hany24@u.nus.edu

Abstract. The increasing occurrence of earth surface anomalies (ESA) underlines the importance of timely and accurate detection of such events. Therefore, researchers have utilized satellite imagery for large-scale detection and developed advanced deep learning methods. However, the performance is hindered by inadequate labeled data and the complexity of semantic information in satellite imagery. To this end, we propose a hierarchical geometry-to-semantic fusion graph neural network (GNN) framework. Specifically, our method employs two branches to extract geoentities and construct graphs at different levels. Then, a hierarchical graph attention network (GAT) is used to mine complex semantic information from graphs, facilitating accurate and rapid detection of ESA. To fill the gap of the lack of benchmark datasets, we create a composite dataset ESAD based on existing datasets for ESA detection. Extensive experiments demonstrate that the proposed method is effective for accurate ESA detection, outperforming many baseline methods.

Keywords: Earth surface anomalies · Hierarchical fusion · Graph neural network · Satellite imagery

1 Introduction

Earth surface anomalies (ESA) refers to sudden events on earth's surface caused by natural and human factors (e.g., natural disasters) [1]. In recent years, numerous ESA have caused significant loss of life and property damage, emphasizing the need for research in Humanitarian Assistance and Disaster Response (HADR) [2]. Timely detection of ESA is of great significance for early rescue and loss reduction [3]. The integration of satellite imagery and deep learning methods provides a large-scale, accurate solution for HADR challenges and has gained considerable attention. However, the scarcity of labeled data and the complexity of satellite imagery hinder the performance of these methods [4].

To achieve accurate detection of ESA, several methods have attempted to introduce additional temporal or modal data [5,6]. However, these methods entail demands for data availability and preprocessing, posing challenges for rapid response. More efficient ESA detection can be achieved via single-image based methods, including deep visual feature-based supervised learning methods like Convolutional Neural Network (CNN), but the overfitting problem caused by limited labeled data is challenging to address. Hence, certain studies have sought to gain insights from the success of unsupervised learning in industrial anomaly detection [7,8]. However, satellite imagery is much more complex than industrial images, leading to unsatisfactory model performance. Meanwhile, abnormal samples can provide valuable information and should not be completely ignored. Instead, we should strive to learn more useful information from limited data [9].

In order to address the aforementioned problems, we propose a two-stage framework, i.e. Hierarchical Geometry-to-Semantic Fusion Graph Neural Network (GNN), which solely utilizes a single image for detecting ESA, thereby reducing data requirements and decreasing preprocessing and inference time. In the first stage, the graph is constructed through a two-branch module that obtains geoentities and their relationships from a single satellite image. In the second stage, a hierarchical graph attention network (GAT) and attention-based fusion module are used to extract the graph-level embedding. Inspired by how human interpret satellite imagery, proposed method aims at simulating the human brain's ability to learn higher-level information from objects and their relationships. Meanwhile, existing datasets mostly focus on single or a few classes of ESA, e.g., flooding, hurricane, and landslide, while a large-scale, multi-class dataset is still lacking. To bridge this gap, we create a composite dataset ESAD, based on existing distributable datasets, xBD [10], Multi³Net [11], and Sichuan Landslide and Debrisflow [12]. Extensive experiments demonstrate that our method is effective for accurate ESA detection, outperforming many baseline methods.

2 Related Work

2.1 Earth Surface Anomalies Detection and Diagnosis

The increasing frequency of ESA has attracted attention. Current methods primarily utilize satellite imagery for large-scale research, and according to the focus of research, these methods can be divided into post-hoc analysis and rapid response. The former aims to study the process, intensity, impact, and other factors of ESA and conducts comprehensive analysis to aid the authorities in better evaluation, planning and action. Particularly, the release of the large-scale building damage assessment dataset xBD [10] has led to many related studies, such as change detection and multimodal fusion [5,13,14]. The rapid response task focuses on timeliness, striving to obtain detection results as early as possible during or after a disaster [2,15,16]. To this end, EmergencyNet [15] utilizes single UAV images for rapid detection of several classes of ESA. However, it is better suited for on-site applications and challenging to scale up. In [16], the author proposes a lightweight network to rapidly extract landslide areas from

single satellite imagery. In [2], flood areas were extracted with limited resources for rapid response under on-orbit conditions. Nonetheless, these methods are only applicable to a single type of ESA. Given the aforementioned problems, our method aims to rapidly detect ESA using binary classification on a single satellite image.

