ISPRS Journal of Photogrammetry and Remote Sensing 227 (2025) 99–124

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

SAGRNet: A novel object-based graph convolutional neural network for diverse vegetation cover classification in remotely-sensed imagery

Baoling Gui [a], Lydia Sam [a,*], Anshuman Bhardwaj [a], Diego Soto Gómez [b], Félix González Peñaloza [c], Manfred F. Buchroithner [d], David R. Green [a]

[a] School of Geosciences, University of Aberdeen, King's College, Aberdeen AB24 3UE, UK
[b] International Research Center in Critical Raw Materials for Advanced Industrial Technologies (ICCRAM), University of Burgos, Centro de I+D+I, Plaza Misael Bañuelos s/n, 09001 Burgos, Spain
[c] Evenor-Tech, Avda. República Argentina 27B A6, 41011 Sevilla, Spain
[d] Technische Universität Dresden 01069 Dresden, Germany

ARTICLE INFO

Keywords:
Object-based classification
Graph convolutional
Vegetation mapping
Deep learning
Remote sensing

ABSTRACT

Growing global population, changing climate, and shrinking land resources demand for quicker, efficient, and more accurate methods of mapping and monitoring vegetation cover in remote sensing datasets. Many deep learning-based methods have been widely applied for semantic segmentation tasks in remote sensing images of vegetated environments. However, most existing models are pixel-based, which introduces challenges such as high time consumption, cumbersome implementation, and limited scalability. This paper presents the SAGRNet model, a Graph Convolutional Neural Network (GCN) that incorporates sampling aggregation and self-attention mechanisms, while leveraging the ResNet residual network structure. A key innovation of SAGRNet is its ability to fuse features extracted through diverse algorithms, enabling comprehensive representation and enhanced classification performance. The SAGRNet model demonstrates superior performance over leading pixel-based neural networks, such as U-Net++ and DeepLabV3, in terms of both time efficiency and accuracy in vegetation image classification tasks. We achieved an overall mapping accuracy of ~90 % using SAGRNet, compared to ~87% and ~85% from U-Net++ and DeepLabV3, respectively. Additionally, it offers more convenience in data processing. Furthermore, the model significantly outperforms cutting-edge graph-based convolutional networks, including Graph U-Net (achieved overall accuracy ~65%) and TGNN (achieved overall accuracy ~75%), showcasing exceptional generalization capability and classification accuracy. This paper provides a comprehensive analysis of the various processing aspects of this object-based GCN for vegetation mapping and emphasizes its significant potential for practical use. The model's versatility can also be expanded to other image processing domains, offering unprecedented possibilities of information extraction from satellite imagery. The code for practical application experiment is available at https://github.com/baoling123/GCN-remote-sensing-classification.git.

1. Introduction

Fast changing global climate, increasing extreme weather events, and highest ever pollution levels are impacting vegetation cover. Moreover, the fast-growing global population is putting additional pressure on land resources and food security, necessitating quick and efficient monitoring of agricultural and other vegetated regions, especially within a diverse land cover setting. Effective vegetation monitoring and planning are crucial to ensuring sustainable land use and food security (Su and Zhang, 2023; Wang et al., 2022). Global deforestation and food scarcity are escalating issues. The Food and Agriculture Organisation of the United Nations (FAO) Global Forest Resources Assessment 2020 reported a net loss of 178 million hectares of forest area worldwide from 1990 to 2020, primarily attributed to deforestation and land use change (2020). The UN World Food Programme indicates that around 282 million individuals in 59 nations and territories are experiencing acute hunger in 2023, marking a rise of 24 million from the prior year (2024). These statistics indicate the ongoing deterioration of

* Corresponding author.
E-mail address: lydia.sam@abdn.ac.uk (L. Sam).

https://doi.org/10.1016/j.isprsjprs.2025.06.004
Received 12 February 2025; Received in revised form 19 May 2025; Accepted 5 June 2025
Available online 12 June 2025

global forest resources and increasing food insecurity, underscoring the necessity for immediate measures to safeguard forests and enhance food security. Satellite remote sensing plays a key role in vegetation monitoring, offering high spatial, spectral, and temporal resolutions, which provide robust data support for rapid, large-scale information acquisition and land use modelling (Campos-Taberner et al., 2023; Gui et al., 2024a). Vegetation cover monitoring is particularly vital for understanding crop supply, health, and their interactions with surrounding ecosystems (Huang et al., 2024). These data form the foundation and key resource for analysing the current state of vegetation, with many studies and analyses relying on vegetation classification maps (Kganyago et al., 2024). Therefore, the rapid and efficient generation of accurate vegetation land use classification maps has become a critical focus in the field of remote sensing assisted agronomy.

With the advancement of machine learning and deep learning algorithms, a wide array of techniques has been employed for vegetation mapping. Early approaches, such as traditional pixel-based classification methods, primarily relied on individual spectral information and classification algorithms to process remote sensing images. To enhance classification accuracy, researchers have progressively improved these methods by integrating more sophisticated algorithms, such as Random Forest (Ok et al., 2012), SVM (Pereira et al., 2022), XGBoost (Saini and Ghosh, 2021), and K-means (Javidan et al., 2023), all of which have been widely recognised for their effectiveness in vegetation mapping. However, as very high spatial-resolution imagery has become more prevalent, the limitations of relying solely on individual spectral information have become evident. Modern applications increasingly demand methods that not only analyse spectral data but also incorporate surrounding texture and contextual information, recognising that most land use types cannot be fully captured by a single spectrum alone, especially if captured at high spatial resolutions. This gap is partially bridged by object-based classification methods, which emphasise the texture of elements and land cover at boundaries. Unlike pixel-based approaches, object-based classification aggregates units with similar textures into irregular objects (Dingle Robertson and King, 2011; Duro et al., 2012). These aggregated objects are then classified using algorithms like Random Forest and SVM (Li et al., 2016; Peña et al., 2014). However, traditional object-based classifiers have certain drawbacks: the segmented objects can be too large or too constrained, and the performance and computing requirements of the classifier heavily influence overall classification accuracy (Ma et al., 2017). As a result, object-based image classification is yet to experience widespread development and usability.

Conversely, pixel-based deep learning image classification algorithms have been widely applied across various domains (Gui et al., 2024c; Liu et al., 2024; Xu et al., 2024),. Numerous convolutional neural network (CNN) models have been developed and refined for vegetation mapping. For example, the U-Net segmentation network has been used for agricultural shed segmentation (Wang et al., 2023), while Khan et al. (2023) utilised a long-short-term memory network (LSTM) to extract changing features of agricultural crops for crop mapping. Additionally, Alcantara et al. (2012) applied statistical texture learning in a pyramid scene parsing network to monitor abandoned suburban farmland. Other deep learning networks, such as the Deeplab series (Gui et al., 2024c; Niu et al., 2019), YOLO (Ajayi et al., 2023), and Faster R-CNN (Quan et al., 2019), have also been employed, achieving improved classification results for vegetation cover in various contexts. The self-learning capability of deep learning algorithms presents numerous opportunities for feature classification and land cover extraction. Their strong feature extraction and nonlinear classification capabilities make them powerful tools for vegetation mapping, particularly when we need differentiating among various vegetation classes such as woodlands, grasslands, and arable lands. However, these strengths also come with significant computational demands. From data preprocessing, segmentation model selection, and parameter tuning to final prediction and the production of large-scale land use classification maps, deep learning requires stringent processes and standards (Alzubaidi et al., 2021). Furthermore, deep learning models demand high-quality training samples, with strict requirements for label accuracy, training set form, size, and format (Gui et al., 2024b). Even minor errors can result in significant training inaccuracies. Moreover, pixel-based deep learning algorithms often function as a “black box,” where feature computation is highly complex, and many irrelevant features are extracted for computation and classification (Dobson, 2023). The single-cell scale computation also leads to lengthy training and prediction times (Li et al., 2024). These limitations hinder the full potential of deep learning algorithms for practical vegetation classification applications.

To overcome these limitations, there has been a growing focus on applying object-based deep learning in remote sensing image classification. Object-based classification targets larger entities formed through clustering and segmentation, which significantly reduces data volume compared to pixel-based methods (Blaschke, 2010). This approach is particularly practical for vegetation land, where well-defined boundaries and clear textures facilitate object-based classification. To address the shortcomings of previous classifiers, researchers have enhanced the nonlinear classification capabilities by incorporating deep neural networks, thereby improving stability and accuracy. For instance, Su and Zhang (2023) introduced a multi-scale CNN and transformer for small-scale crop classification. Pan et al. (2021) developed a deep semantic segmentation neural network, tested on standard datasets, which demonstrated improved stability and accuracy. Zhang et al. (2018) utilised convolutional neural networks with spatial correlation attributes between objects to achieve high-resolution urban land cover mapping. While these studies have mitigated the limitations of earlier object-based approaches by improving classifier performance, they still fall short of pixel-based deep learning in terms of incorporating neighbourhood information into the analysis. Moreover, the irregular shapes of object-based segments pose challenges for feature extraction using convolutional computation, which is more straightforward in pixel-based deep learning.

To further address this issue, the advent of Graph Convolutional Neural Network (GCN) offers a novel solution for object-based deep learning. Unlike traditional convolutional neural networks, which typically operate on relatively regular graph structures and extract node features using convolutional kernels, GCN are designed to handle irregular topological networks, such as social networks. GCN can accommodate varying numbers of nodes, each with different neighbours, making it difficult for traditional CNN to process such structures. By enabling neighbourhood analysis through node counting in irregularly segmented objects, GCN effectively solve this challenge. Subsequent research has increasingly focused on object-based GCN. For example, several recent studies have made significant strides in this area. Zhang and Wang (2024) proposed a progressive feature fusion framework based on GCN, applying it to standard datasets for remote sensing scene classification with impressive results. Similarly, Yang et al. (2024) designed an adaptive multi-frequency graph feature learning module to extract multiscale and neighbourhood information in GCNs, which was also tested and validated on standard datasets, demonstrating the advantages of their model. Additionally, Song et al. (2024) achieved significant performance gains by constructing a hybrid skeletal network that exchanges and fuses information through multi-layer interactions between CNN and GNN branches, leading to notable model improvements. Chen et al.(2024) further enhanced GCN-based image classification by introducing residuals and multi-scale feature learning for hyperspectral image classification, achieving better results on three benchmark datasets. Despite these advances, the existing models have primarily been tested on standard test sets covering small areas, leaving questions about their suitability and accuracy in practical applications. Moreover, current research has not sufficiently addressed the early-stage calculation of node feature attributes for the original segmented objects. Fully integrating the advantages of pixel-based deep learning in object-based feature extraction remains a challenge that warrants further

exploration.

To address these issues, our study introduces a novel object-based graph convolution neural network model that aims to make the following contributions:

1. Aggregate important feature attribute computation methods and reduce data redundancy by selecting features of higher importance through SHapley Additive exPlanations (SHAP) analysis.

2. Construct a lightweight graph convolutional neural network (SAGRNet) to achieve semantic segmentation of remote sensing images for vegetation mapping with less computational elimination cost, high accuracy and strong generalisation ability.

3. Enhance the representation of spatial and contextual relationships by integrating multi-scale feature fusion and attention mechanisms, allowing the model to effectively capture intricate patterns and interactions in heterogeneous vegetation landscapes, thus improving classification performance even in challenging scenarios.

4. Promote the practical application of object-based graph convolutional network model in image classification.

The remainder of the paper is organized as follows. In Section 2, we discuss the limitations of traditional object-based classification methods and their implications for vegetation cover mapping and emphasises the necessity and effectiveness of preliminary feature calculation. Section 3 introduces the proposed SAGRNet model, detailing its architecture, including the integration of graph convolution, multi-scale feature fusion, and attention mechanisms. And focuses on the experimental setup, including data preprocessing, graph construction, model training and evaluation, and adaptation testing of the model in different landscape contexts around the world. Section 4 presents the experimental results, comparing SAGRNet's performance with state-of-the-art pixel-based and object-based classification models, comparison and verification test results. Furthermore, it demonstrates the classification ability of the algorithm proposed in this paper in different global landscape backgrounds. In the following Section 5, we provide a detailed discussion of the results of this paper, emphasising the scalability, generalisability and sensitivity of the algorithms in this paper for practical application in different contexts, along with a discussion of its strengths and limitations. Finally, Section 6 concludes the paper by summarizing the contributions of this study and proposing potential future research directions in vegetation cover classification using advanced object-based deep learning frameworks.

2. Study area and data

2.1. Study area

To facilitate ample result validation, our study focuses on diversly vegetated areas comprising woodlands, farmlands, and grasslands situated adjacent to other landcover classes (e.g., built-up and water), located in the northeast of Scotland, UK (as shown in Fig. 1). The vegetated environment in northern Scotland is significantly shaped by

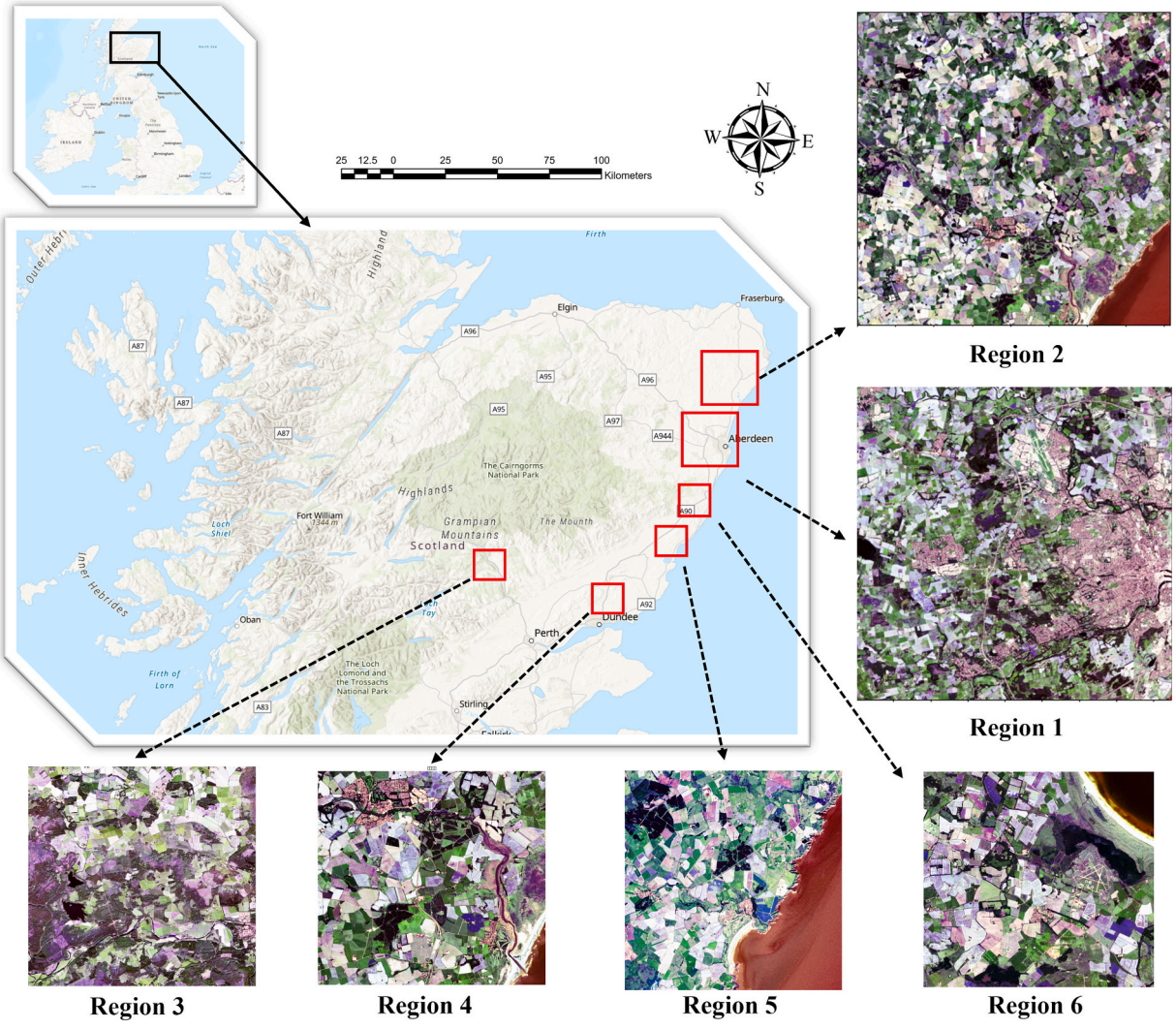

Fig. 1. Schematic map of the study area.

its topography, climate, and history. The region is predominantly hilly and mountainous, characterized by a cold and rainy climate, which limits the extent of arable land. Pastureland, suitable for cattle and sheep farming, dominates the landscape. Despite the limited arable land, some cereals and vegetables are cultivated in the lowlands and plains, and it can be challenging to differentiate these farmlands from the surrounding vegetation cover using traditional image classifiers. Grasslands are widely used for sheep and upland cattle grazing, while woodlands are found in the upland and hilly areas.

For this study, we selected six study areas, representing varying vegetation cover, for testing and validation. Two of these areas were used to test the algorithm of this project, to test the classification effect and operational efficiency of the algorithm in different environmental backgrounds. To capture the different vegetation layouts, the coverage of these two areas varies (Region 1 and 2). One study area emphasizes a single vegetation cover, with predominant land uses including woodland, grassland, and arable land. The other area covers a larger portion of the urban environment, adding complexity to the imagery. Urban environments introduce various uncontrollable factors that challenge image segmentation, fully testing the adaptability of the model proposed in this study to complex environments. In addition, in order to verify the generalisation ability of the algorithm in this paper, we selected another four research areas for algorithm verification to test whether the algorithm has sufficient ability for promotion and application (Region 3 to 6). The classification types in this study include grassland, arable land, woodland, water bodies and urban areas.

2.2. Data

We used Sentinel-2 Level-2A products with less than 8% cloud coverage as our raw data, and these images were both geometrically and atmospherically corrected. Given the significant relative spectral difference between grassland and arable land in summer, we acquired satellite images from 23 September 2022. Six bands—Band 2, 3, 4, 5, 8, and 12—were selected for analysis. The selection of bands 2, 3, 4, 5, 8, and 12 is predicated on their efficacy and pertinence in vegetation analysis. These bands encompass the visible, near-infrared, and shortwave infrared areas, enabling the capture of spectral features of vegetation, particularly in evaluating plant health and differentiating among various vegetation species. Choosing these essential bands enhances classification accuracy and diminishes data redundancy, thereby lowering computing complexity. This method has demonstrated efficacy in investigations concerning remote sensing vegetation categorisation (Abbaszad et al., 2024). For the labels, we utilized results from the UK Centre for Ecology and Hydrology (UKCEH) (UK Centre for Ecology & Hydrology (ceh.ac.uk)), which provided 10 m resolution land cover maps of the UK with an overall accuracy exceeding 90 %.

3. Methodology

To comprehensively analyse the classification potential of object-based GCN for vegetated land, this research is organised into three key directions. First, we thoroughly evaluate the capabilities of traditional object-based classification algorithms for vegetation cover, focussing on the importance of different feature attributes in individual objects involved in classification. This analysis will provide a foundation for feature selection in object-based GCNs and help reduce time-consuming processes and overfitting errors caused by the inclusion of excessive feature attributes. Second, we detail the fusion of the ResNet residual model with a node-connected attention mechanism in the proposed developed novel graph convolutional neural network model (SAGRNet), which is further tested for a diverse vegetation setting in north-eastern Scotland, UK. Finally, we will compare the SAGRNet model with several mainstream pixel-based and object-based neural network models to identify the strengths and limitations of the proposed approach. In addition, to evaluate the applicability of the model proposed in this paper in different regions and climates, five regions around the world were selected for testing and analysis to verify the model's generalisation ability in different landscapes. In this study, the term 'vegetation' refers broadly to semi-natural vegetation, such as forests and grasslands, and farm crops. While these categories share common green spectral characteristics, they differ significantly in their visual appearance, management practices, and spectral variability. These distinctions are addressed in the feature selection and classification steps of our methodology, aligning with LU/LC classification principles.

3.1. Object-based analysis of vegetation classification

This section presents a comprehensive assessment of classic object-based classification methods in vegetation cover classification, emphasising the significance of various feature properties in individual objects. This stage is crucial to the whole workflow, primarily aimed at establishing a dependable foundation for feature selection in following object-based GCN models. In actual classification problems, several characteristics contribute unequally to classification accuracy. Uncritically employing all feature qualities can result in the inefficient use of computational resources, heightened model complexity, and an augmented danger of overfitting.

Object-based image classification is an image analysis method that decomposes an image into multiple objects or regions with similar attributes through image segmentation and then classifies these objects. The process begins with image segmentation, where the image is divided into regions sharing similar spectral, texture, or shape features. Next, features are extracted from each segmented object, such as spectral, texture, and shape features. Classification is then performed based on these extracted features using methods like decision trees, support vector machines, and others. Finally, post-processing, such as merging or smoothing, is applied to the classification results to produce the final classification map (Blaschke, 2010).

This part of our workflow begins with the initial segmentation of the image using the Felzenszwalb algorithm, a classical method proposed by Felzenszwalb and Huttenlocher (2004a) for image segmentation. The algorithm is designed to quickly and efficiently divide an image into a series of interconnected regions, or "hyperpixels," that are visually similar in colour, texture, or luminance characteristics. Felzenszwalb's algorithm performs segmentation based on graph theory. The core of Felzenszwalb is to determine whether to merge two regions (components) C_1 and C_2 through a region merging strategy. The merging conditions are as follows:

$$If\,\omega(e) \le \min(Int(C_1)+\tau(C_1), Int(C_2)+\tau(C_2)), then merge C_1 \text{ and } C_2 \quad (1)$$

where $\omega(e)$ is the minimum edge weight connecting two regions. $Int(C)$ indicates the internal variation within region C, defined as the maximum edge weight (i.e., maximum internal variation) of that region:

$$Int(C) = \max\omega(e), \epsilon \in MST(C) \quad (2)$$

where MST stands for minimum spanning tree. $\tau(C)$ denotes the sensitivity threshold term of the region, which is used to control the merging scale, and is defined as:

$$\tau(C) = \frac{k}{|C|} \quad (3)$$

where k is the hyperparameter that controls the overall segmentation scale (the larger the value, the coarser the segmentation); $|C|$ indicates the number of pixels in the region (region size).

This algorithm represents an image as an undirected weighted graph, where the nodes correspond to pixels, the edges represent connections between neighbouring pixels, and the edge weights reflect the similarity between these pixels, usually based on differences in colour or

luminance. Initially, the algorithm constructs a graph where each pixel is a node and the similarity between pixels, computed from pixel value differences, determines the weight of the edges. The algorithm then applies a minimum spanning tree approach to group pixels, assigning those with high similarity to the same region. The construction of the minimum spanning tree ensures that each edge added to the tree has the smallest weight, indicating that the connected pixel pairs are visually most similar. Once the spanning tree is constructed, the algorithm determines whether certain connections should be cut based on a threshold, which is typically related to the overall size of the image and the differences in edge weights within a region. This threshold controls the degree of similarity within regions and the distinction between regions. To prevent over-segmentation, the algorithm may also merge small, segmented regions by checking region size and edge weights. The Felzenszwalb algorithm is widely used in computer vision tasks for image segmentation due to its speed, ease of implementation, and low sensitivity to parameters.

We implemented the algorithm in Python by directly calling the existing Felzenszwalb algorithm (Felzenszwalb and Huttenlocher, 2004b). It is important to note that this algorithm's segmentation process requires fine-tuning two key parameters. The 'scale' parameter controls the similarity measure between different regions of the image during segmentation, corresponding to the k value in Equation (3), while 'min_size' parameter sets the minimum allowable size for the segmented regions. These parameters must be carefully analysed and tested based on the specific requirements of the task. To ensure test accuracy and balance between precision and data volume, the 'scale' parameter was set primarily based on the size of the largest farmland plot, ultimately being set to 100 to ensure that relatively uniform plots are well segmented. Additionally, to ensure that smaller, fragmented feature types are also accurately classified, the 'min_size' parameter was adjusted to its relatively lowest possible value, i.e., 5. After completing the segmentation, it is essential to calculate the potential feature attribute values for each segmented object and convert the key attributes into metrics that can be used for classification or analysis. This allows for more effective differentiation and identification of various object classes. To leverage existing feature attribute calculation methods, we summarised most of these methods and thoroughly analysed them from six perspectives: spectral features, geometric features, colour features, texture features, shape features, and frequency-domain features. The description of each feature attribute and its calculation method is detailed in Table 1. Considering that some of the features were calculated for a single band, the total number of feature dimensions involved in this study was 87. We first performed correlation analyses for these 87 dimensions using Pearson correlation coefficients, and feature attributes with correlation coefficients greater than 0.9 were subsequently eliminated. We further analysed their importance in each classification type by applying the SHapley Additive exPlanations (SHAP) algorithm through random sampling.

SHAP is a method for interpreting machine learning models, rooted in cooperative game theory (Du et al., 2024). It determines the significance of each feature by calculating its incremental contribution to the model's predicted result. At its core, SHAP uses Shapley values to allocate the influence of each feature on the final prediction across various feature combinations, ensuring this allocation is fair and consistent. SHAP offers several key advantages. First, it is model agnostic, meaning it can be applied to a wide range of machine learning models. Second, it provides explanations for specific predictions, offering a local interpretation for individual samples. Additionally, SHAP can perform global feature importance analysis, giving insights into the overall importance of features across the entire model. Moreover, it effectively captures the combined impacts in intricate models by considering the interplay of features, which is especially crucial for comprehending features with many dimensions or models that exhibit non-linear behaviour (Shetty et al., 2024). We then selected the top five most important attributes as the standard feature attribute set for prediction, which were used in object-based classification and prediction. For classification, we selected the XGBoost algorithm (Chen and Guestrin, 2016) as our classifier using feature attribute values for classification. The dataset was split into training and test sets, with a 7:3 ratio.

3.2. Object-based implementation of deep GCN models

This section systematically delineates the implementation method of an object-based graph convolutional neural network (GCN) model, encompassing graph building, graph structure specification, and the phases of model training, testing, and validation. Our objective is to transform the object properties of remote sensing photos into a graph structure compatible with GCN processing, therefore enabling the model to effectively capture the spatial and semantic connections among objects. The document delineates the underlying architecture of the proposed SAGRNet method, which not only augments the comprehensiveness of feature representation but also markedly increases the model's generalisation capability and precision in intricate vegetation cover classification tasks. Furthermore, during the testing and verification process, we thoroughly assess the model's performance across various situations, offering a solid foundation for model optimisation and practical implementation. Detailed description of each module provided below.

3.2.1. Graphical construction

A GCN is a deep learning model particularly suited for processing non-Euclidean spatial data, such as feature graph structures in remote sensing images. In the context of remote sensing image classification, GCN enables the model to effectively capture spatial features and their contextual information by treating objects as nodes in a graph and constructing the graph structure based on their spatial proximity. The structure of GCN shares many similarities with CNN, including graph convolutional layers, activation functions, and finally a fully connected or classification layer. GCN gradually fuses and enhances local to global features through layer-by-layer convolution. However, during the initial training phase, GCN requires the construction of a graph for training purposes. This involves building the features of irregular objects and the connections between different objects to form a matrix that can be processed by the GCN network. The graph construction for an object-based GCN primarily includes the following components, as illustrated in Fig. 2:

(1) Nodes: In remote sensing images, each node typically represents an object or region, such as feature units obtained through segmentation algorithms. The features of a node can include the spectral, geometric, and texture characteristics of the object.

(2) Edges: Edges represent the relationships or connections between nodes. During graph construction, edges are usually determined based on spatial proximity (e.g., Euclidean distance between nodes). Neighbouring pairs of nodes form edges, reflecting the spatial relatedness between them.

(3) Adjacency Matrix (AM): The adjacency matrix can be binary-coded or contain weighted real numbers, representing the connectivity between nodes in the graph. A value of 1 in the matrix indicates an edge between the corresponding pair of nodes, while a value of 0 indicates no connection. The adjacency matrix is fundamental to the graph convolution operation, guiding the propagation of information through the graph.

(4) Degree Matrix: The degree matrix represents the degree of each node in the graph, indicating the number of edges connected to that node.

(5) Node Feature Matrix: The node feature matrix stores the feature information of each node, with each row corresponding to a node and each column corresponding to a feature. These features can include spectral information, texture, geometric shape, and more. The node feature matrix is used as input for graph convolution, where feature extraction is performed through the graph convolution layer.

Table 1
Description of feature calculation.

Feature type	Subset of features	Formula	Instruction	Output dimension
Spectral Features (Jia, 2006)	Mean Intensity	$\frac{1}{N}\sum_{i=1}^{N} I_i$	For each band, the average intensity of all pixels in the region was calculated separately, and then the average intensity of the six bands was averaged. Where I_i: the intensity value of the ith pixel. N: the total number of pixels in the region.	6
	Std Intensity	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_i - MI)^2}$	The standard deviation of the intensities of all pixels in the region was calculated separately, and then the standard deviation was averaged over the six bands.	6
	Max Intensity	$\max(I_1, I_2, \cdots, I_N)$	The maximum intensity of all pixels in the region is calculated separately, and then the maximum intensities of the six bands are averaged.	6
	Min Intensity	$\min(I_1, I_2, \cdots, I_N)$	The minimum intensity of all pixels in the region is calculated separately, and then the minimum intensities of the six bands are averaged.	6
	Median Intensity	$\mathrm{Med}(I_1, I_2, \cdots, I_N)$	The median intensities of all pixels in the region were calculated separately, and then the median intensities of the six bands were averaged.	6
Geometric Features (Haralick and Shapiro, 1992; Twogood and Sommer, 1982)	Area	*Total number of pixels in the region*	The area is the total number of all pixels in the region.	1
	Perimeter	*Length of the boundary of the region*	The perimeter is the total length of the area boundary.	1
	Eccentricity	$\sqrt{1-(\frac{minoraxislength}{majoraxislength})^2}$	Eccentricity describes the ellipticity of the region, i.e. the deviation of the shape relative to a perfect circle.	1
	Solidity	$\frac{Area}{ConvexArea}$	Solidity is the ratio of the area of a region to its bump area and indicates how well the region fills the bump.	1
	Aspect Ratio	$\frac{Widthoftheboundingbox}{Hightoftheboundingbox}$	The aspect ratio is the ratio of the width to the height of the area bounding box.	1
	Convex Area	*Area of the convex hull of the region*	The convex envelope area is the area of the smallest convex polygon that encloses the region.	1
	Extent	$\frac{Area}{Areaoftheboundingbox}$		1
Colour Features (Swain and Ballard, 1991)	Colour Entropy	$-\sum_{i=1}^{K} p_i log_2(p_i)$	The colour entropy measures the uniformity of the colour distribution within the region, the higher the value, the more uniform the colour distribution. Where, p_i: probability of colour value i in the region. i: number of possible colour values in the region.	1
Texture Features (Haralick et al., 1973)	Texture Contrast	$\sum_{i,j}(i-j)^2 P(i,j)$	Contrast measures the sharpness of the image texture, the larger the value, the stronger the contrast of the texture. Where $P(i,j)$ denotes the value of position (i,j) in the grey scale covariance matrix (GLCM) of the image.	1
	Texture Dissimilarity	$\sum_{i,j}\lvert i-j\rvert P(i,j)$	Dissimilarity measures the local variation of grey values in an image, with larger values indicating greater differences in textures.	1
	Texture Homogeneity	$\sum_{i,j}\frac{P(i,j)}{1+\lvert i-j\rvert}$	Homogeneity measures the local similarity of an image, with larger values indicating greater uniformity of texture.	1
	Texture Energy	$\sqrt{\sum_{i,j}P(i,j)^2}$	Energy measures the repetitiveness of the texture in the image, with larger values indicating a more regular texture.	1
	Texture Correlation	$\sum_{i,j}\frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i\sigma_j}$	Correlation measures the linear dependence between the grey levels of an image, with larger values indicating a stronger linear relationship in the texture. Where μ_i, μ_j denote the mean values of i and j in GLCM. σ_i, σ_j denote the standard deviation of i and j in GLCM.	1
	Texture ASM	$\sum_{i,j}P(i,j)^2$	ASM measures the symmetry of the image texture, with larger values indicating a more regular texture.	1
	Texture Entropy	$\sum_{i,j}P(i,j)log_2(P(i,j))$	Texture entropy measures the randomness of the image texture, with larger values indicating higher complexity of the texture.	1
Shape Features (Haralick and Shapiro, 1992)	Convexity	$\frac{PerimeteroftheConvexHull}{perimeteroftheRegion}$	Convexity indicates how smooth the boundary of the region is relative to its convex envelope, with values closer to 1 indicating that the shape of the region is closer to convex.	1
	Aspect Ratio Shape	$\frac{MajorAxisLength}{MinorAxisLength}$	The shape aspect ratio indicates the aspect ratio of the region, with larger values indicating a narrower and longer shape of the region.	1
	Eccentricity Shape	$\sqrt{1-(\frac{MajorAxisLength}{MinorAxisLength})^2}$	Shape eccentricity describes the elliptical shape of the region, with values closer to 1 indicating that the region is closer to a straight line.	1
	Compactness	$\frac{Perimeter^2}{4\pi \times Area}$	Tightness measures how tightly the region is shaped, with smaller values indicating that the region is closer to being round.	1
Frequency Domain Features (Mallat, 1989)	Mean FFT	$\frac{1}{N}\sum_{i=1}^{N} F_i$	The average FFT measures the average intensity of all frequency components in the region and reflects the overall frequency characteristics of the region. Where F_i: Magnitude value of the i th frequency component after Fourier transform. i: Total number of frequency components.	1
	Std FFT	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - MeanFFT)^2}$	The standard deviation FFT measures the degree of variation in the strength of the frequency components and indicates fluctuations in the frequency characteristics of the region.	1
	Wavelet Mean	$\frac{1}{N}\sum_{i=1}^{N} W_i$	The wavelet transform mean measures the average intensity of all wavelet components in the region and reflects the local frequency characteristics of the region. Where W_i denotes the amplitude value of the i th component after wavelet transform.	1

(*continued on next page*)

Table 1 (*continued*)

Feature type	Subset of features	Formula	Instruction	Output dimension
	Wavelet Std	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(W_i - WaveletMean)^2}$	The standard deviation of the wavelet transforms measures the degree of variation in the intensity of the wavelet components and indicates fluctuations in the local frequency characteristics of the region.	1
Statistical Features (Duda et al., 2001)	Mean Value	$\frac{1}{N}\sum_{i=1}^{N} V_i$	The mean value measures the average of all pixel values in the region. Where V_i: the value of the i th pixel in the region.	6
	Std Value	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(V_i - MeanValue)^2}$	The standard deviation measures the degree of fluctuation in pixel values within the region.	6
	Skewness	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\frac{V_i - MeanValue}{\text{Std Value}})^3}$	Skewness measures the symmetry of the distribution of pixel values within a region, with positive skewness indicating that the distribution is right-skewed and negative skewness indicating that the distribution is left-skewed.	6
	Kurtosis	$\frac{1}{N}\sum_{i=1}^{N}(\frac{V_i - MeanValue}{\text{Std Value}})^4 - 3$	Kurtosis measures the peakiness of the distribution of pixel values within a region, with larger values indicating a sharper distribution.	6
	Covariance Eigenvalue	$\frac{1}{N}\sum_{i=1}^{N}(F_i - MeanValue) \times (V_i - MeanValue)^T$	The covariance matrix of the pixel values in the region is calculated and then the eigenvalues of the covariance matrix are extracted. These eigenvalues reflect the covariance structure of the pixel values in the region.	6
Structural Features (Young, 1983)	Skeleton Area	$Areaoftheskeletonoftheregion$	The skeleton area is the area of the region's skeleton and represents the basic shape of the region.	1
	Porosity	$\frac{Areaofthevoids}{TotalArea}$	Porosity indicates the proportion of voids within the region, with larger values indicating a more porous region.	1

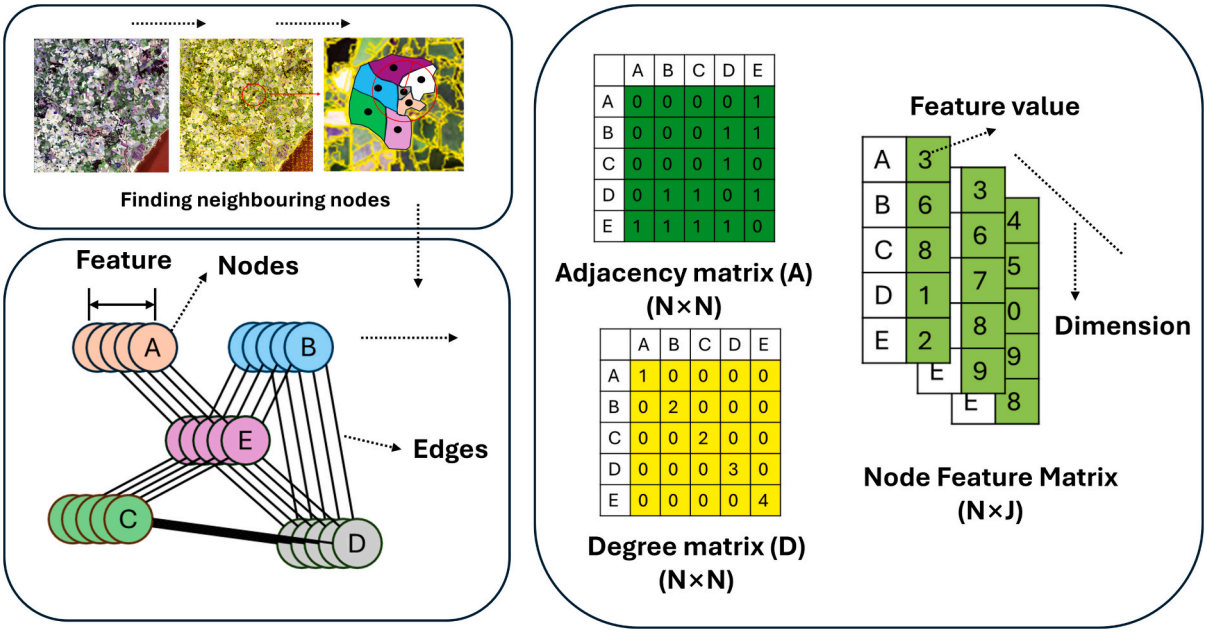

Fig. 2. Graph construction process. The yellow lines in the finding neighbouring nodes diagram indicate the dividing lines of the individual objects. Here N denotes the number of nodes and J denotes the number of features. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(6) Graph Laplacian Matrix: The graph Laplacian matrix is a variant of the adjacency matrix, often used to normalise the graph structure and stabilise the graph convolution operation. It balances the influence between nodes by subtracting the degree matrix.

To construct the graph, the initial object boundaries are obtained using the segmentation method introduced in Section 3.1, which segments the image into spatially coherent regions suitable for object-level feature aggregation and graph construction. Then, remote sensing image data is transformed into a graph structure (as shown in Fig. 2). The process begins by using the 'regionprops' function to calculate the centroid of each region, representing the geometric centre of each node in the image (Chen et al., 2021). The distance between these centroids is then used to determine the connectivity between nodes. Specifically, a KDTree data structure is employed to accelerate the computation of distances between centroids, and the query ball tree method is used to identify the neighbours of each node within a specified radius. This radius, which might be set to 5 object units, is determined based on the extent and complexity of the actual segmentation. By traversing each node and its neighbour list, an adjacency matrix A is constructed, where $a_{ij} = 1$ if node i and node j are neighbours, and 0 otherwise. To save memory and computational resources, this adjacency matrix is stored using sparse matrices. After obtaining the adjacency matrix A, the graph Laplacian matrix L is computed. The Laplacian matrix L is calculated as the difference between the degree matrix D and the adjacency matrix $A(L = D - A)$, where D is a diagonal matrix with elements D_{ii} representing the degree of node i (i.e., the number of edges connected to it). The graph Laplacian matrix is used to describe the overall structure of the graph and the global relationships between nodes.