2.2 Graph Neural Network in Satellite Imagery

GNN can explicitly capture relationships between nodes, edges, and graphs to learn higher-order semantic information, such as semantic relationships and constraints, enabling further exploitation of images. Consequently, some studies incorporate it into satellite imagery, especially hyperspectral image semantic segmentation. CEGCN [17] utilizes superpixels to construct the graph and combines CNN and GNN to achieve accurate land cover classification. In [18,19], various strategies are employed to construct multi-scale graphs for extracting high-order semantic information, resulting in a significant improvement in semantic segmentation. Recently, GNN has also been introduced in satellite imagery classification. H-GCN [20] uses GNN to leverage relationships between objects such as forests and ponds for remote sensing scene classification, while MLRSSC-CNN-GNN [21] combined object detection with GNN for complex scene understanding. By incorporating GNN based on deep visual features, these methods have achieved better performance in complex visual tasks. GNN can simulate the human ability of image interpretation and knowledge formation to some extent, thus facilitating the use of complex satellite imagery for high-level semantic reasoning. Inspired by this, GNN is integrated into our method for ESA detection. To the best of our knowledge, it is the first exploration of GNN in this field, which can provide a reference and benchmark for future research.

3 Our Method

The framework of our method is illustrated in Fig 1. The proposed method consists of two stages: graph generation and earth surface anomalies detection, which will be detailed in the following sections. For an input of high spatial resolution satellite image comprised of RGB bands, the first stage generates hierarchical graphs and assignment matrix through two segmentation branches. Subsequently, a hierarchical GAT is utilized to update the node features and extract graph embeddings for each level. An attention-based feature fusion module then combines them to yield the graph-level feature vector, which is finally processed through a Multi-Layer Perceptron (MLP) for ESA detection.

3.1 Notations

In the first stage, the task is to obtain an attributed, undirected graph $G := (V, E, H)$ from an input single satellite image X . Here, V , E , and H represent nodes, edges, and node features, respectively. Each node $v \in V$ is characterized