After completing the graph construction, we obtain the node feature matrix, the graph Laplacian matrix, and the labelling data, which is

extracted from the original labelled image and represents the category information of each node (object). Once the training data is prepared, the node feature matrices are first normalised to ensure consistent value ranges across different features, thereby avoiding issues in model training caused by varying magnitudes between features. Simultaneously, the labelled data is encoded and transformed to make the classification task numerically computable. Finally, the graph structure data (including the adjacency matrix and graph Laplacian matrix) is paired with the node feature matrices to construct the data objects required for GCN training. These preprocessing steps are crucial for ensuring the quality of the input data and the stability of the model training.

3.2.2. Graph convolutional neural network

In this section, we emphasise the modular design of SAGRNet to elucidate its distinctive benefits in enhancing the vegetation classification task. The primary objective of this section is to provide a comprehensive explanation of how SAGRNet overcomes the constraints of conventional object-based and pixel-based classification models when processing intricate graphics data by integrating multiple feature extraction algorithm structures.

SAGRNet combines a sampling and aggregation mechanism (Hamilton et al., 2017) (SAGEConv) with a graph attention mechanism (Veličković et al., 2017) (GATConv), emulating the ResNet feature extraction method to efficiently classify objects in remote sensing imagery. The model integrates two key operations of GCN, SAGEConv and GATConv, and enhances feature representation and gradient flow through a deep residual network structure (ResNet) (He et al., 2015).

sample and aggregate convolutional layer, aggregates information from a node's neighbours in a more efficient manner. Unlike traditional graph convolution, which typically aggregates information from all neighbouring nodes—resulting in significant computational overhead in large-scale graphs—SAGEConv reduces this complexity by sampling a subset of neighbouring nodes for aggregation.

Initially, the node limits the amount of computation by sampling neighbouring nodes. The features of these sampled neighbouring nodes are then integrated into the current node's features using a mean-value aggregation function, producing a new feature representation (as shown in Fig. 3(a)). Next, this new feature is transformed through a linear transformation followed by a nonlinear activation (e.g., ReLU). This process continues layer by layer, with each layer integrating more neighbourhood information, ultimately generating more discriminative node features. The features of the sampled neighbouring nodes are aggregated through operations such as averaging and pooling to generate new features for the target node:

$$h_{\mu_{(i)}}^{(l+1)} = Aggregate(\{e_{ji}h_j^l, \forall j \in \mu_{(i)}\}) \tag{4}$$

$$h_i^{(l+1)} = norm(\sigma(W \cdot concat(h_j^l, h_{\mu_{(i)}}^{(l+1)}) \tag{5}$$

where e_{ji} is the scalar weight of the edges from node j to node i. Equation (1) represents the characteristics of the aggregated neighbour nodes, where $\mu_{(i)}$ denotes the set of neighbours of the node i, e_{ji} is the scalar weight of the edges from the node j to the node i, and is used to weight the characteristics h_j^l of the neighbour nodes. Equation (2) updates the representation of the current node h_j^l. The representation of the current node is updated using the features of the current node and the aggregated neighbour information $h_{\mu_{(i)}}^{(l+1)}$. Where W is the learnable weight matrix, σ is the activation function. Finally, the features are normalised to ensure the stability of the training process. This process means that the features of each node are updated at every layer, and this update depends on the aggregated information from its neighbours. While the structure of the graph (i.e., the neighbourhood relationships) remains unchanged, the features of the nodes are updated at each layer through convolution operations. As a result, after neighbourhood aggregation, the feature values of some nodes may decrease or even approach zero. However, this does not imply that the nodes themselves are deleted or that their connectivity relations are removed.

GATConv, the Graph Attention Convolutional Layer, introduces a self-attention mechanism into graph neural networks. The core idea of this mechanism is to assign different weights to each node and its neighbours, allowing for more nuanced aggregation of neighbouring node information (as shown in Fig. 3(b)).

In traditional graph convolution, node feature updates are typically achieved by simply averaging or summing the weighted features of neighbouring nodes (as shown in Fig. 4(b)). However, in GATConv, this aggregation is enhanced by computing attention weights G_{ij} between the neighbouring nodes and the target node. The attention mechanism is defined as:

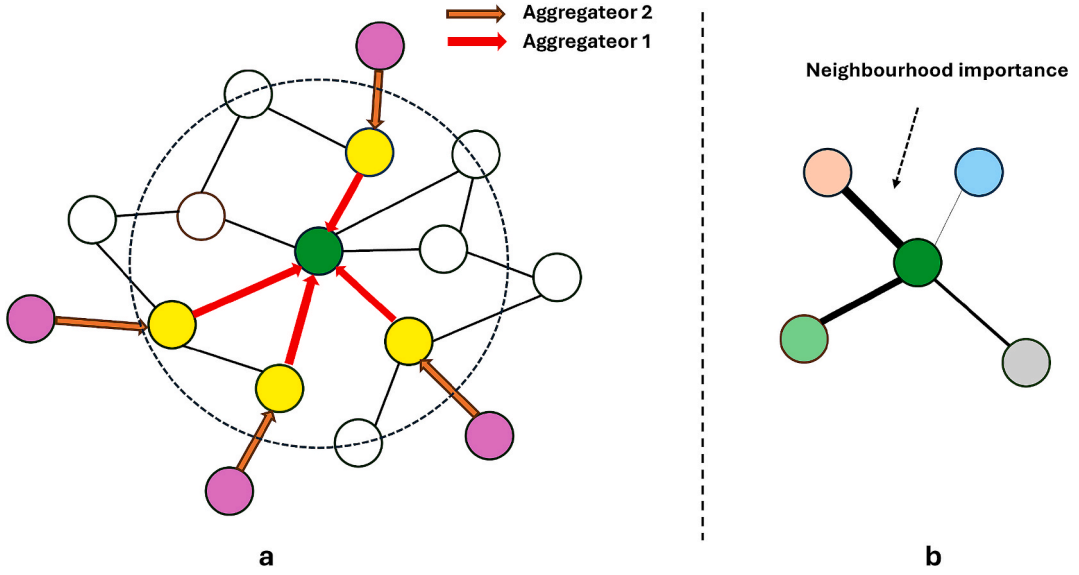

Fig. 3. (a), (b) Schematic representation of the sampling aggregation method and the self-attention mechanism, respectively. In (b), the connecting lines of different thicknesses in the diagram indicate different weight sizes, which indicate the importance between different surrounding nodes and the central node.

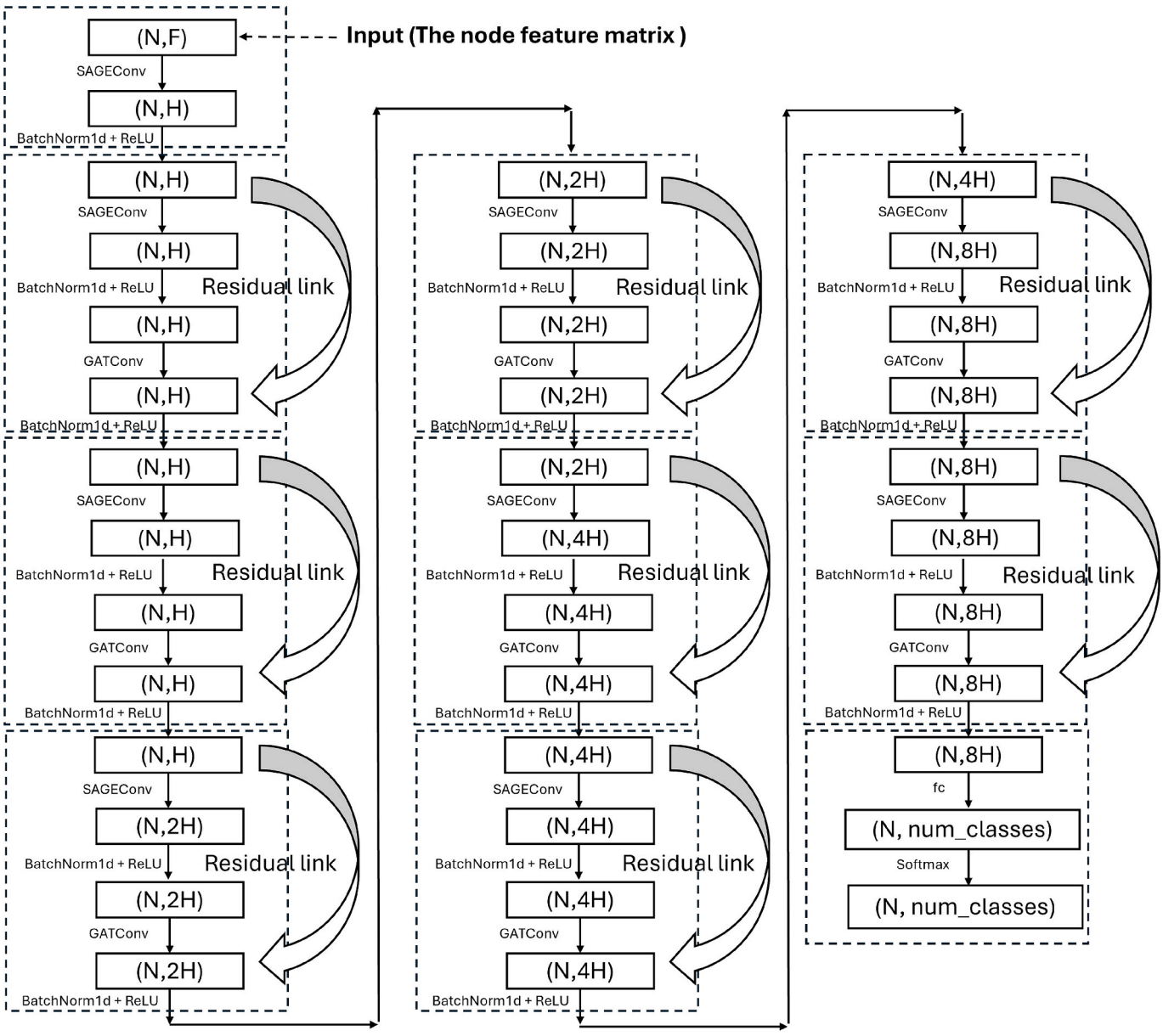

Fig. 4. Structure of SAGRNet, Where N is the number of nodes and F is the number of features per node. The edge index matrix is of the form (2, E), where E is the number of edges. E plays a role in SAGEConv and GATConv.

$$G_{ij} = \frac{\exp\left(LeakyReLU(a^T[Wh_i \| Wh_j])\right)}{\sum_{k\in\mu_{(i)}} \exp\left(LeakyReLU(a^T[Wh_i \| Wh_k])\right)} \tag{6}$$

where a is a learnable attention vector, W is a learnable weight matrix, and h_i and h_j are the feature vectors of the target and neighbour nodes, respectively. As described in Eq. (3), a linear transformation of the node features is first performed, where the feature x_i of each node i is mapped to a new feature space using a linear transformation matrix W, resulting in the transformed feature Wh_i. Similarly, the feature x_j of node j is transformed using the same matrix to obtain Wh_j. The attention weight e_{ij} is then calculated by concatenating the transformed features Wh_i and Wh_j and applying a learnable vector a and the LeakyReLU activation function. The attention weight e_{ij} represents the importance score of node i relative to node j. Additionally, $a^T[Wh_i \| Wh_j]$ in the equation denotes that after concatenating Wh_i and Wh_j, they are mapped by a linear model (defined by the vector a) and undergo a LeakyReLU nonlinear transformation. Finally, the normalized attention weight G_{ij} is obtained by softmax normalizing the attention weight e_{ij}, determining how much the feature of node j contributes during the feature update process of node i. The normalization is done by calculating the relative magnitude of e_{ij} among all neighboring nodes. The denominator in the equation represents the normalization operation performed across all neighbouring nodes k.

$$H_i = \sum_{j\in\mu_{(i)}} G_{ij} h_j \tag{7}$$

SAGRNet design, the combination of SAGEConv and GATConv is embedded within the residual block, mimicking the structure of ResNet18 (as shown in Fig. 4). Initially, SAGEConv performs the aggregation of the node's features, and then GATConv further weights and aggregates the neighbouring features using an attention mechanism. Finally, the input features are directly added to the convolved output, which is subsequently nonlinearly transformed by the ReLU activation function. This design allows the network to maintain focus on the original features while extracting deep features, thereby enhancing

information flow and improving the model's expressive power.

3.2.3. Implementation

The training and evaluation of the model are conducted on Google Cloud's TPU v2 platform, which offers up to 180 TFLOPs of floating-point computational power, providing high throughput and low latency. This platform is particularly well-suited for training large-scale deep learning models. All experiments were conducted on Google Cloud TPU v2 devices, each with 8 cores and 64 GB high-bandwidth memory. This environment ensures high-throughput parallel processing and efficient graph operations. In the experiments, the model's training parameters are set as follows: the learning rate is 1e-4, the weight decay coefficient is 5e-4, and the Adam optimizer is used. During the training process, 70 % of the data is sampled for training, while the remaining 30 % is used for prediction and evaluation of the model. During model training, the proposed SAGRNet exhibited significantly reduced computational time compared to baseline models. Specifically, SAGRNet required approximately 4 min per regional dataset (10 km × 10 km) to reach convergence within 1000 epochs, while U-Net++ and DeepLabV3 typically required over 30–40 min under identical conditions. This 8–10 × speedup can be attributed to the model's lightweight graph-based architecture, reduced parameter complexity, and early convergence behaviour. Furthermore, the reproducibility of our training process is ensured through fixed random seeds, consistent hyperparameters, and publicly accessible hardware infrastructure. This facilitates the direct replication of training efficiency and model performance under identical computational conditions.

3.2.4. Comparison test

To demonstrate the significant advantages of the models proposed in this paper, we have selected several classical object-based and pixel-based models for comparison, focusing on their runtimes, F-score, and IOU results.

Graph U-Net (Gao and Ji, 2019) is a graph neural network architecture inspired by U-Net, which enables multi-scale representation learning of graph data by combining pooling and up sampling operations. During graph convolution, Graph U-Net adaptively pools the graph, gradually reducing its size to capture global information, and reconstructs node features during the up-sampling phase. This structure is particularly suited for tasks that require a combination of global and local information. To enhance its performance, we introduce tunable pooling ratios (e.g., 0.6, 0.4, 0.2), allowing the model to flexibly learn different levels of graph structure information.

Transformers with Graph Neural Networks (TGNN) (Yun et al., 2022) is an innovative approach that combines the advantages of Transformers and GNN. The model replaces the traditional graph convolutional layer with TransformerConv, leveraging the Multi-Head Attention mechanism to achieve more accurate information aggregation. This makes it particularly effective for complex graph-structured data. Unlike GATConv, which is based on local attention and focuses on fixed neighbours, TGNN employs a global self-attention mechanism. This mechanism dynamically adjusts dependencies between nodes based on feature similarity, enabling it to capture global relationships in the graph. In other words, TGNN's attention mechanism operates across the entire graph structure, not just within neighbouring nodes, allowing for more flexible feature aggregation. The model architecture consists of multiple TransformerConvs per layer, with residual concatenation and BatchNorm to enhance training stability.

UNet++ is an enhanced pixel-based network architecture derived from U-Net (Zhou et al., 2018). UNet++ improves the fusion of multi-scale features by introducing a series of densely nested skip connections between the encoder and decoder. Compared to the traditional U-Net, UNet++ has a more complex structure, designed to improve segmentation accuracy and fine-grained information capture by gradually narrowing the semantic gap between features. The network excels in handling small target segmentation and images with multi-scale features, making it widely used in various image segmentation tasks.

DeepLabV3 is an advanced pixel-based semantic segmentation model proposed by Google (Chen et al., 2018). The model significantly enhances the segmentation of multi-scale objects by incorporating techniques such as atrous (dilated) convolution and the Atrous Spatial Pyramid Pooling (ASPP) module. Atrous convolution expands the receptive field of the convolutional kernel without increasing computational complexity, allowing the model to capture more contextual information. The ASPP module further integrates features at different scales, enabling the model to perform exceptionally well in complex scenes and with multi-scale objects. Due to structural differences, pixel-based algorithms require the raw data to be uniformly segmented in advance, generating regular 516x516 square raster. In contrast, object-based algorithms do not require this preprocessing. Additionally, both pixel-based algorithms use ResNet18 as the feature extraction backbone. Other parameters are kept consistent with the ResNetAmGCN training, although some adjustments may be necessary to optimise prediction accuracy.

Since many of the object-based neural networks mentioned earlier are still in the testing phase, we have selected relatively more mature models that we have tested and successfully implemented for comparison with the models presented in this paper. Although these models have shown good performance on standard datasets, practical applications often reveal challenges, particularly related to parameter tuning and optimizer adjustments. To address these issues, we have carefully adjusted the parameters and model architectures to better suit the training and prediction requirements. Our goal was to minimize the loss function during training and ensure that the prediction performance is optimized.

For sample construction, the training and test sets are distributed in a 7:3 ratio. The benchmark for accuracy calculation is the standard label set. Accuracy is evaluated by comparing the predicted labels with the benchmark label set when calculating the F1 score and IOU, ensuring a comprehensive assessment of overall classification accuracy.

To validate the performance of the mapping, we calculated the confusion matrix, Precision, Recall and F1 score. The F1 score is the reconciled average of Precision and Recall:

$$P = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} (TP_i + FP_i)} \tag{8}$$

$$R = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} (TP_i + FN_i)} \tag{9}$$

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

The intersection ratio (IoU) is a commonly used evaluation metric in image segmentation and target detection tasks to measure the degree of overlap between the predicted results and the true results. The value of IoU is between 0 and 1, with larger values indicating that the predicted results are closer to the true results. The following is the formula:

$$IOU = \frac{\sum_{i=1}^{N} \left| A_p^i \cap A_g^i \right|}{N \sum_{i=1}^{N} \left| A_p^i \cup A_g^i \right|} \tag{11}$$

3.2.5. Module ablation test

In constructing an ablation test for the SAGRNet model, we focus on analysing the impact of different components on model performance. The ablation test systematically evaluates the contribution of key components by removing or replacing them step-by-step. First, we remove either the SAGEConv or GATConv layer and compare the results with the base model. Second, we ablate the ResNet18 residual connections to assess the role of residuals in addressing training stability and gradient vanishing issues. It is crucial that all other hyperparameters and the training process remain consistent during the tests to ensure fairness and

that any variations in the results are a true reflection of the specific components' influence. Through these tests, we can determine the contribution of each module to model accuracy and convergence speed, allowing for further optimization based on the results.

3.3. Global generalization test

To further evaluate the generalization capability of the proposed SAGRNet model, we conducted additional experiments across five globally distributed urban fringe areas: Guangzhou (China), Durban (South Africa), Sydney (Australia), New York City (United States), and Porto Alegre (Brazil), as show in Fig. 5. These areas were carefully selected to represent diverse ecological backgrounds, vegetation structures, urbanization intensities, and land cover complexities, providing a robust basis for assessing the transferability of our method beyond the original experimental region.

All satellite data used in this section were obtained from Sentinel-2 Level-2A surface reflectance products, ensuring consistent spatial and spectral resolution across all sites. To minimize seasonal variation and enhance comparability across regions, all images were selected from December 2022 without clouds, aligning with the low-vegetation or dormancy period in most regions. To ensure a reliable and standardized ground truth, we adopted the Esri 2022 Global Land Cover dataset (https://livingatlas.arcgis.com/landcover/). This dataset offers globally consistent land cover labels at a 10-meter resolution, derived from Sentinel-2 imagery using a supervised deep learning pipeline. Its classification schema, spatial consistency, and compatibility with Sentinel-2 make it a suitable benchmark for evaluating model performance in diverse geographic settings.