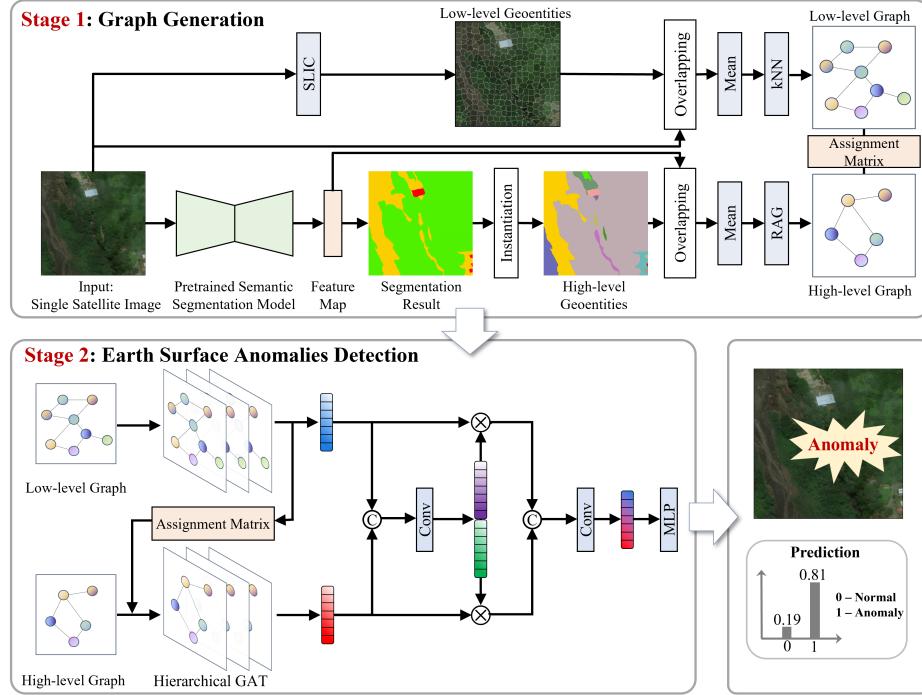


Fig. 1. Overview of the hierarchical geometry-to-Semantic fusion graph neural network.

by a feature vector $h(v) \in \mathbb{R}^d$, thus, $H \in \mathbb{R}^{|V| \times d}$, with d indicating the length of feature vectors and $|\cdot|$ denoting the cardinality of a set. An edge between two nodes $u, v \in V$ is represented by $e_{u,v}$. The graph structure is determined by a symmetric adjacency matrix $A \in \mathbb{R}^{|V| \times |V|}$, where $A_{u,v}$ equals 1 if $e_{u,v} \in E$. The neighborhood of a node $v \in V$ is defined as $N(v) := \{u \in V | e_{u,v} \in E\}$.

3.2 Graph Generation

In order to fully leverage the advantages of graphs and GNN, it is crucial to convert an image into graph. Most existing methods employ the Simple Linear Iterative Clustering (SLIC) [22] algorithm to generate superpixels as nodes, reducing noise and computational efficiency. However, SLIC may only capture a fraction of semantic entities, such as dividing a road into segments, making it challenging to represent higher-level semantic objects and their relationships. To this end, we introduce a semantic segmentation branch that utilizes larger geoentities as nodes to construct a graph. The combination of generated graphs from different levels can enable more comprehensive graph representations.

Low-level Graph. The construction of a low-level graph involves three steps: superpixel segmentation, node feature extraction, and topology configuration, as shown in Fig 1. Firstly, for an input X , we employ SLIC algorithm to obtain a superpixel map S_1 that containing $|V_{low}|$ geoentities serve as nodes. Secondly, by superimposing S_1 onto X , we calculate channel-wise mean values of superpixels, which are used to serve as feature vectors $h(v) \in \mathbb{R}^{d_1}$ for each node. Lastly, we build the initial topology with k-Nearest Neighbor(kNN) algorithm. Meanwhile, we consider the Euclidean distance between superpixels, and the graph is pruned by removing edges longer than the threshold τ_{dist} . Formally, for node v , an edge $e_{v,u}$ is built if

$$u \in \{w | D(v, w) \leq d_k \wedge D(v, w) < \tau_{dist}, \forall w, v \in V_{low}\} \quad (1)$$

where d_k denotes the k -th smallest in the distance matrix $D(v, w)$. Following this process, the low-graph topology is constructed and the low-graph obtained can be formulated as $G_{low} := \{V_{low}, E_{low}, H_{low}\}$.

High-level Graph. To emulate the human brain’s ability to interpret images with high-level semantic understanding, we introduce a semantic segmentation branch. Similar to low-level graph generation, we use a pre-trained semantic segmentation model to obtain the preliminary segmentation map S_2 and feature map F_s . Then, S_2 is instantiated so that each connected component is independent to obtain \hat{S}_2 . Secondly, we overlay \hat{S}_2 onto F_s and calculate the mean values of connected components to obtain a feature vector $h(v) \in \mathbb{R}^{d_2}$ for each node. Finally, we employ Region Adjacency Graph (RAG) algorithm to construct a high-level graph, which can be formulated as $G_{high} := \{V_{high}, E_{high}, H_{high}\}$.