To ensure consistency across test sites, we reclassified the original land cover into five unified categories, aligning with the classes used in our prior experiments. In selecting the study areas, we carefully considered seasonal factors to improve spectral separability. For instance, winter imagery was preferred in some regions to better distinguish cropland from grassland. Additionally, we conducted thorough visual inspections of the Esri label data to assess and confirm its accuracy. Despite being officially published, we observed notable misclassifications in several areas. These quality control measures were implemented to minimize the influence of external factors and ensure that model performance is evaluated solely based on classification capability.

The selected regions span five continents and cover a wide spectrum of climate zones, vegetation formations, and anthropogenic pressures. Each area includes complex urban-natural transition zones that are typically challenging for land cover classification. The Table 2 below outlines the main characteristics of each selected region and summarizes their respective land use and landscape composition:

Each 10 km × 10 km area was processed following the same workflow described in earlier sections. Sentinel-2 bands were selected for spectral analysis, followed by object-based segmentation using the Felzenszwalb algorithm. Spectral, textural, and geometric features were extracted for each object, and the Esri land cover map was used to assign labels for training and evaluation. The classification test was also performed using the model proposed in this paper. The relevant parameters were set to remain consistent with those in Section 3.2.3.

4. Result

4.1. Object-based image classification and feature importance analysis

To assess the importance and impact of input features on the model's classification predictions, this study uses SHAP values to rank the contribution of all features through uniform random sampling. Fig. 6 illustrates the distribution of feature importance for different classification types based on the XGBoost classifier. Each row represents a feature, with the horizontal axis showing the SHAP value. Features are ranked according to the average absolute SHAP value, with the most important feature at the top. Wider sections indicate a large cluster of samples, while each dot represents a sample: red indicates a higher feature value, and blue indicates a lower feature value.

To ensure that all features contribute significantly to the classification of different types, and to reduce the potential impact of excessive features, the study first analysed the correlation between all computed feature attributes and excluded highly correlated features, as demonstrated in Fig. 7. The 24 features with low correlation were retained. Subsequently, these features were subjected to SHAP analysis and the

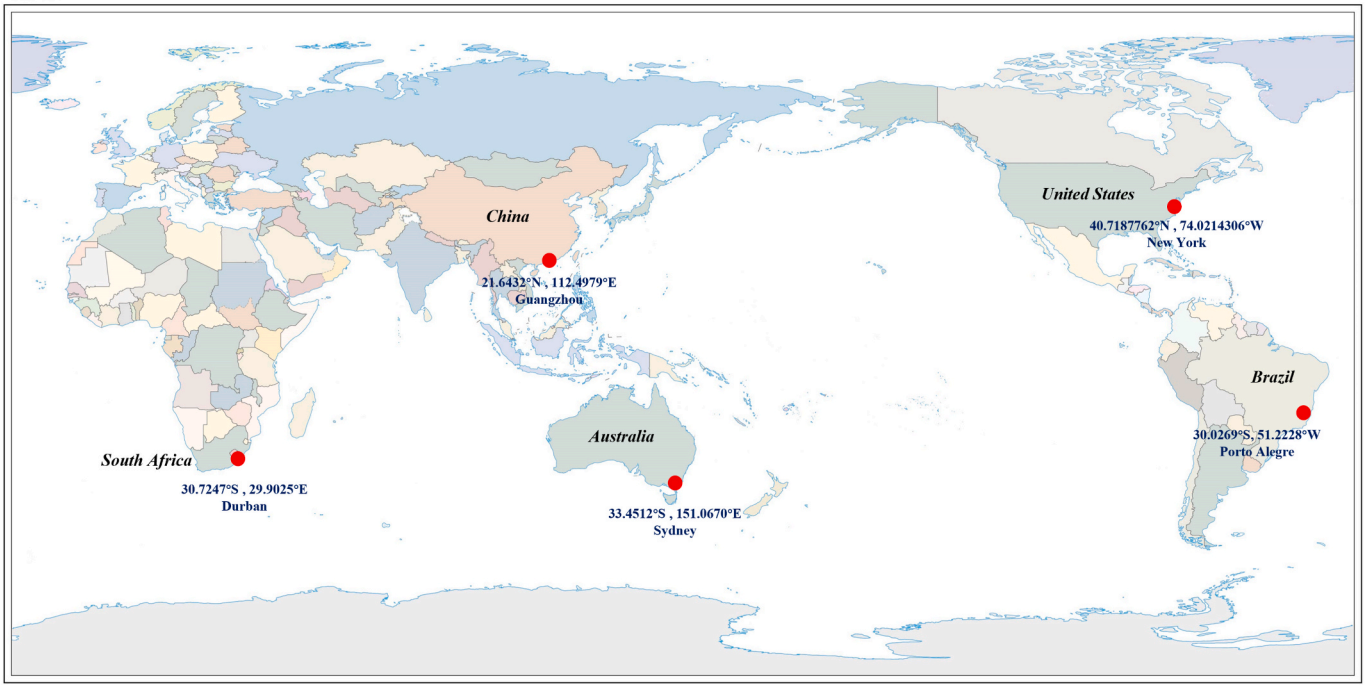

Fig. 5. Global generalization test area location.

Table 2
Detailed information on the cities tested (Gui et al., 2025).

Region	Country	Dominant Land Cover Types	Climate Zone	Landscape Description
Guangzhou	China	Dense Urban, Paddy Fields, Evergreen Forest	Subtropical Monsoon	Urban areas are densely concentrated with strong spectral contrast against surrounding vegetation. Though isolated buildings may occur in green zones, urban features are still clearly distinguishable. However, spectral confusion can occur between subtropical forests and managed green spaces, while arable land remains relatively separable.
Durban	South Africa	Grassland, Savannah, Low-density Urban	Humid Subtropical	Central urban areas show moderate building density, decreasing toward the periphery. The outskirts feature bare surfaces gradually transitioning into grassy terrain. Vegetation types are spatially intertwined, with fine-scale fragmentation and scattered patches of shrubland or grassland within other categories, creating a highly irregular landscape.
Sydney	Australia	Urban, Coastal Shrubland, Dry Forest	Temperate Subtropical	Urban layout is relatively organized with substantial vegetated buffers. Clear spectral differences are seen among grassland, arable land, and trees, though some isolated vegetation patches exist. The landscape is moderately structured, aiding in visual and algorithmic distinction.
New York City	United States	Urban, Deciduous Forest, Wetlands	Temperate Continental	Ecological layout is relatively homogeneous. Urban and green zones exhibit spatial regularity. Land cover categories such as arable land, woodland, and urban features are spectrally distinct, simplifying classification and improving structural clarity.
Porto Alegre	Brazil	Urban, Mixed Cropland, Atlantic Forest	Humid Subtropical	Urban areas show strong spectral contrast with vegetation, but inter-class overlap exists among grassland, arable land, and woodland areas. Land patches are generally larger, though vegetation categories are difficult to distinguish due to similar spectral responses in certain regions.

top five importance rankings in the corresponding feature analysis for each type were used as the calculated metrics for individual object feature attributes. After summarizing and combining these features, we selected texture contrast, covariance eigenvalue band 2, aspect ratio, mean intensity band 1, area, minimum intensity band 5, covariance eigenvalue 5, texture correlation, mean intensity band 5, standard deviation (std) value, std intensity band 4, texture homogeneity, and covariance eigenvalue band 3 as key metrics for object property calculation.

Fig. 8 presents the effect of feature calculation and object-based classification, with F1 and IOU scores of 0.72 and 0.61, respectively. Visually, the overall classification performance is promising, but upon closer inspection, several large errors are evident. While the classification of water bodies and urban environments is relatively accurate, more noticeable misclassifications occur in forest land, grassland, and cultivated land. Furthermore, these misclassification areas tend to appear in large blocks, which is closely tied to the object segmentation size. This highlights the importance of classification accuracy in object-based classification, as even minor misclassifications can result in substantial errors across larger areas.

4.2. Classification prediction results based on the SAGRNet model

As shown in Table 3, the F1 and IOU scores of the SAGRNet model for classification prediction of different regions were evaluated after completing training, performed well in the classification task, achieving an accuracy of 0.9. Region 1 corresponds to an area with more urban development, while the region 2 corresponds to one with less urban area, as depicted in Fig. 1. The results indicate that the SAGRNet model performs relatively poorly in regions with more urban areas. Regarding the classification accuracy across different types, the main errors are concentrated in urban classification. The SAGRNet model performed well in the classification task, achieving an accuracy of 0.9, demonstrating its reliability in most object classifications. However, a deeper analysis of the classification results reveals that, in addition to overall classification errors for certain objects, the model exhibits errors in more complex object regions and along object boundaries. Specifically, in comparison with the labelled data, the model struggles with some object boundaries and sporadic pixels, showing more noticeable smoothing and classification errors. These discrepancies become particularly evident when zooming in on the images for visualization, as shown in Figs. 9 and 10. This issue may be attributed to the smooth transition of the model during the initial segmentation process, reflecting its limitations in handling fine details and noise. Additionally, it is worth noting

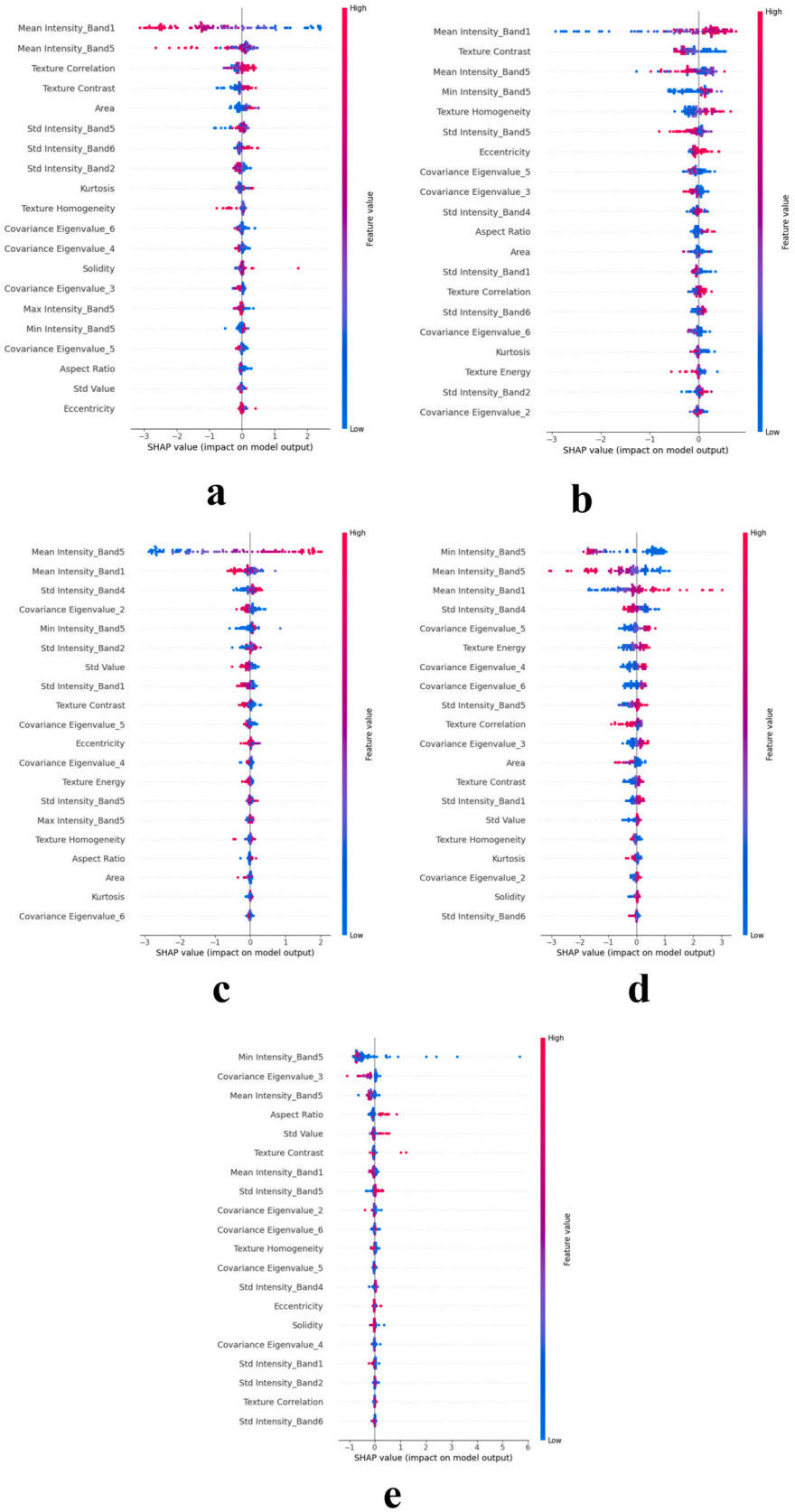

Fig. 6. Scatterplot of SHAP feature density. Graphs (a)(b)(c)(d)(e) correspond to the features analysis results for Woodland, Grassland, Cropland, Urban, and Water, respectively.

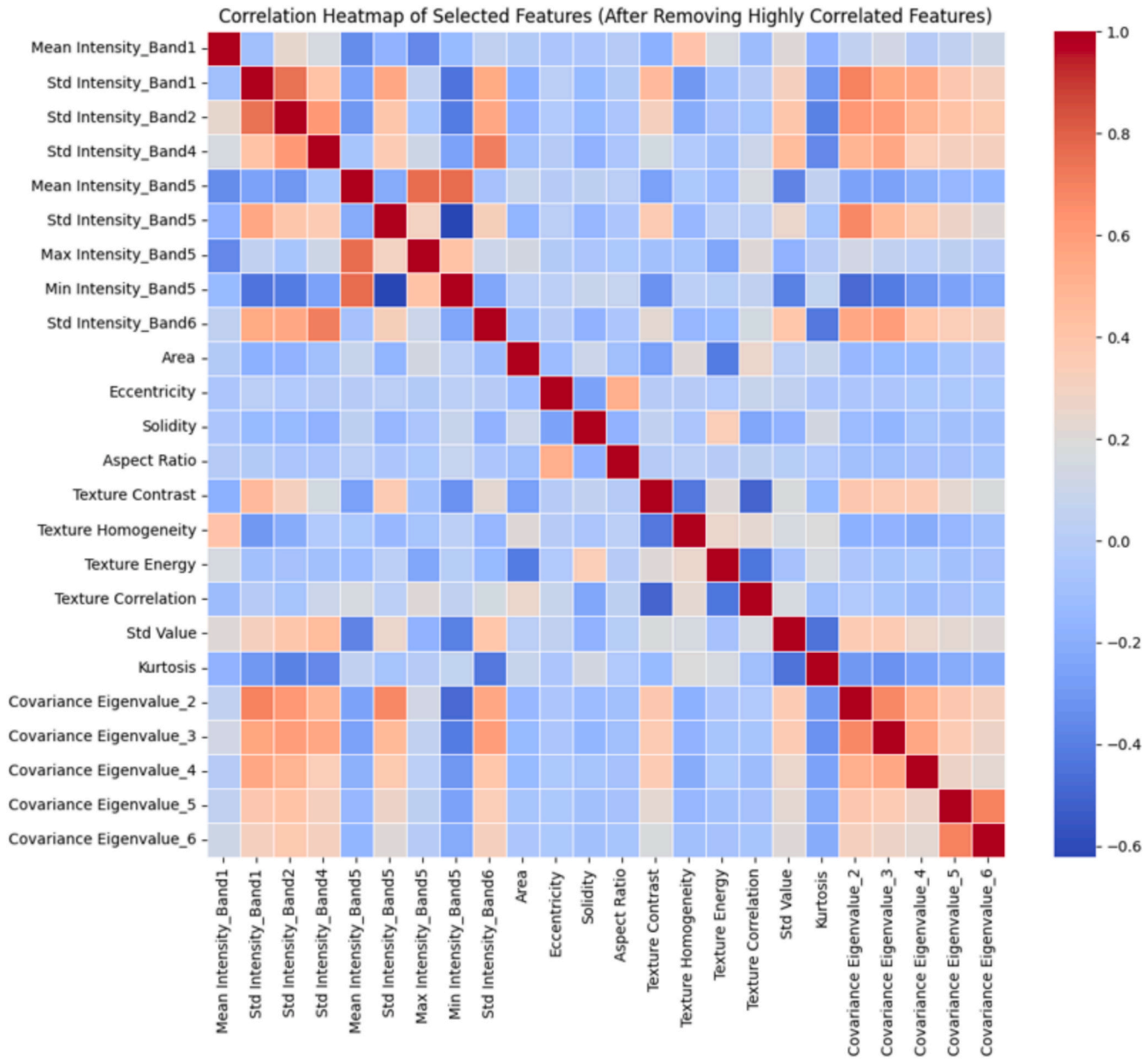

Fig. 7. Feature attribute correlation analysis hotspot map.

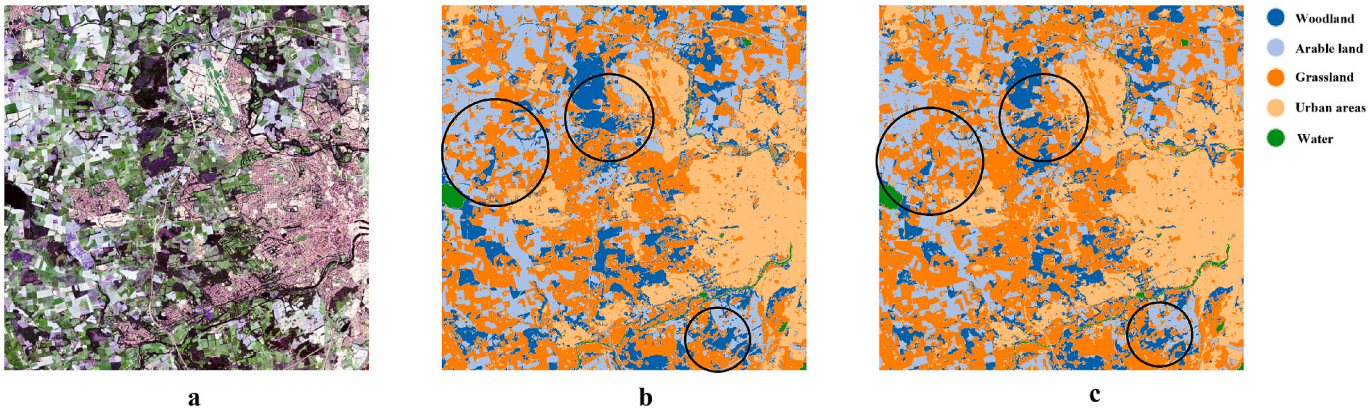

Fig. 8. Object-based classification results using the XGBoost classifier. Where a indicates the test impact RGB visualization, b indicates the classification result based on the traditional object algorithm, and c indicates the label classification diagram. The circles indicate several areas with obvious errors.

Table 3
Accuracy evaluation results based on the SAGRNet model.

	Classification type	F1 score	IOU	Total F1score
Region 1	Woodland	0.87	0.79	0.89
	Grassland	0.92	0.87	
	Arable land	0.92	0.85	
	Urban	0.87	0.84	
	Water	0.84	0.80	
Region 2	Woodland	0.92	0.85	0.92
	Grassland	0.94	0.86	
	Arable land	0.93	0.82	
	Urban	0.88	0.81	
	Water	0.94	0.83	

that the labelled data used in this study was sourced from an official dataset, which may have been generated using traditional pixel-based methods, introducing noise and errors into the dataset. This suggests that, while the model performs well at a macro level, the quality of the dataset and the labelling strategy may have impacted the accuracy evaluation.