Assignment Matrix. From a semantic perspective, objects in satellite imagery can be considered as hierarchical geoentities, ranging from low-level geometry, such as trees, to high-level semantic, such as forests. Inspired by [23], we utilize an assignment matrix to achieve a jointly representation of hierarchical graph. Intra-level topological structures already captured in the previous graph generation phase, i.e., G_{low} and G_{high} . To exploit the inter-level information, we use an assignment matrix $A_{low \rightarrow high} \in \mathbb{R}^{|V_{low}| \times |V_{high}|}$ to represent the topological relationships between geoentities of different levels. Specifically, for the i^{th} low-level geoentity and the j^{th} high-level geoentity, the corresponding assignment can be formulated as

$$A_{low \rightarrow high}[i, j] = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ low-level geoentity} \in j^{\text{th}} \text{ high-level geoentity} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Finally, each low-level geoentity is assigned to one and only one high-level geoentity. Formally, the final hierarchical graph of each input X can be formulated as $G_{joint} := \{G_{low}, G_{high}, A_{low \rightarrow high}\}$.

3.3 Earth Surface Anomalies Detection

Hierarchical GAT. In the second stage, we aim to leverage G_{joint} to extract high-order semantic information for ESA detection. The relationships among geoentities are often diverse, and assuming that all neighboring nodes make equal contributions during message aggregation is inappropriate. Therefore, we adopt the GAT [24] as backbone. Specifically, the updatation of node features can be formulated as Eq.(3) and Eq.(4):

$$h_i^{(l)} = \sigma \left(\sum_{j \in N_i} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l-1)} + b^{(l)} \right), \quad (3)$$

$$\alpha_{ij}^{(l)} = \frac{\exp \left(\text{LeakyReLU} \left(\mathbf{a}^{(l)T} \left[W^{(l)} h_i^{(l-1)} || W^{(l)} h_j^{(l-1)} \right] \right) \right)}{\sum_{k \in N_i} \exp \left(\text{LeakyReLU} \left(\mathbf{a}^{(l)T} \left[W^{(l)} h_i^{(l-1)} || W^{(l)} h_k^{(l-1)} \right] \right) \right)} \quad (4)$$

where $h_i^{(l)}$ denotes the feature vector of node v_i in the l^{th} layer, $\sigma(\cdot)$ is the sigmoid function, N_i represents some neighborhood of node v_i , $\alpha_{ij}^{(l)}$ is the attention coefficient, $W^{(l)}$ and $b^{(l)}$ are the weight matrix and bias term of the l^{th} layer, $\mathbf{a}^{(l)}$ is the learnable parameter vector, and $||$ denotes the concatenation operation.

Given the varying lengths of node features ($d_1 \neq d_2$), we employ two GATs with different input layers, namely $f_{\text{GAT}}^{\text{low}}$ and $f_{\text{GAT}}^{\text{high}}$, to update node features. Following the $\hat{G}_{low} = f_{\text{GAT}}^{\text{low}}(G_{low})$, the low-level node embeddings, $\hat{h}_{low}(v) | v \in V_{low}$, and $A_{low \rightarrow high}$ are used to initialize the node features of high-level graph, which can complement the loss of low-level geometric information caused by the large areas and mean operations of high-level geoentities, i.e.,

$$h_{\text{high}}(w) = \left[H_{\text{high}}(w) || \sum_{v \in M(w)} \hat{h}_{low}(v) \right] \quad (5)$$

where $M(w) := v \in V_{low} | A_{low \rightarrow high}(v, w) = 1$ is the set of nodes in G_{low} mapping to a node $w \in V_{high}$.

After that, G_{high} is likewise processed by $f_{\text{GAT}}^{\text{high}}$. Finally, readout operation and MLP are utilized to extract feature vectors \mathbf{x}_{low} and \mathbf{x}_{high} of different levels from \hat{G}_{low} and \hat{G}_{high} , where $\mathbf{x}_{low}, \mathbf{x}_{high} \in \mathbb{R}^{d_3}$.

Attention-based Feature Fusion. We propose an attention-based feature fusion module to adaptively merge \mathbf{x}_{low} and \mathbf{x}_{high} into a feature vector \mathbf{z} , which serves as a representation of input satellite imagery X . In the end, \mathbf{z} will be fed into an MLP for ESA detection. Formally in Eq.(6)

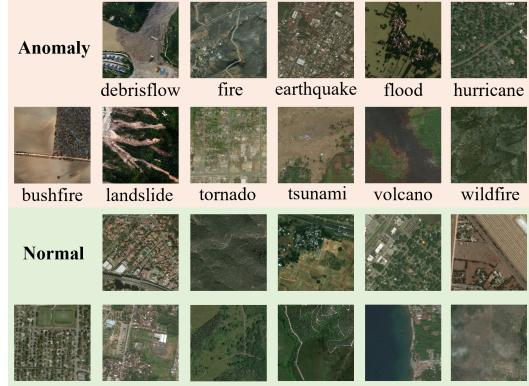
$$\mathbf{z} = f \left([x_{low} \odot a_{low} || a_{high} \odot a_{high}] \right) \quad (6)$$

$$[a_{low}, a_{high}] = [\sigma(f([x_{low}, x_{high}]), 1 - \sigma(f([x_{low}, x_{high}]))] \quad (7)$$

where a_{low} and a_{high} are attention weights calculated according to Eq.(7), and \odot denotes element-wise product. The function $f(\cdot)$ represents a convolutional layer that maps a vector of length d_3 to m .

Table 1. Statistics of ESAD

Class	Num
Anomaly	Flood
	Landslide
	Debrisflow
	Hurricane
	Wildfire
	Earthquake
	Volcano
	Tornado
	Tsunami
	Fire
Bushfire	
Normal	
Normal	

**Fig. 2.** Examples of ESAD Dataset

4 Experiments

4.1 Dataset

The existing ESA datasets mainly focus on single-class detection or post-hoc analysis, which is insufficient for our research needs. We propose ESAD, a composite dataset to fill the gap of large-scale multi-class datasets. Specifically, ESAD is composed of three publicly datasets: xBD, Multi³Net and Sichuan Landslide and Debrisflow. xBD comprises paired pre- and post-disaster images and is currently the largest dataset for building damage assessment. Multi³Net includes post-disaster high spatial resolution images after Hurricane Harvey, while Sichuan Landslide and Debrisflow contains images of four geohazards. The ESAD contains 13,058 ESA samples classified into 11 image-level classes as shown in Fig 2. ESAD is integrated by carefully filtering through visual interpretation. Additionally, all the samples are preprocessed to have a resolution ranging from 0.2m to 1m and a size of 1024×1024 . The detailed statistics of ESAD is illustrated in Table 1.

4.2 Baseline Methods and Implementation Details

To verify the performance of the proposed method, we use ResNet-50, MobileNetV3 and ViT-B/32 as visual feaure-based baseline methods. As for graph classification, HGP-SL [25] is one of the state-of-the-art method, while HACT-Net [23] is an advanced method in medical image analysis that also uses hierarchical graph representations. Moreover, to verify the effectiveness of two branches, GAT-Low and GAT-High only use feature vector generated by their respective branches. We evaluate the attention-based feature fusion module using Concat-GAT, which only uses concatenation to obtain the graph-level embedding.

We divided the dataset into training, testing, and validation sets in a 6:2:2 ratio. The semantic segmentation branch uses HRNet pre-trained on the DeepGlobe Land Cover dataset. Each method is trained from scratch for 200 epochs,

Table 2. Quantitative result of comparison methods

Method	OA	Recall	AIT	Params
ResNet-50	91.05	90.42	16.63ms	25.61M
MobileNetV3	88.40	88.08	14.98ms	3.8M
ViT-B/32	93.71	93.40	16.67ms	88.21M
HGP-SL-Low	66.85	66.87	2.04ms	0.07M
HGP-SL-High	61.64	61.58	0.28ms	0.14M
HAUT-Net	74.53	75.42	2.47ms	0.79M
GAT-Low	65.15	66.80	2.88ms	0.09M
GAT-High	62.32	68.06	3.42ms	0.21M
Concat-GAT	78.41	77.25	6.92ms	1.02M
Our method	83.89	83.86	6.04ms	1.01M

with an initial learning rate of 0.0001 and Adam optimizer. The loss function is set to BCELoss. All experiments are conducted using the PyTorch framework in Ubuntu 18.04 and trained on a NVIDIA 3090 GPU with 24GB memory.