4.3. SAGRNet model outperforms most other networks

From a quantitative perspective, as shown in Table 4, SAGRNet outperforms other models in terms of F1-score and IoU, with relatively low time consumption in both regions. In contrast, U-Net++ and DeeplabV3 show lower F1-scores and IoU values, with DeeplabV3′s F1-score in region 2 being particularly low at 0.81, coupled with significantly longer processing times. While Graph U-Net demonstrates faster performance and smoother loss convergence, its F1-score and IoU are considerably lower than those of TGNN. Although TGNN has moderately longer runtime, it still falls short of SAGRNet in F1-score and IoU.

From a visualization standpoint (as shown in Figs. 11 and 12), all models except Graph U-Net maintain relatively strong results. Overall, SAGRNet and U-Net++ maintain high levels of accuracy, with SAGRNet excelling at segmenting intricate details and ensuring smooth transitions in complex regions (as shown in Figs. 13 and 14). It avoids sporadic single-pixel noise while maintaining a high level of accuracy. In comparison, U-Net++ performs slightly less effectively in certain details, particularly in classifying larger objects, where it is outperformed by SAGRNet. DeeplabV3, although consistent over large areas, misses fine details and tends to over-smooth the results, leading to the misclassification of smaller objects. While TGNN guarantees reasonable accuracy overall, it noticeably loses detail in more complex scenes.

4.4. Module ablation experiment

Based on the results of the ablation test. From a quantitative perspective, as shown in Table 5, the combination of SAGEConv, GATConv, and ResNet18 provides the best performance, yielding a high F1-score and IoU, which indicates that the model effectively captures details while maintaining a reasonable time cost. Removing SAGEConv resulted in a significant decline in F1-score and IoU, particularly in complex regions, although it reduced the time cost. Similarly, removing GATConv caused a moderate decrease in the model's performance. When using only SAGEConv and ResNet18, the classification accuracy decreased noticeably, despite the reduced running time. In summary, the combination of SAGEConv, GATConv, and ResNet18 achieves the optimal balance between higher accuracy and efficiency, whereas

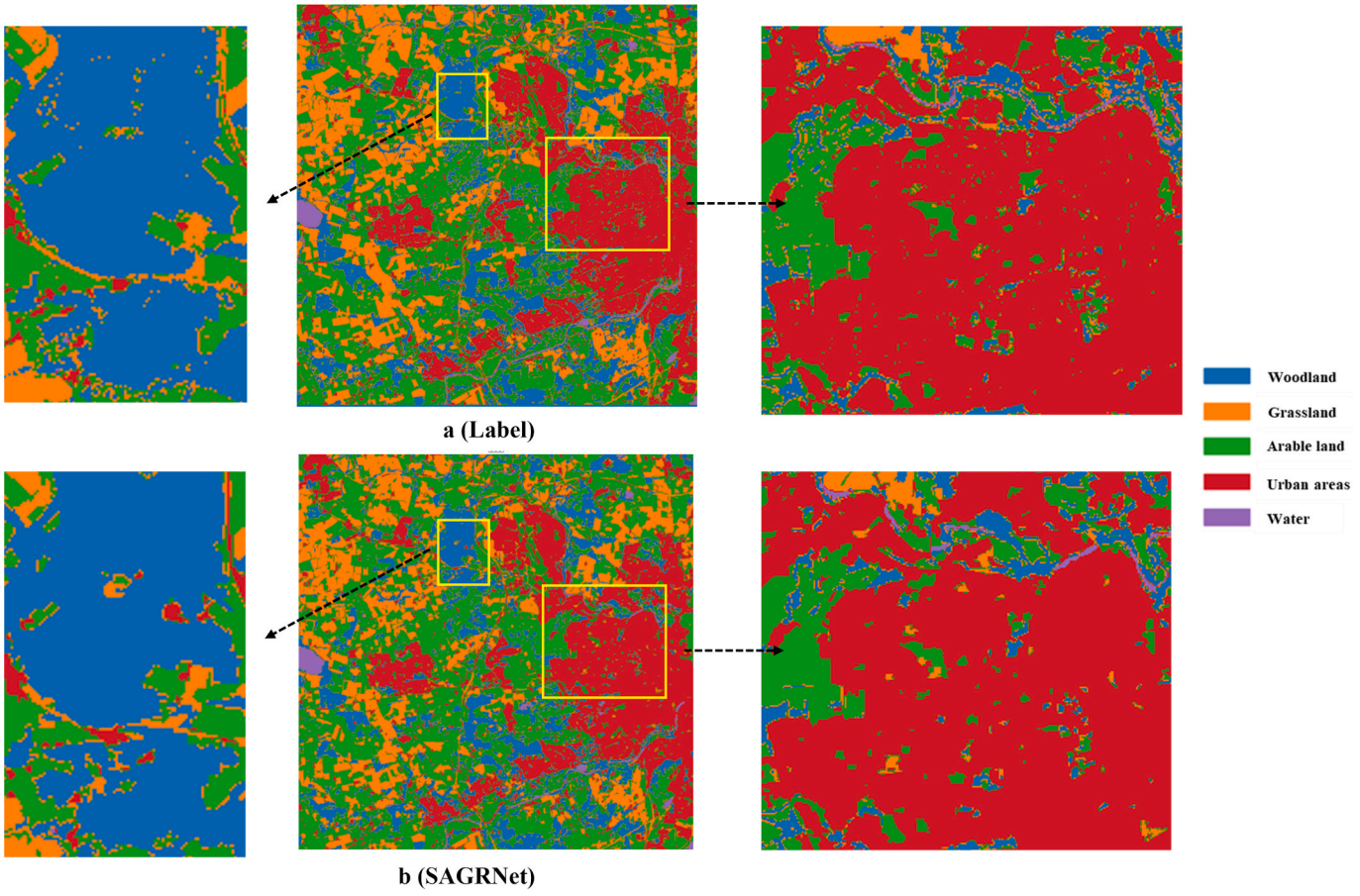

Fig. 9. Comparison of label with predictions based on SAGRNet model for region 1. where a is the label classification graph and b is the classification result based on SAGRNet.

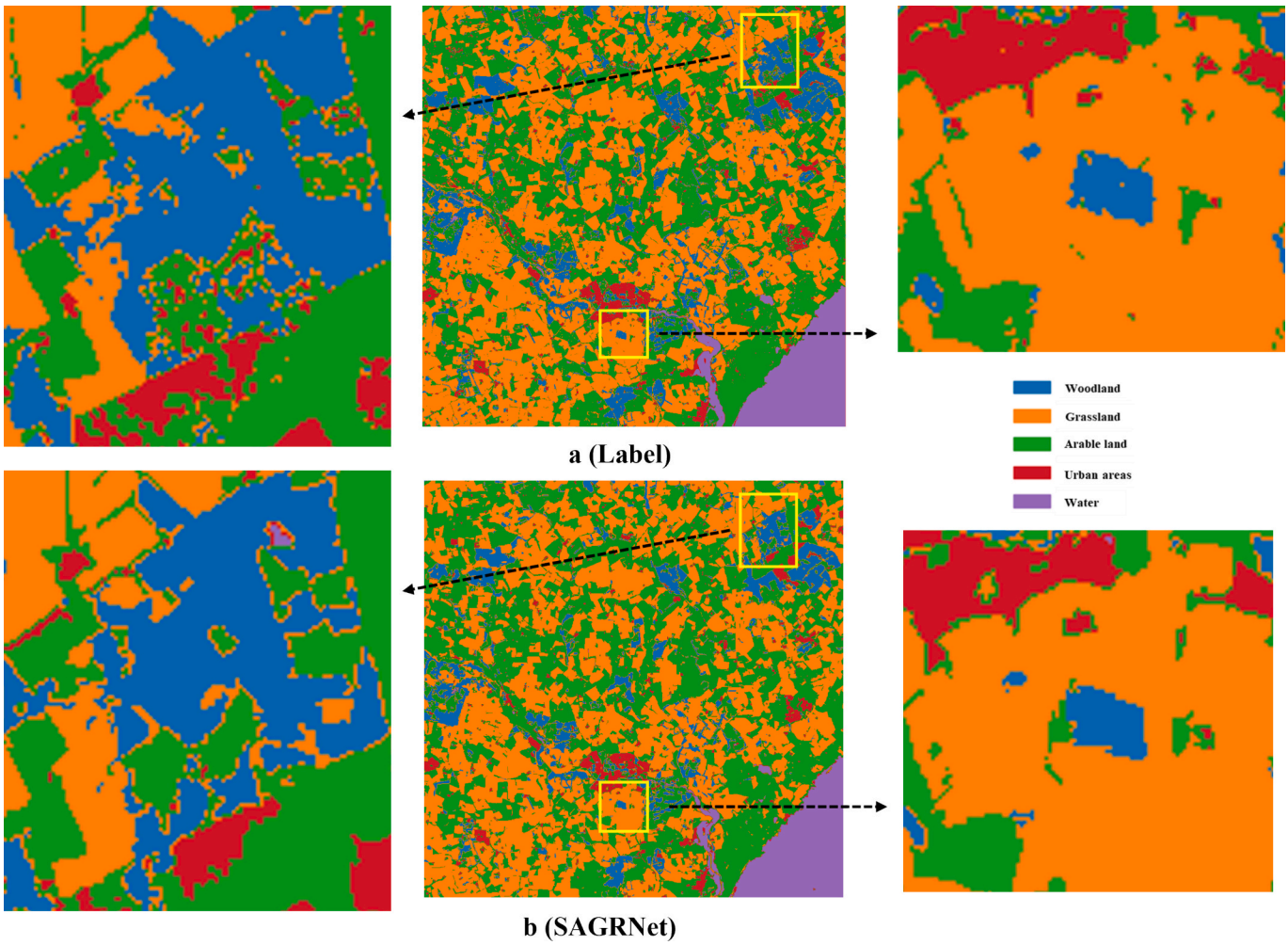

Fig. 10. Comparison of label with predictions based on SAGRNet model for region 2. where a is the label classification graph and b is the classification result based on SAGRNet.

Table 4
Comparison of accuracy assessment of different algorithms.

	Region 1			Region 2		
	Time	F1-score	IOU	Time	F1-score	IOU
SAGRNet	21 min	0.89	0.82	9 min	0.92	0.84
U-Net++	235 min	0.87	0.79	213 min	0.90	0.82
DeeplabV3	276 min	0.83	0.71	245 min	0.81	0.69
Graph U-Net	15 min	0.67	0.43	7 min	0.62	0.38
TGNN	28 min	0.81	0.73	15 min	0.71	0.62

removing any of these components negatively impacts the model's overall performance. From a qualitative point of view, as shown in Figs. 15 and 16, they perform consistently on classification smoothing, with most of the error sources coming from misclassification of individual objects, while the model that integrates all modules has fewer classification errors compared to the others.

4.5. Generalization and robustness

To evaluate the generalization ability and robustness of the proposed method, we selected four additional vegetation study areas in the UK as test data using the same model settings. As shown in Fig. 17, the visual consistency between predictions and reference labels remains high, and most boundaries and vegetation structures are well preserved. From a visual inspection perspective, the model was able to delineate mixed vegetation patterns and structural transitions in a reasonable manner, including complex patches and small-scale changes. Quantitative evaluations further support this, with all F1 scores are above 0.9 and IoU values exceeding 0.8, confirming the model's reliability and generalization capability in unseen regions. These results reflect the overall strong performance of the model, though slight errors are still present. Most of the residual misclassifications are concentrated near class boundaries or in areas where the reference labels themselves exhibit smoothing or structural generalization. Additionally, due to the use of object-based segmentation, some very small features may be missed or overly simplified. Misjudgements may also occur with some objects with small differences in spectral characteristics. Nonetheless, the observed errors are minor and do not affect the overall classification structure, reinforcing the effectiveness of SAGRNet across diverse landscapes.

4.6. Results of global regional classification

Results show that the proposed SAGRNet model consistently achieved strong classification performance across all five test sites. Specifically, the model attained F1-scores of 0.91 (Guangzhou), 0.89 (Durban), 0.93 (Sydney), 0.90 (New York City), and 0.89 (Porto Alegre), with corresponding IoU scores of 0.80, 0.78, 0.84, 0.81, and 0.79 respectively. These results indicate that the model performed reliably and with high accuracy across a range of urban-ecological transition zones with complex land cover compositions.

From a visual inspection perspective (as show in Fig. 18),

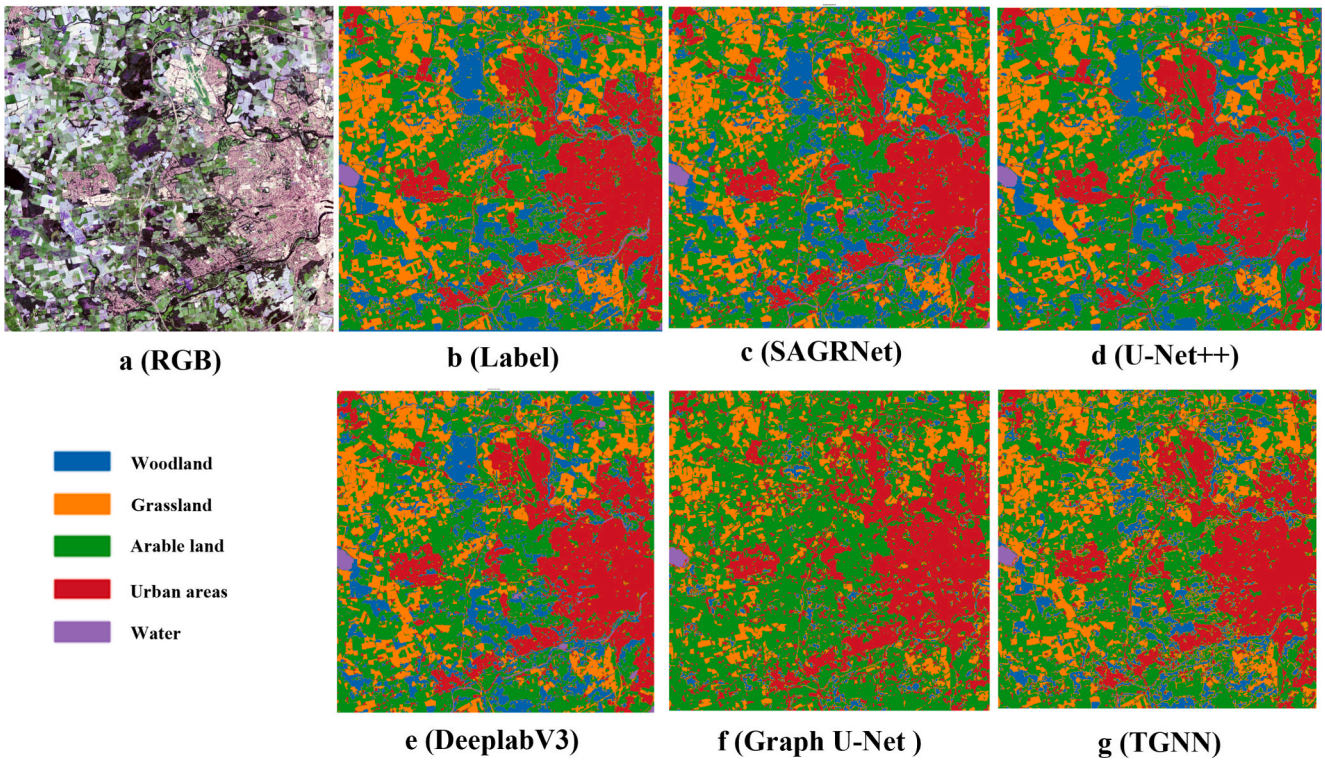

Fig. 11. Comparison of classification results of different algorithms (Region 1). Where a is the original image visualization, and b to g represent the classification results based on SAGRNet, U-Net++, DeepLabV3, Graph U-Net, and TGNN, respectively.

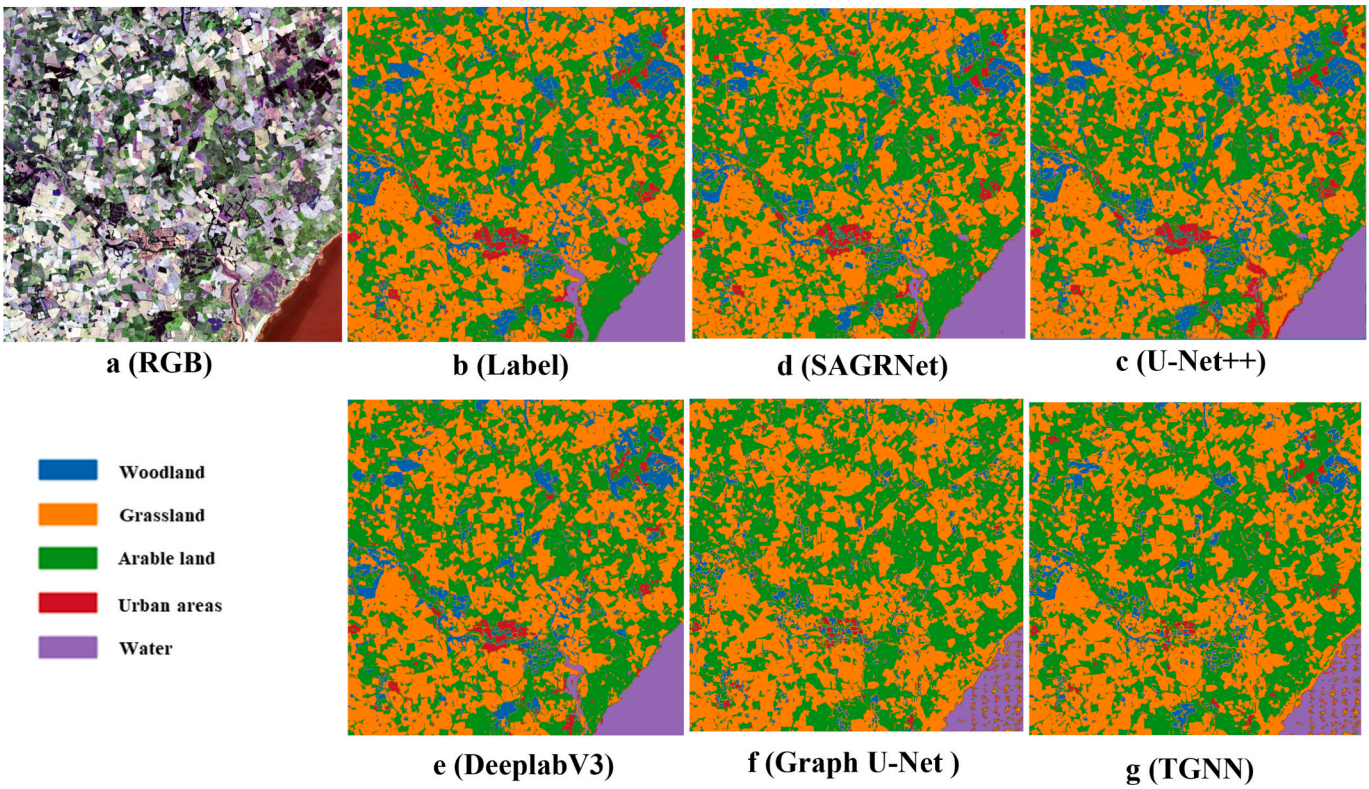

Fig. 12. Comparison of classification results of different algorithms (Region 2). Where a is the original image visualization, and b to g represent the classification results based on SAGRNet, U-Net++, DeepLabV3, Graph U-Net, and TGNN, respectively.