4.3 Experimental Results

The quantitative results of comparison methods are illustrated in Table 2, where Overall Accuracy (OA), Recall, Average Inference Time (AIT) and Parameters are used to evaluate the comprehensive performance. As shown in Table 2, the visual feature-based methods achieve the best accuracy, with only our framework from the GNN-based methods is competitive. Notably, GNN-based methods generally have fewer parameters and faster inference speed, even outperforming MobileNetV3, which is specifically designed for limited resource conditions. In comparison to advanced GNN based methods, our method achieved significant performance improvements with similar parameters, with accuracy and recall increasing by 9.36% and 8.44%, respectively, without a significant increase in the number of parameters. By fusing features of different levels, the proposed method achieved better performance compared to GAT-Low and GAT-High. Additionally, our attention-based feature fusion module better incorporates hierarchical information than Concat-GAT, resulting in accuracy and recall improvements of 5.48% and 6.61%. Overall, our method strikes a good balance between accuracy and efficiency, making it more suitable for ESA detection while saving valuable time and resources for downstream tasks.

5 Conclusion

We propose a framework called the Hierarchical Geometry-to-Semantic Fusion Graph Neural Network for ESA detection. The proposed framework is designed based on the human brain’s ability in comprehending and interpreting images. It leverages GNN to learn high-order semantic information from the satellite

imagery. To fill the gap of a lack of benchmark datasets, we created the ESAD dataset based on existing related datasets. Extensive experiments demonstrate that our method achieves a good balance between accuracy and efficiency, which is more suitable for ESA detection with high timeliness requirements. In future work, we will further explore brain-inspired models for better performance and extend our method to on-orbit real-time ESA detection task.

Acknowledgements. This research is partially supported by the National Natural Science Foundation of China Major Program (Grant No. 42192580, 42192583), Hubei Province Natural Science Foundation (Grant No. 2021CFA088 and 2020CFA003), the Science and Technology Major Project (Grant No. 2021-AAA010), and Wuhan University - Huawei Geoinformatics Innovation Laboratory. The numerical calculations in this paper had been supported by the super-computing system in the Supercomputing Center of Wuhan University.

References

1. Qiao, W.: Research framework of remote sensing monitoring and real-time diagnosis of earth surface anomalies. *Acta Geodaetica et Cartographica Sinica* **51**(7), 1141 (2022)
2. Mateo-Garcia, G., Oprea, S., Smith, L., Veitch-Michaelis, J., Baydin, A.G., Backes, D., Gal, Y., Schumann, G.: Flood detection on low cost orbital hardware. Artificial Intelligence for Humanitarian Assistance and Disaster Response Workshop at NeurIPS (2019)
3. Weber, E., Papadopoulos, D.P., Lapedriza, A., Ofli, F., Imran, M., Torralba, A.: Incidents1m: a large-scale dataset of images with natural disasters, damage, and incidents. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022)
4. Rui, X., Cao, Y., Yuan, X., Kang, Y., Song, W.: Disastergan: Generative adversarial networks for remote sensing disaster image generation. *Remote Sensing* **13**(21), 4284 (2021)
5. Weber, E., Kané, H.: Building disaster damage assessment in satellite imagery with multi-temporal fusion. *arXiv preprint arXiv:2004.05525* (2020)
6. Saha, S., Shahzad, M., Ebel, P., Zhu, X.X.: Supervised change detection using prechange optical-sar and postchange sar data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **15**, 8170–8178 (2022)
7. Tilon, S., Nex, F., Kerle, N., Vosselman, G.: Post-disaster building damage detection from earth observation imagery using unsupervised and transferable anomaly detecting generative adversarial networks. *Remote sensing* **12**(24), 4193 (2020)
8. Liu, Q., Klucik, R., Chen, C., Grant, G., Gallagher, D., Lv, Q., Shang, L.: Unsupervised detection of contextual anomaly in remotely sensed data. *Remote Sensing of Environment* **202**, 75–87 (2017)
9. Ding, C., Pang, G., Shen, C.: Catching both gray and black swans: Open-set supervised anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7388–7398 (2022)
10. Gupta, R., Goodman, B., Patel, N., Hosfelt, R., Sajeev, S., Heim, E., Doshi, J., Lucas, K., Choset, H., Gaston, M.: Creating xbd: A dataset for assessing building damage from satellite imagery. In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition Workshops. pp. 10–17 (2019)

11. Rudner, T.G., Rußwurm, M., Fil, J., Pelich, R., Bischke, B., Kopačková, V., Bilinski, P.: Multi3net: segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 702–709 (2019)
12. Zeng, C., Cao, Z., Su, F., Zeng, Z., Changxi, Y.: A dataset of high-precision aerial imagery and interpretation of landslide and debris flow disaster in sichuan and surrounding areas between 2008 and 2020. *China Scientific Data* **7**(2) (2022)
13. Xu, J.Z., Lu, W., Li, Z., Khaitan, P., Zaytseva, V.: Building damage detection in satellite imagery using convolutional neural networks. arXiv preprint arXiv:1910.06444 (2019)
14. Lee, J., Xu, J.Z., Sohn, K., Lu, W., Berthelot, D., Gur, I., Khaitan, P., Koupparis, K., Kowatsch, B., et al.: Assessing post-disaster damage from satellite imagery using semi-supervised learning techniques. Artificial Intelligence for Humanitarian Assistance and Disaster Response Workshop at NeurIPS (2020)
15. Kyrkou, C., Theocharides, T.: Emergencynet: Efficient aerial image classification for drone-based emergency monitoring using atrous convolutional feature fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **13**, 1687–1699 (2020)
16. Niu, C., Gao, O., Lu, W., Liu, W., Lai, T.: Reg-sa-unet++: A lightweight landslide detection network based on single-temporal images captured postlandslide. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **15**, 9746–9759 (2022)
17. Liu, Q., Xiao, L., Yang, J., Wei, Z.: Cnn-enhanced graph convolutional network with pixel-and superpixel-level feature fusion for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **59**(10), 8657–8671 (2020)
18. Jiang, J., Ma, J., Liu, X.: Multilayer spectral–spatial graphs for label noisy robust hyperspectral image classification. *IEEE Transactions on Neural Networks and Learning Systems* **33**(2), 839–852 (2020)
19. Xi, B., Li, J., Li, Y., Song, R., Xiao, Y., Du, Q., Chanussot, J.: Semisupervised cross-scale graph prototypical network for hyperspectral image classification. *IEEE Transactions on Neural Networks and Learning Systems* (2022)
20. Gao, Y., Shi, J., Li, J., Wang, R.: Remote sensing scene classification based on high-order graph convolutional network. *European Journal of Remote Sensing* **54**(sup1), 141–155 (2021)
21. Liang, J., Deng, Y., Zeng, D.: A deep neural network combined cnn and gcn for remote sensing scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **13**, 4325–4338 (2020)
22. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(11), 2274–2282 (2012)
23. Pati, P., Jaume, G., Foncubierta-Rodriguez, A., Feroce, F., Anniciello, A.M., Scognamiglio, G., Brancati, N., Fiche, M., Dubruc, E., Riccio, D., et al.: Hierarchical graph representations in digital pathology. *Medical Image Analysis* **75**, 102264 (2022)
24. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
25. Zhang, Z., Bu, J., Ester, M., Zhang, J., Yao, C., Yu, Z., Wang, C.: Hierarchical graph pooling with structure learning. arXiv preprint arXiv:1911.05954 (2019)