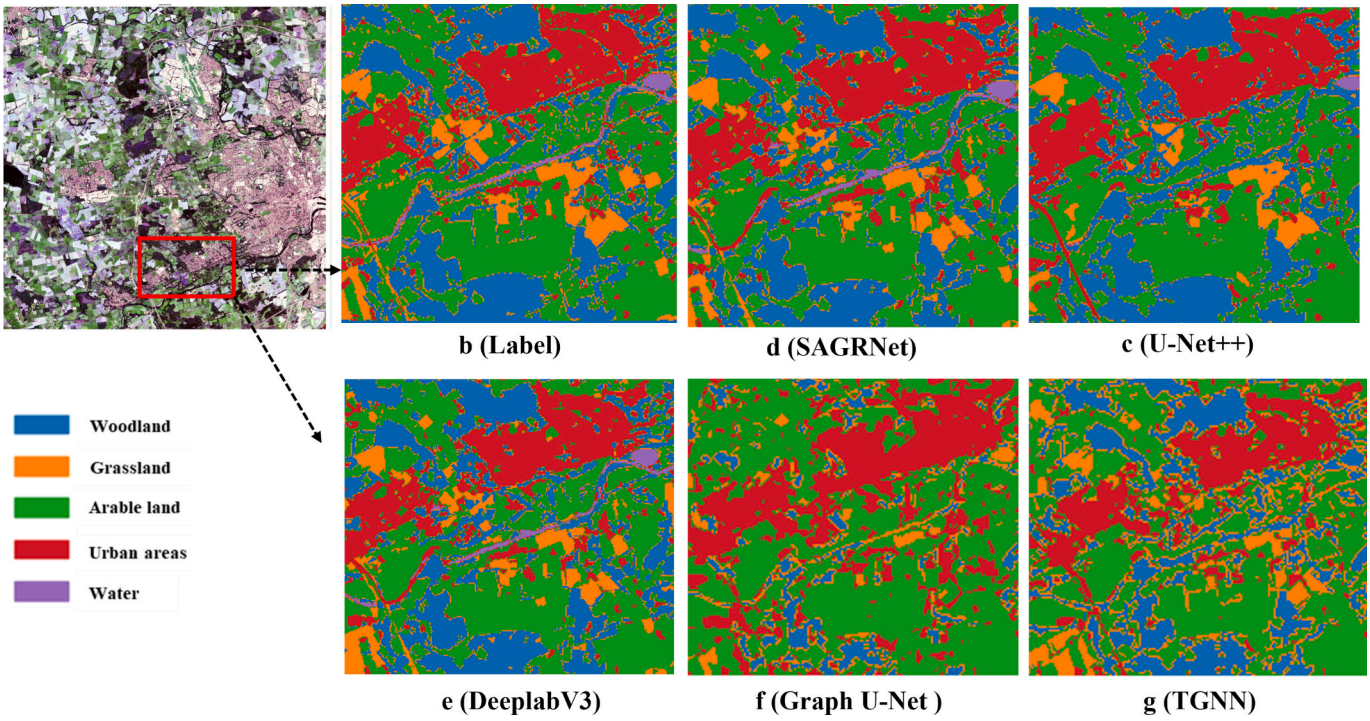

Fig. 13. Comparison of classification results of different algorithms (partial region comparison in region 2). Where a is the original image visualization, and b to g represent the classification results based on SAGRNet, U-Net++, DeepLabV3, Graph U-Net, and TGNN, respectively.

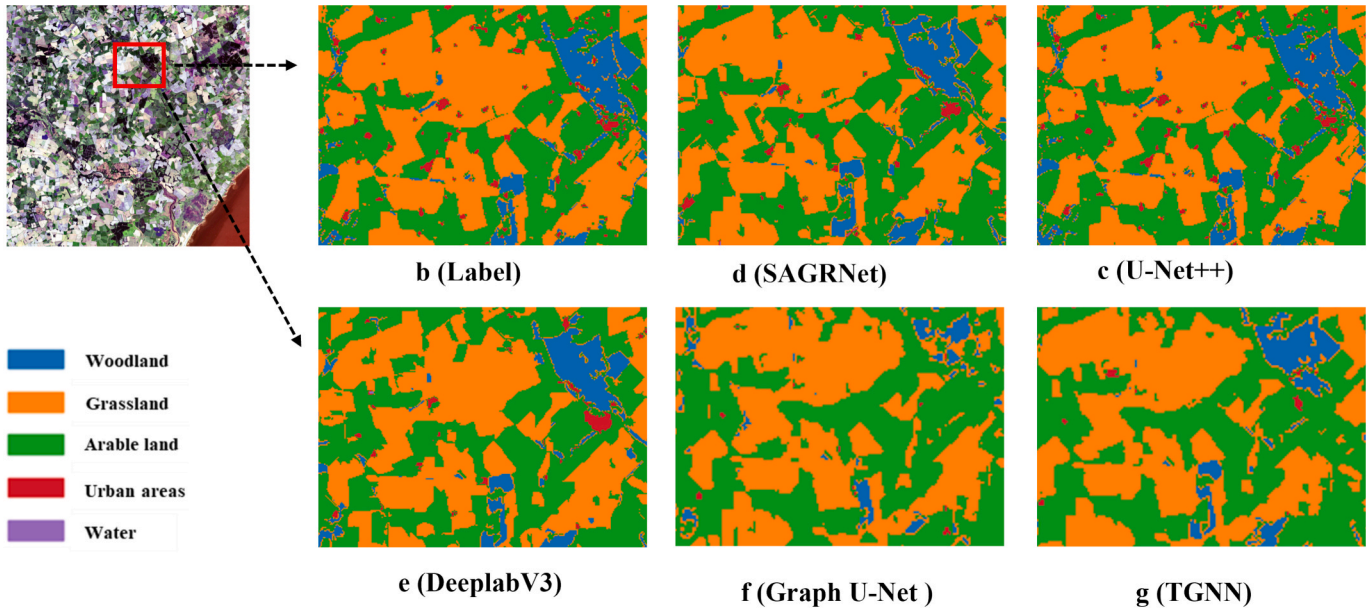

Fig. 14. Comparison of classification results of different algorithms (partial region comparison in region 2). Where a is the original image visualization, and b to g represent the classification results based on SAGRNet, U-Net++, DeepLabV3, Graph U-Net, and TGNN, respectively.

Table 5
Comparison of accuracy ratings of ablation experiments.

SAGEConv	GATConv	ResNet18	Region 1 Time	F1-score	IOU	Region 2 Time	F1-score	IOU
√	√	√	21 min	0.89	0.82	9 min	0.92	0.84
	√	√	17 min	0.83	0.67	7 min	0.86	0.74
√	√		8 min	0.79	0.68	3 min	0.81	0.72
√		√	15 min	0.85	0.70	6 min	0.87	0.75

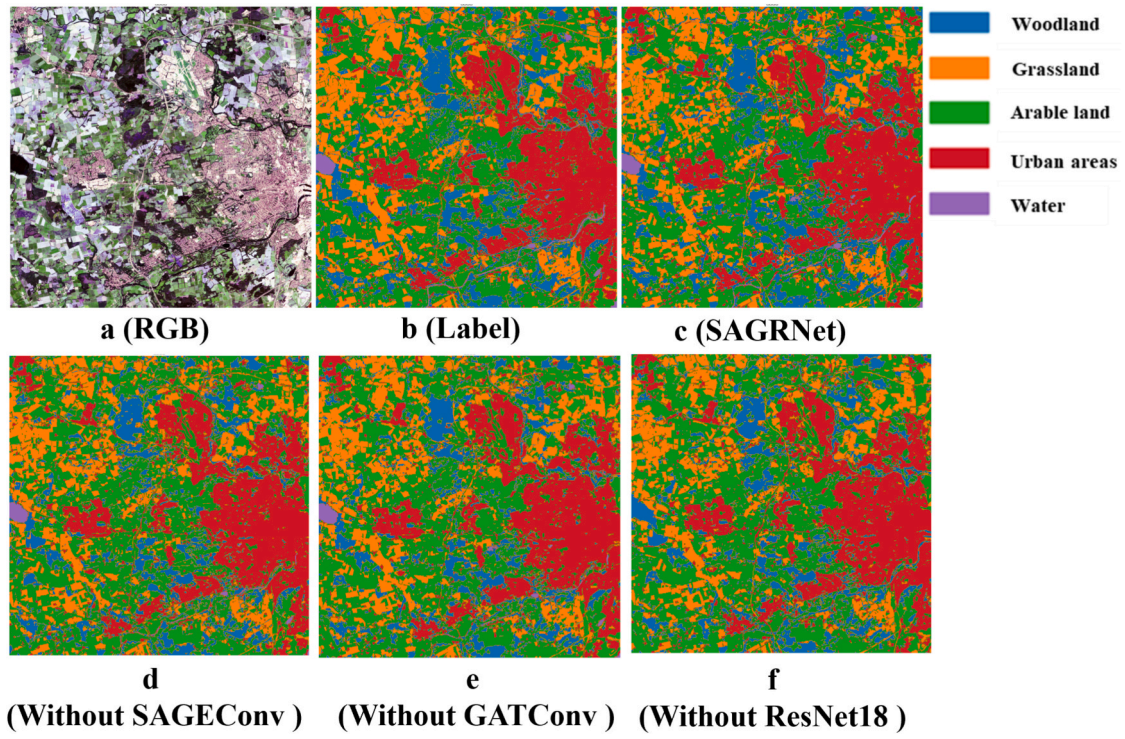

Fig. 15. Comparison of different ablation results of different algorithms (partial region comparison in region 1). Where a is the original image visualization, and b to g represent the classification results of different feature extraction module combinations.

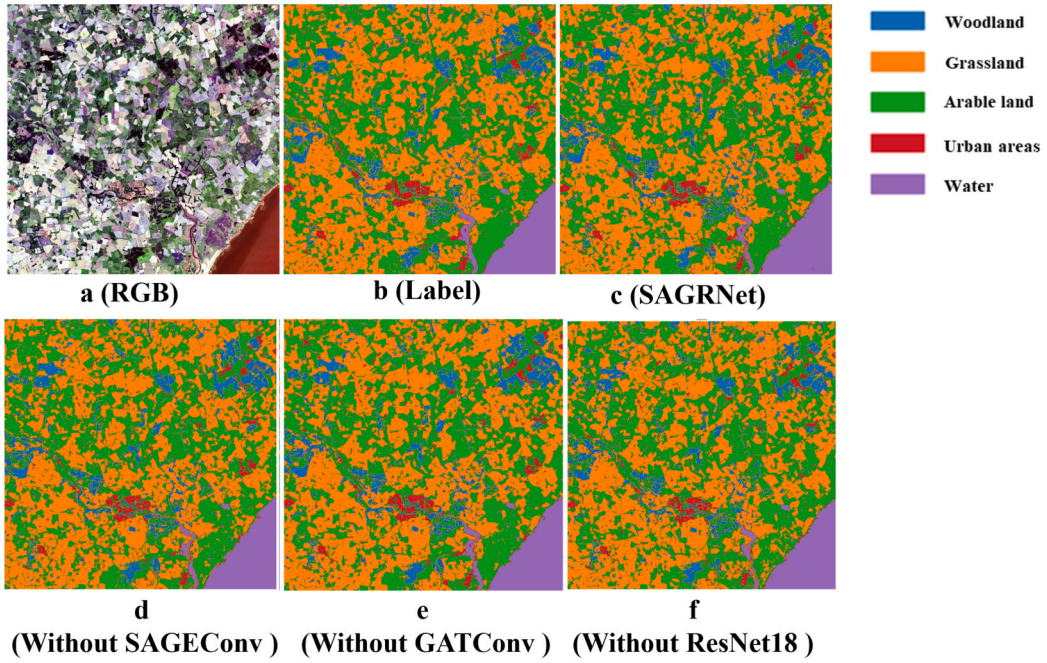

Fig. 16. Comparison of different ablation results of different algorithms (partial region comparison in region 2). Where a is the original image visualization, and b to g represent the classification results of different feature extraction module combinations.

classification errors were found to be closely related to landscape structure and segmentation quality. In Guangzhou, some confusion occurred between subtropical woodlands and managed green spaces, especially where vegetation was dense, and shading was present. Durban exhibited more fragmented patterns, and many errors arose in transitional zones between bare land and sparse grassland, where object boundaries were irregular, and vegetation types intermingled. In Sydney, although the overall layout was relatively structured, several small vegetation clumps within urban and shrubland backgrounds were not effectively segmented, leading to local misclassifications. In New York City, most areas were clearly distinguishable due to well-defined urban, woodland, and arable zones, though occasional confusion still occurred between wetlands and adjacent woodlands. In Porto Alegre, while urban regions were clearly detected, the visual distinction

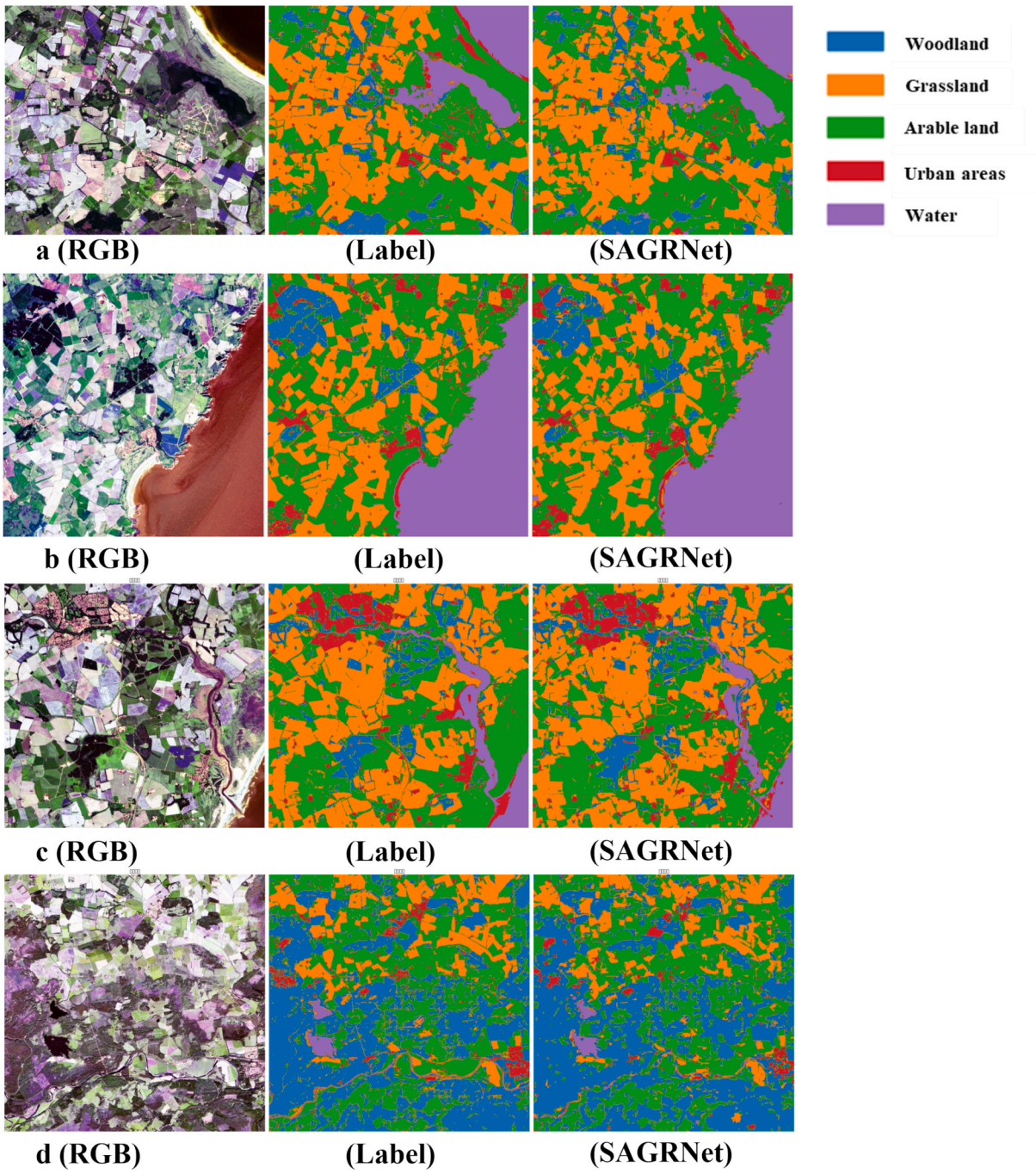

Fig. 17. Classification results of the four additional study areas. Where a to d respectively indicate the classification results of the verification areas region 3 to 6.

between arable land, grassland, and woodland remained challenging due to similar spectral characteristics in mixed areas. Across all five regions, a recurring issue was the presence of small or isolated vegetation objects that were not captured well by the segmentation algorithm, resulting in boundary ambiguity and reduced object-level classification precision. In addition, we observed that part of the classification deviation was associated with systematic errors in the label data itself. Although the Esri Global Land Cover dataset provides relatively high-quality labels, some boundaries between classes—especially at urban-natural transition zones—exhibited excessive smoothing or generalized transitions. This often led to ambiguity near class edges where the model output was visually more precise than the reference label. Such discrepancies, while small, inevitably affect the final accuracy evaluation and should be considered when interpreting classification performance.

5. Discussion

Despite the widespread use of pixel-based deep learning for image semantic segmentation, achieving both high accuracy and short time consumption remains a critical goal for practical applications. Networks that perform well in both areas hold great promise for real-world use. The SAGRNet proposed in this study offers significant advantages for vegetation cover mapping, particularly in complex and heterogeneous landscapes. By leveraging its object-based approach and multi-scale feature fusion, the model excels at distinguishing subtle variations in vegetation types that are often challenging for pixel-based methods. where it achieves higher F1-scores and IoU values. Compared to other

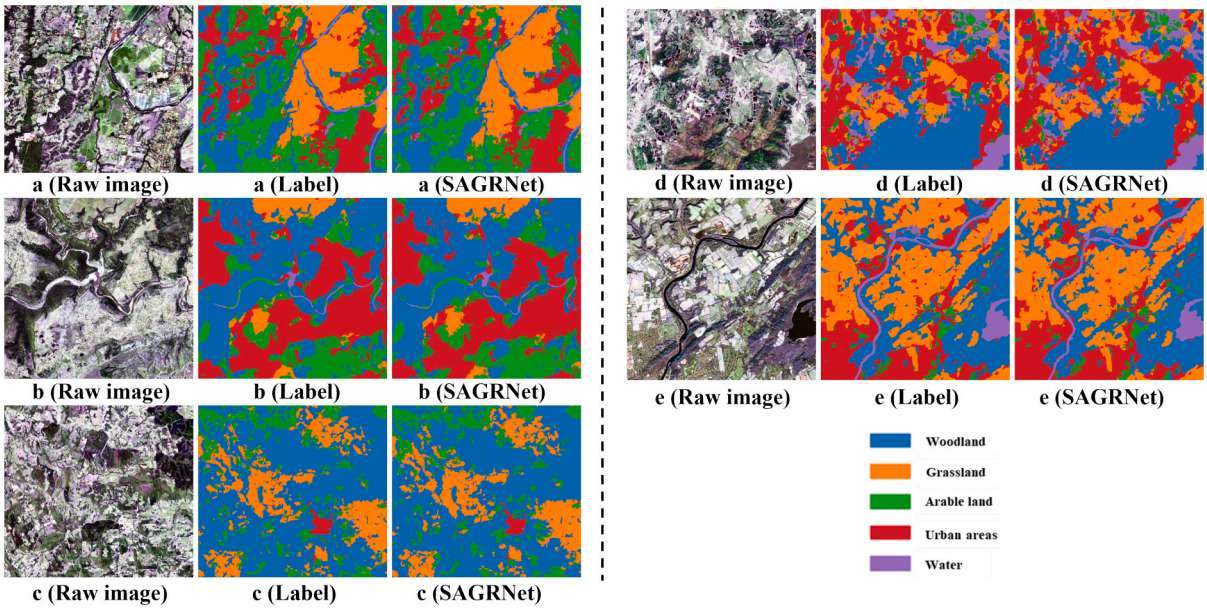

Fig. 18. Graph of the test results for the five study areas. Where a to e indicate the classification results of Guangzhou, Durban, Sydney, New York City, and Porto Alegre, respectively.

models, SAGRNet excels in both classification accuracy and computational efficiency. When compared with pixel-based models like U-Net++ and DeeplabV3, SAGRNet not only shows improvements in accuracy but also performs more efficiently in terms of computation time. Additionally, it exhibits clear application advantages over other object-based graph convolutional networks. These strengths were further validated through ablation testing, and the model's generalization and applicability offer strong evidence of its potential for broader application. The following is a detailed discussion and analysis of several important aspects of the model:

5.1. *Label uncertainty and evaluation robustness*

In this study, the training and evaluation of the model relied on land cover labels derived from the Esri 2021 Global Land Cover dataset in global-scale experiments and UKCEH land cover maps. Both datasets are generated through deep learning and conventional pixel-based classification algorithms. Although they are rigorously constructed and demonstrate high overall accuracy (exceeding 90 % in the case of UKCEH), they inevitably exhibit systematic labelling errors, particularly in areas where land cover classes are spatially heterogeneous or where class boundaries are transitional and fuzzy, such as in urban-natural interface zones.

These systematic errors may lead to mismatches between the actual land cover structure and the labelled data used for model supervision. Such inconsistencies can affect both the training process—by introducing noise in the optimization targets—and the evaluation results—by penalizing predictions that are visually or structurally correct but deviate from imperfect reference labels.

To mitigate the impact of such label-related noise on model performance, we adopted several strategies: (1) a random and spatially uniform sampling approach was used to extract 70 % of labelled samples for training, which reduces the influence of localized mislabelling; (2) model learning was conducted at the object level using aggregated features from spatially coherent regions, which inherently suppresses the effects of random pixel-level noise; and (3) the model is optimized to approximate the dominant distribution patterns in the training data rather than directly memorizing local label artifacts. Furthermore, stratified validation based on the remaining 30 % of samples ensured that performance metrics are robust and not disproportionately affected by specific regions of poor label quality.

Despite these measures, we acknowledge that certain small-scale misclassifications may persist, particularly where ground truth boundaries are imprecise or smoothed. In such cases, it is possible that the model's predictions reflect more precise class transitions than the reference labels themselves. This highlights an inherent limitation in supervised remote sensing tasks where ground truth quality may impose an upper bound on achievable accuracy.

It is also important to note that our objective differs from conventional model development focused solely on benchmark datasets. SAGRNet is designed with practical application scenarios in mind, where real-world labels are often imperfect, and regionally specific discrepancies are common due to acquisition time, vegetation cycle, and land management practices. We emphasize algorithm deployment and scalability across diverse regions, rather than optimization on a fixed test set. In such contexts, the challenge lies in balancing training label alignment with real-world consistency—a point also raised in recent literature (Tong et al., 2025; Valle et al., 2023).

In practice, we observed that SAGRNet occasionally produces classification results that deviate slightly from the training labels but align more closely with visual ground patterns, particularly near fuzzy class boundaries. This is consistent with the notion that well-structured models may partially absorb or compensate for systemic label noise, leading to reduced metric performance but improved semantic consistency (Kumar et al., 2024). Such behavior underlines the necessity of interpreting classification metrics alongside visual validation and contextual understanding, especially in large-area vegetation mapping tasks where perfect reference labels are unattainable. Future work could explore the integration of multi-source label refinement, uncertainty-aware training schemes, or semi-supervised learning to further alleviate these issues.

5.2. *Feature attribute selection*

Traditional object-based testing serves as a crucial foundation for this study, offering a theoretical framework and guidance for validating the first feature screening. We conducted a detailed analysis of various mainstream feature computation methods and identified 12 key features that significantly influence object-based classification. While some characteristics are prominent in conventional approaches, we do not

directly assume that these characteristics would inherently be beneficial in novel models. This phase of the research offers a preliminary screening framework to direct feature selection for future practical implementations. It also enables us to further ascertain if these features can influence the ultimate object-based categorisation outcomes throughout the model creation phase. The significance of the attributes may be constantly assessed and modified in subsequent tests and practical application optimisations to ensure their flexibility. This step establishes the foundational framework for proposing feature selection for the algorithm by identifying features closely associated with the vegetation classification task, minimising data redundancy and the inclusion of irrelevant features, thus enhancing computational efficiency and mitigating the risk of overfitting. Therefore, subsequent research can continue to explore the impact of multiple object features on different parameters or model settings to explore richer and more accurate methods for calculating initial feature attributes.

While SHAP analysis was used in this study to identify the most informative features and reduce data redundancy, we acknowledge that the current feature selection process may involve a degree of subjectivity. Nevertheless, we emphasize that the pool of features initially constructed in our study already represents a comprehensive collection of attribute types commonly used in remote sensing-based object classification tasks. These features were selected based on extensive literature review and prior benchmarking results, ensuring a broad and balanced representation of object characteristics.

However, we recognize that it is possible for alternative features—such as object shape descriptors, multi-angle reflectance properties, or contextual indices—to further enhance the discriminative power of specific vegetation types. Exploring such extensions could be a promising direction for future research.

Regarding the impact of feature selection on model performance, there is an important trade-off between classification accuracy and computational efficiency. While including more features may increase the representational capacity, it also introduces risks of overfitting and significantly increases both feature computation time and memory consumption. Conversely, aggressive feature pruning may improve speed but risk discarding valuable semantic information. Our current use of SHAP aims to strike a balance between these extremes. Future work could explore adaptive feature selection strategies that jointly optimize classification performance and computational overhead.

5.3. Interpretation of model performance

In the SAGRNet model, each module collaborates to enhance its performance in remote sensing image classification tasks. SAGEConv contributes to runtime savings and maintains classification accuracy through efficient sampling and neighbourhood aggregation. Instead of aggregating the entire neighbourhood, as in traditional graph convolution, SAGEConv uses a “sampling” approach, selecting a subset of a node’s neighbours for aggregation. This significantly reduces computational load while preserving rich information, allowing for faster training. Additionally, SAGEConv’s aggregation operations (e.g., mean-value aggregation) progressively integrate more contextual information through multi-layer convolution, resulting in diverse and detailed node representations within the graph structure. This layer-by-layer aggregation not only preserves important features but also captures global structural information, thereby enhancing classification robustness and accuracy.

GATConv further improves model accuracy through its self-attention mechanism. While traditional graph convolution assigns equal weight to all neighbouring nodes, GATConv dynamically adjusts the influence of each neighbour by calculating attention weights between the target node and its neighbours. This allows the model to prioritize more important nodes, making it highly flexible in handling regions with complex and heterogeneous graph structures. This self-attention mechanism significantly boosts the model’s generalization ability and classification accuracy, particularly in areas with substantial structural differences.

ResNet18, inspired by pixel-based convolution, effectively addresses the issue of gradient vanishing in deep networks through residual connections, ensuring stable feature transfer and preventing information loss. It also enhances feature extraction by fusing shallower and deeper layers, thus preserving detailed information. Compared to traditional graph convolutional networks, ResNet18 is more efficient in processing graph-structured data, and its deeper convolutional design further strengthens feature extraction, allowing the model to capture more complex graph-structured features and improve classification performance. In the overall model, the combination of SAGEConv, GATConv, and ResNet18 ensures comprehensive aggregation, enhancement, and retention of information throughout the layer-by-layer transfer process. This synergy enhances the model’s performance.

5.4. Model sensitivity analysis

To further analyse the sensitivity of the proposed model to various parameters and environmental scenarios, the model’s adaptability was tested and evaluated from three perspectives.

5.4.1. The effect of the minimum size setting of the initial segmented object on the model

In the study of object-based GCN for vegetation remote sensing land classification, the minimum size setting of the initial segmented objects is a crucial parameter that significantly impacts model performance. To assess its effect on computation time and classification accuracy, we conducted tests in two different environments of the same area: a complex region with more urban coverage (shown in Fig. 1 as region1) and a simpler vegetation region with less urban coverage (shown in Fig. 1 as region2). We tested a range of minimum size parameters—100, 80, 60, 40, 20, 10, 5, and 2—for the initial segmentation objects in these different environments and statistically analysed how varying these parameters influenced computation time and accuracy. The results, depicted in Fig. 19, show that the minimum size parameter has a pronounced effect in complex regions with higher urban coverage. Specifically, as the minimum size decreases, the training time increases exponentially. While classification accuracy improves with smaller object sizes, this gain is offset by a sharp rise in computational costs. In contrast, vegetation regions with less urban coverage are less sensitive to the minimum size setting. Although training time still increases as object size decreases, the rate of increase is significantly lower compared to the complex region. Similarly, the improvement in classification accuracy is not as substantial. After the minimum size reaches 10, the growth in accuracy begins to level off, indicating that in simpler vegetation regions, the benefits of finer object segmentation are limited, while the computational cost remains relatively stable.

As the minimum size parameter decreases, segmented objects become smaller, allowing the model to capture more subtle local features. This detailed segmentation is particularly important for improving performance in complex urban environments. Urban areas are densely structured with buildings, roads, and green spaces, featuring fuzzy and diverse boundaries. In these heterogeneous regions, larger object segmentation can mix pixels from different categories into a single object, leading to inaccurate classification. Smaller segmented objects, on the other hand, can better capture local variations and details, improving classification accuracy. In complex urban environments, feature types often have similar spectral characteristics, requiring the model to meticulously distinguish small differences. This level of differentiation relies on finer object segmentation to enhance detail resolution. By contrast, in simpler vegetation areas, feature textures tend to be more regular and well-defined. For example, the boundaries of farmland, vegetation, and water bodies are typically clearer, making larger segmented objects sufficient to capture the main characteristics of these categories. In such cases, overly fine

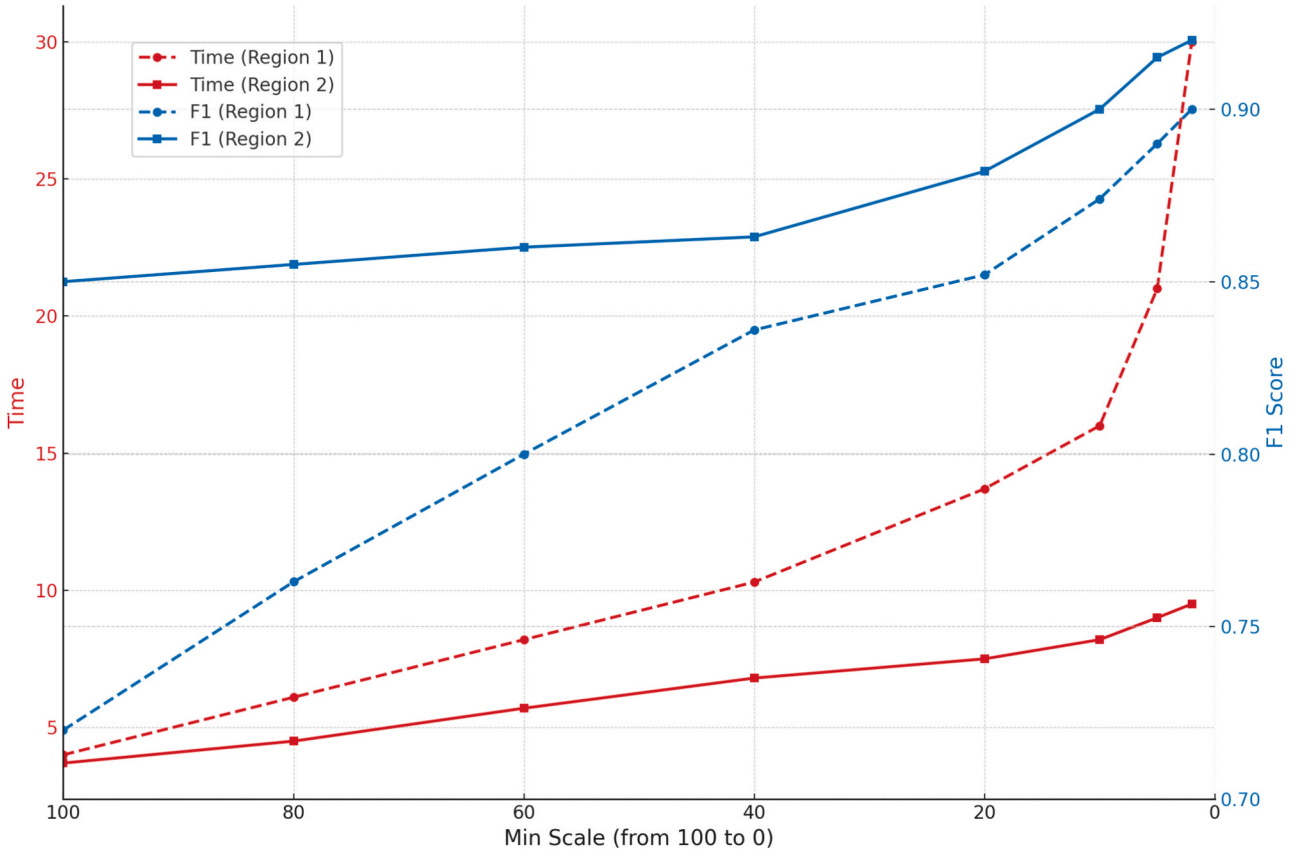

Fig. 19. Comparison of time and F1 score across two regions with varying min scale parameters.

segmentation does not significantly enhance classification accuracy, as larger objects can already classify distinct features effectively. Moreover, smaller object sizes in these simpler environments may introduce unnecessary computational overhead without providing substantial gains in accuracy.

The minimum size of the initial segmentation object not only influences classification accuracy but also directly impacts the computational cost of the GCN. As the size of the segmented objects decreases, the number of objects increases dramatically, affecting both graph construction and feature computation. With more segmented objects, the number of graph nodes rises accordingly, and the connections (edges) between nodes grow exponentially. The feature attributes (e.g., spectra, textures) and neighbourhoods of each node require more intensive computation, leading to a sharp increase in processing time. Each segmented object corresponds to one node, and as the number of objects grows, the dimension of the feature matrix expands. In larger object segmentations, edge effects are relatively weak, as the objects typically contain more complete class features. However, as the segmented objects become smaller, edge pixels and mixed categories become more prominent, placing greater demands on local feature extraction in the GCN. While smaller objects capture more detailed local features, they may also introduce inconsistencies at boundaries and increase noise. In complex environments, such local inconsistencies necessitate higher computational power to handle diverse edge information. Conversely, in vegetation environments—where edges are typically more regular and well-defined—the model requires less robustness to handle edge effects. Therefore, overly fine segmentation in these areas increases computational cost without proportional accuracy improvements. A core feature of GCNs is their ability to propagate information through the neighbourhood matrix, capturing spatial dependencies between nodes. In urban environments, smaller segmented objects imply more intricate interdependencies between nodes, requiring the model to handle longer chains of dependencies. In contrast, dependencies in vegetation areas are usually more localized and direct, meaning larger segmented objects are sufficient to capture these spatial relationships.

In practical applications, a balance between accuracy and computational efficiency should be struck based on the specific scenario. In complex urban environments, a moderate reduction in object size can improve accuracy, while in simpler vegetation areas, smaller object sizes may introduce unnecessary computational burden. Therefore, a larger object size is recommended to optimize computational efficiency. In the future, adaptive segmentation strategies could be explored to automatically adjust object sizes based on environmental complexity, ensuring both accuracy and resource efficiency. Additionally, combining different scales of segmented objects for feature fusion may further enhance the model's performance in complex environments.

5.4.2. The effect of neighbourhood distance of graph construction on the model

The effect of neighbourhood distances in graph construction, particularly in models based on GCN, is an important issue that warrants in-depth exploration. Expanding the neighbourhood distance allows each node to access information from nodes farther away, theoretically providing more context. However, according to our tests, this additional information from distant neighbouring nodes did not improve the model's accuracy.

Our analysis suggests that when nodes aggregate features from surrounding nodes, the added information may not always be relevant and can introduce noise. In the context of remote sensing image classification, where object sizes are typically large, objects from different categories often have distinct spectral features. Enlarging the neighbourhood can cause the model to mix features from different categories, ultimately reducing classification accuracy. Furthermore, as

more node features are incorporated into training, model complexity increases. With limited training data, the model risks capturing noise or subtle changes specific to the dataset, leading to overfitting. Expanding the neighbourhood distance thus increases the likelihood of learning irrelevant features without gaining useful information. Additionally, the nature of graph convolution involves local smoothing of node features. When the neighbourhood distance becomes too large, node features may become overly similar, reducing the model's ability to differentiate between nodes. In remote sensing image classification, this manifests as blurred boundaries, making it difficult for the model to capture subtle differences between categories accurately.

Nevertheless, this does not imply that more distant neighbourhoods should be entirely disregarded. In future research, incorporating jump connections or mixing local and global features could help integrate neighbourhood information at different scales. This approach would allow for expanded neighbourhoods while maintaining fine-grained representations of local features, potentially improving the model's accuracy.

5.4.3. The impact of ResNet depth on the model

To study the effect of ResNet depth on model performance, we designed four additional models: Res34, Res50, Res101, and Res152, along with the baseline Res18. As expected, the time cost for training increased dramatically with network depth. However, we observed that increasing the depth of ResNet did not significantly improve the model's accuracy. In some cases, deeper networks even led to a decrease in performance. This suggests that for large-scale remote sensing tasks, simply increasing network depth may not be the optimal approach, especially when the accuracy gains are marginal, and the additional depth introduces diminishing returns. In future work, it may be more effective to incorporate advanced algorithms such as ResNeXt, DenseNet, or EfficientNet, which optimize model performance by adaptively balancing network depth and width, introducing attention mechanisms, or leveraging multi-layer feature fusion. Exploring these cutting-edge techniques could help avoid performance degradation while reducing time costs, ensuring that the model remains both efficient and accurate when handling large-scale remote sensing image classification tasks.

5.5. Advantages

In this study, the proposed object-based model demonstrates significant advantages in several aspects, particularly in terms of accuracy, time efficiency, and data input flexibility. Firstly, the model optimizes computation time while maintaining high accuracy. While some deep learning models achieve improved accuracy by increasing network complexity, this often comes at the cost of significant computational overhead. In contrast, this model achieves efficient classification tasks with reduced computation time through a well-designed architecture, without compromising accuracy.

One of the model's key advantages is the simplicity of data input. Traditional pixel-based models typically require images to be adjusted to a fixed size, involving strict preprocessing and segmentation, which can introduce several issues. In pixel-based models, images are often cropped or scaled, leading to potential loss of useful information or pixel distortion. By contrast, the object-based model eliminates these strict size constraints, retaining complete image information and avoiding preprocessing errors at the pixel level. This flexibility allows the model to handle input data of varying sizes and scales more effectively, thanks to the inherent structure of SAGRNet. In GCNs, relationships between objects are expressed through graph nodes and their neighbourhood information, unlike CNN, which require a fixed grid and local region-based convolution operations. In GCN, the convolution operation is based on the neighbourhood of graph nodes, where each node updates its features using information from neighbouring nodes. This flexibility enables the SAGRNet to adapt to different image sizes during training and prediction, without the need for complex preprocessing, which is often required in pixel-based CNN.

Additionally, the five urban-fringe test regions – Guangzhou, Durban, Sydney, New York City, and Porto Alegre – spanned a wide spectrum of landscape heterogeneity, and the results indicate that this heterogeneity has a tangible impact on classification performance. Heterogeneous areas like Durban (with its fine-scale fragmented grassland–savannah mosaic) and Porto Alegre (with spectrally overlapping vegetation classes) posed greater challenges, showing slightly lower IoU scores (~0.78–0.79) compared to more structured landscapes. In contrast, regions with more homogeneous or organized layouts (e.g. Sydney and New York, where urban and vegetated zones are spatially distinct) achieved higher accuracy (IoU up to ~0.84). Nonetheless, SAGRNet's performance remained robust across all sites, with F1-scores hovering around 0.90 in each region. This consistency underscores the advantages of the SAGRNet model's structural flexibility and spatial-contextual learning ability: its graph-based architecture can accommodate irregular object shapes/sizes and leverage neighbourhood context to better disambiguate classes in complex scenes. In practice, these design strengths (including multi-scale feature fusion and self-attention) allowed SAGRNet to distinguish subtle vegetation differences even in areas with intermingled or transitional land cover.

5.6. Limitations and future work

While SAGRNet demonstrates strong performance in terms of accuracy and efficiency across diverse vegetation types and geographic regions, several limitations should be acknowledged. First, the model's dependence on an external image segmentation algorithm introduces sensitivity to the segmentation quality. Our experiments revealed that overly coarse segmentation may lead to the loss of small or fragmented vegetation patches, whereas overly fine segmentation can produce jagged boundaries and excessive noise. Currently, segmentation parameters require manual adjustment across scenes. A promising future direction is to develop an end-to-end pipeline with adaptive segmentation parameter tuning, though this must be balanced against potential increases in model complexity and training cost.

Second, although the graph-based architecture of SAGRNet improves spatial contextual reasoning, its scalability to very large-scale remote sensing datasets (e.g., country-level or continental mosaics) may be constrained by graph memory and computational complexity. Hierarchical graph representations or regional subgraph partitioning strategies could be incorporated to extend its applicability in large-area mapping tasks.

Third, while SHAP-based feature importance analysis improves interpretability to some extent, the model's internal decision mechanisms remain complex and opaque. Future efforts could explore attention heatmaps, gradient-based saliency, or contrastive explainability frameworks to provide more intuitive explanations of classification behaviour (An and Joe, 2022; Ismail et al., 2021).

Otherwise, SAGRNet, like most supervised models, depends on deterministic training labels. In scenarios where reference labels exhibit spatial ambiguity or semantic inconsistency, uncertainty-aware training frameworks and weakly supervised methods may help improve robustness and better reflect true class boundaries (Ren et al., 2023; Tabarisaadi et al., 2024).

Finally, some limitations were observed in the most complex landscapes – for example, small or isolated vegetation patches in highly fragmented areas were occasionally missed by the initial segmentation, leading to boundary ambiguity and local misclassifications. To address such cases, targeted improvements could be explored: an adaptive segmentation refinement strategy, which dynamically adjusts object granularity to capture fine-scale patches, would likely reduce segmentation omissions (Farooq and Bazaz, 2020; Zhang et al., 2020). Similarly, parameter tuning for complex regions (e.g. refining graph connectivity or model thresholds in fragmented environments) may enhance classification of difficult transitional zones. Looking forward, future work

might implement multi-scale segment fusion and region-specific architectural tuning – for instance, tailoring certain model components or training hyperparameters to a region's characteristic landscape structure – to further boost SAGRNet's performance in high-heterogeneity areas. Such adaptive refinements, combined with SAGRNet's inherent spatial-context learning, promise even more robust vegetation mapping in the face of complex, fragmented landscapes (Gui et al., 2025).

6. Conclusion

This study proposed SAGRNet, a lightweight and efficient object-based graph convolutional neural network for vegetation cover classification. By integrating multiple GCNs and feature extraction complementary modules within a unified object-centric structure, the model effectively captures both spectral signatures and spatial-contextual dependencies to enhance classification robustness and efficiency. Experimental results across northeastern Scotland and diverse global regions demonstrate that SAGRNet achieves superior classification accuracy and significantly faster training time compared to baseline models. The model shows strong generalization across ecological zones with varying landscape complexity, vegetation types, and label qualities.

SAGRNet is particularly applicable to a wide range of land cover classification tasks. It shows excellent adaptability in structured environments such as agricultural and managed lands, while also maintaining high robustness in complex and irregular surface landscapes like urban fringes and mixed vegetation zones. These characteristics make it suitable for integration into large-scale applications such as land resource surveys, ecological monitoring platforms, national land cover mapping programs, and environmental change analysis frameworks.

Looking forward, future work may explore extensions of SAGRNet to multi-temporal and multi-sensor imagery, enabling seasonal and long-term vegetation dynamics monitoring. The incorporation of adaptive segmentation, uncertainty-aware learning, and hierarchical graph structures may enhance performance on large-scale datasets and under uncertain labelling conditions. Additionally, embedding the model within cloud-native or edge-computing frameworks could further support its deployment for nationwide ecological surveillance and global-scale land cover monitoring.

Overall, SAGRNet offers a robust, interpretable, and deployable framework for next-generation vegetation classification and geospatial intelligence in remote sensing.

CRediT authorship contribution statement

Baoling Gui: Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lydia Sam:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Data curation, Conceptualization. **Anshuman Bhardwaj:** Writing – review & editing, Supervision, Resources, Methodology, Data curation, Conceptualization. **Diego Soto Gómez:** Writing – review & editing. **Félix González Peñaloza:** Writing – review & editing. **Manfred F. Buchroithner:** Writing – review & editing. **David R. Green:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors acknowledge BBSRC International Partnership Award (Ref: RG17324-19) and the University of Aberdeen to fund the research activities. We would like to thank the UK Centre for Ecology and Hydrology (UKCEH) for providing the high-resolution land cover maps of the United Kingdom, which served as an essential reference in our regional experiments. We also express our appreciation to ESRI for releasing the Global Land Cover dataset through the Living Atlas platform, which enabled the global-scale validation of our model. These openly accessible and high-quality datasets significantly supported the development and evaluation of our work.

References

Abbaszad, P., Asadzadeh, F., Rezapour, S., Khosravi Aqdam, K., Shabani, F., 2024. Evaluation of Landsat 8 and Sentinel-2 vegetation indices to predict soil organic carbon using machine learning models. Model. Earth Syst. Environ. 10, 2581–2592. https://doi.org/10.1007/s40808-023-01916-x.

Ajayi, O.G., Ashi, J., Guda, B., 2023. Performance evaluation of YOLO v5 model for automatic crop and weed classification on UAV images. Smart Agric. Technol. 5, 100231. https://doi.org/10.1016/j.atech.2023.100231.

Alcantara, C., Kuemmerle, T., Prishchepov, A.V., Radeloff, V.C., 2012. Mapping abandoned agriculture with multi-temporal MODIS satellite data. Remote Sens. Environ. 124, 334–347. https://doi.org/10.1016/j.rse.2012.05.019.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8, 53. https://doi.org/10.1186/s40537-021-00444-8.

An, J., Joe, I., 2022. Attention map-guided visual explanations for deep neural networks. Appl. Sci. 12, 3846. https://doi.org/10.3390/app12083846.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65, 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004.

Campos-Taberner, M., Javier García-Haro, F., Martínez, B., Sánchez-Ruiz, S., Moreno-Martínez, Á., Camps-Valls, G., Amparo Gilabert, M., 2023. Land use classification over smallholding areas in the European Common Agricultural Policy framework. ISPRS J. Photogramm. Remote Sens. 197, 320–334. https://doi.org/10.1016/j.isprsjprs.2023.02.005.

Chen, G., Zhang, Y., Cai, Z., Li, X., 2021. The building recognition and analysis of remote sensing image based on depth belief network. Cogn. Syst. Res. 68, 53–61. https://doi.org/10.1016/j.cogsys.2021.02.002.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Presented at the Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818.

Chen, R., Vivone, G., Li, G., Dai, C., Chanussot, J., 2024. Multi-scale feature learning via residual dynamic graph convolutional network for hyperspectral image classification. Int. J. Remote Sens. 45, 863–888. https://doi.org/10.1080/01431161.2024.2305179.

Chen, T., Guestrin, C., 2016. XGBoost: a Scalable tree Boosting System, in. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Dingle Robertson, L., King, D.J., 2011. Comparison of pixel- and object-based classification in land cover change mapping. Int. J. Remote Sens. 32, 1505–1529. https://doi.org/10.1080/01431160903571791.

Dobson, J.E., 2023. On reading and interpreting black box deep neural networks. Int J Digit Humanities 5, 431–449. https://doi.org/10.1007/s42803-023-00075-w.

Du, H., Li, W., Cardellach, E., Ribó, S., Rius, A., Nan, Y., 2024. Deep residual fully connected network for GNSS-R wind speed retrieval and its interpretation. Remote Sens. Environ. 313, 114375. https://doi.org/10.1016/j.rse.2024.114375.

Duda, R., Hart, P., Stork, G.D., 2001. Pattern Classification. Wiley Interscience.

Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sens. Environ. 118, 259–272. https://doi.org/10.1016/j.rse.2011.11.020.

Farooq, J., Bazaz, M.A., 2020. A novel adaptive deep learning model of Covid-19 with focus on mortality reduction strategies. Chaos, Solitons Fractals 138, 110148. https://doi.org/10.1016/j.chaos.2020.110148.

Felzenszwalb, P.F., Huttenlocher, D.P., 2004. Efficient graph-based image segmentation. Int. J. Comput. Vis. 59, 167–181. https://doi.org/10.1023/B:VISI.0000022288.19776.77.

Gao, H., Ji, S., 2019. Graph U-Nets [WWW Document]. arXiv.org. URL https://arxiv.org/abs/1905.05178v1 (accessed 9.5.24).

Gui, B., Bhardwaj, A., Sam, L., 2025. Comparative analysis of different machine learning algorithms for urban footprint extraction in diverse urban contexts using high-resolution remote sensing imagery. J. Geogr. Sci. 35, 664–696. https://doi.org/10.1007/s11442-025-2339-y.

Gui, B., Bhardwaj, A., Sam, L., 2024a. Evaluating the efficacy of segment anything model for delineating agriculture and urban green spaces in multiresolution aerial and spaceborne remote sensing images. Remote Sens. (Basel) 16, 414. https://doi.org/10.3390/rs16020414.

Gui, B., Bhardwaj, A., Sam, L., 2024b. A novel automated labelling algorithm for deep learning-based built-up areas extraction using nighttime lighting data. Knowl.-Based Syst. 306, 112702. https://doi.org/10.1016/j.knosys.2024.112702.

Gui, B., Sam, L., Bhardwaj, A., 2024c. From roofs to renewables: deep learning and geographic information systems insights into a comprehensive urban solar

photovoltaic assessment for Stonehaven. Energy 360 (1), 100006. https://doi.org/10.1016/j.energ.2024.100006.
Hamilton, W.L., Ying, R., Leskovec, J., 2017. Inductive Representation Learning on Large Graphs [WWW Document]. arXiv.org. URL https://arxiv.org/abs/1706.02216v4 (accessed 8.30.24).
Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics SMC-3 610–621. https://doi.org/10.1109/TSMC.1973.4309314.
Haralick, R.M., Shapiro, L.G., 1992. Computer and Robot Vision, 1st ed. Addison-Wesley Longman Publishing Co., Inc, USA.
He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. https://doi.org/10.48550/arXiv.1512.03385.
Huang, X., Wang, H., Li, X., 2024. A multi-scale semantic feature fusion method for remote sensing crop classification. Comput. Electron. Agric. 224, 109185. https://doi.org/10.1016/j.compag.2024.109185.
Ismail, A.A., Corrada Bravo, H., Feizi, S., 2021. Improving Deep Learning Interpretability by Saliency Guided Training, in: Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 26726–26739.
Javidan, S.M., Banakar, A., Vakilian, K.A., Ampatzidis, Y., 2023. Diagnosis of grape leaf diseases using automatic *K*-means clustering and machine learning. Smart Agric. Technol. 3, 100081. https://doi.org/10.1016/j.atech.2022.100081.
Jia, X., 2006. Remote Sensing Digital Image Analysis: An Introduction, Remote Sensing Digital Image Analysis: An Introduction. Doi: 10.1007/3-540-29711-1.
Kganyago, M., Adjorlolo, C., Mhangara, P., Tsoeleng, L., 2024. Optical remote sensing of crop biophysical and biochemical parameters: an overview of advances in sensor technologies and machine learning algorithms for precision agriculture. Comput. Electron. Agric. 218, 108730. https://doi.org/10.1016/j.compag.2024.108730.
Khan, H.R., Gillani, Z., Jamal, M.H., Athar, A., Chaudhry, M.T., Chao, H., He, Y., Chen, M., 2023. Early identification of crop type for smallholder farming systems using deep learning on time-series sentinel-2 imagery. Sensors 23, 1779. https://doi.org/10.3390/s23041779.
Kumar, V., Singh, R.S., Rambabu, M., Dua, Y., 2024. Deep learning for hyperspectral image classification: a survey. Computer Science Review 53, 100658. https://doi.org/10.1016/j.cosrev.2024.100658.
Li, M., Ma, L., Blaschke, T., Cheng, L., Tiede, D., 2016. A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. Int. J. Appl. Earth Obs. Geoinf. 49, 87–98. https://doi.org/10.1016/j.jag.2016.01.011.
Li, Z., Chen, B., Wu, S., Su, M., Chen, J.M., Xu, B., 2024. Deep learning for urban land use category classification: a review and experimental assessment. Remote Sens. Environ. 311, 114290. https://doi.org/10.1016/j.rse.2024.114290.
Liu, S., Li, H., Jiang, C., Feng, J., 2024. Spectral–spatial graph convolutional network with dynamic-synchronized multiscale features for few-shot hyperspectral image classification. Remote Sens. (Basel) 16, 895. https://doi.org/10.3390/rs16050895.
Ma, L., Li, M., Ma, X., Cheng, L., Du, P., Liu, Y., 2017. A review of supervised object-based land-cover image classification. ISPRS J. Photogramm. Remote Sens. 130, 277–293. https://doi.org/10.1016/j.isprsjprs.2017.06.001.
Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 11, 674–693. https://doi.org/10.1109/34.192463.
Niu, Z., Liu, W., Zhao, J., Jiang, G., 2019. DeepLab-based spatial feature extraction for hyperspectral image classification. IEEE Geosci. Remote Sens. Lett. 16, 251–255. https://doi.org/10.1109/LGRS.2018.2871507.
Ok, A.O., Akar, O., Gungor, O., 2012. Evaluation of random forest method for agricultural crop classification. Eur. J. Remote Sens. 45, 421–432. https://doi.org/10.5721/EuJRS20124535.
Pan, X., Zhang, C., Xu, J., Zhao, J., 2021. Simplified object-based deep neural network for very high resolution remote sensing image classification. ISPRS J. Photogramm. Remote Sens. 181, 218–237. https://doi.org/10.1016/j.isprsjprs.2021.09.014.
Peña, J.M., Gutiérrez, P.A., Hervás-Martínez, C., Six, J., Plant, R.E., López-Granados, F., 2014. Object-based image classification of summer crops with machine learning methods. Remote Sens. (Basel) 6, 5019–5041. https://doi.org/10.3390/rs6065019.
Pereira, G.W., Valente, D.S.M., de Queiroz, D.M., Santos, N.T., Fernandes-Filho, E.I., 2022. Soil mapping for precision agriculture using support vector machines combined with inverse distance weighting. Precision Agric. 23, 1189–1204. https://doi.org/10.1007/s11119-022-09880-9.
Quan, L., Feng, H., Lv, Y., Wang, Q., Zhang, C., Liu, J., Yuan, Z., 2019. Maize seedling detection under different growth stages and complex field environments based on an improved Faster R–CNN. Biosyst. Eng. 184, 1–23. https://doi.org/10.1016/j.biosystemseng.2019.05.002.
Ren, Z., Wang, S., Zhang, Y., 2023. Weakly supervised machine learning. CAAI Trans. Intell. Technol. 8, 549–580. https://doi.org/10.1049/cit2.12216.
Saini, R., Ghosh, S.K., 2021. Crop classification in a heterogeneous agricultural environment using ensemble classifiers and single-date Sentinel-2A imagery. Geocarto Int. 36, 2141–2159. https://doi.org/10.1080/10106049.2019.1700556.
Shetty, S., Schneider, P., Stebel, K., David Hamer, P., Kylling, A., Koren Berntsen, T., 2024. Estimating surface NO2 concentrations over Europe using Sentinel-5P TROPOMI observations and Machine Learning. Remote Sens. Environ. 312, 114321. https://doi.org/10.1016/j.rse.2024.114321.
Song, X., Hua, Z., Li, J., 2024. Context spatial awareness remote sensing image change detection network based on graph and convolution interaction. IEEE Trans. Geosci. Remote Sens. 62, 1–16. https://doi.org/10.1109/TGRS.2024.3357524.
Su, T., Zhang, S., 2023. Object-based crop classification in Hetao irrigation zone by using deep learning and region merging optimization. Comput. Electron. Agric. 214, 108284. https://doi.org/10.1016/j.compag.2023.108284.
Swain, M.J., Ballard, D.H., 1991. Color indexing. Int. J. Comput. Vision 7, 11–32. https://doi.org/10.1007/BF00130487.
Tabarisaadi, P., Khosravi, A., Nahavandi, S., Shafie-Khah, M., Catalão, J.P.S., 2024. An optimized uncertainty-aware training framework for neural networks. IEEE Trans. Neural Networks Learn. Syst. 35, 6928–6935. https://doi.org/10.1109/TNNLS.2022.3213315.
Tong, X.-Y., Dong, R., Zhu, X.X., 2025. Global high categorical resolution land cover mapping via weak supervision. ISPRS J. Photogramm. Remote Sens. 220, 535–549. https://doi.org/10.1016/j.isprsjprs.2024.12.017.
Twogood, R.E., Sommer, F.G., 1982. Digital image processing. IEEE Trans. Nucl. Sci. 29, 1075–1086. https://doi.org/10.1109/TNS.1982.4336327.
Valle, D., Izbicki, R., Leite, R.V., 2023. Quantifying uncertainty in land-use land-cover classification using conformal statistics. Remote Sens. Environ. 295, 113682. https://doi.org/10.1016/j.rse.2023.113682.
Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y., 2017. Graph Attention Networks [WWW Document]. arXiv.org. URL https://arxiv.org/abs/1710.10903v3 (accessed 8.30.24).
Wang, L., Wang, J., Liu, Z., Zhu, J., Qin, F., 2022. Evaluation of a deep-learning model for multispectral remote sensing of land use and crop classification. The Crop Journal, Crop Phenotyping Studies with Application to Crop Monitoring 10, 1435–1451. https://doi.org/10.1016/j.cj.2022.01.009.
Wang, Q., Chen, W., Tang, H., Pan, X., Zhao, H., Yang, B., Zhang, H., Gu, W., 2023. Simultaneous extracting area and quantity of agricultural greenhouses in large scale with deep learning method and high-resolution remote sensing images. Sci. Total Environ. 872, 162229. https://doi.org/10.1016/j.scitotenv.2023.162229.
Xu, J., Li, K., Li, Z., Chong, Q., Xing, H., Xing, Q., Ni, M., 2024. Fuzzy graph convolutional network for hyperspectral image classification. Eng. Appl. Artif. Intel. 127, 107280. https://doi.org/10.1016/j.engappai.2023.107280.
Yang, J.-Y., Li, H.-C., Yang, J.-H., Pan, L., Du, Q., Plaza, A., 2024. Multifrequency graph convolutional network with cross-modality mutual enhancement for multisource remote sensing data classification. IEEE Trans. Geosci. Remote Sens. 62, 1–14. https://doi.org/10.1109/TGRS.2024.3356510.
Young, I., 1983. Image analysis and mathematical morphology, by J. Serra. Academic Press, London, 1982, xviii + 610 p. $90.00. Cytometry 4, 184–185. https://doi.org/10.1002/cyto.990040213.
Yun, S., Jeong, M., Yoo, S., Lee, S., Yi, S.S., Kim, R., Kang, J., Kim, H.J., 2022. Graph transformer networks: learning meta-path graphs to improve GNNs. Neural Netw. 153, 104–119. https://doi.org/10.1016/j.neunet.2022.05.026.
Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2018. An object-based convolutional neural network (OCNN) for urban land use classification. Remote Sens. Environ. 216, 57–70. https://doi.org/10.1016/j.rse.2018.06.034.
Zhang, C., Wang, B., 2024. Progressive feature fusion framework based on graph convolutional network for remote sensing scene classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 17, 3270–3284. https://doi.org/10.1109/JSTARS.2024.3350129.
Zhang, T., Li, Yu., Li, Y., Sun, S., Gao, X., 2020. A self-adaptive deep learning algorithm for accelerating multi-component flash calculation. Comput. Methods Appl. Mech. Eng. 369, 113207. https://doi.org/10.1016/j.cma.2020.113207.
Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2018. UNet++: A Nested U-Net Architecture for Medical Image Segmentation [WWW Document]. arXiv.org. URL https://arxiv.org/abs/1807.10165v1 (accessed 9.5.24).