Article

Building Function Recognition Using the Semi-Supervised Classification

Xuejing Xie [1,2], Yawen Liu [3], Yongyang Xu [1,2,4,*], Zhanjun He [4], Xueye Chen [1], Xiaoyun Zheng [1] and Zhong Xie [4]

1 Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen 518034, China
2 National Engineering Research Center of Geographic Information System, China University of Geosciences, Wuhan 430074, China
3 State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
4 School of Computer Science, China University of Geosciences, Wuhan 430074, China
* Correspondence: yongyangxu@cug.edu.cn

Abstract: The functional classification of buildings is important for creating and managing urban zones and assisting government departments. Building function recognition is incredibly valuable for wide applications ranging from the determination of energy demand. By aiming at the topic of urban function classification, a semi-supervised graph structure network combined unified message passing model was introduced. The data of this model include spatial location distribution of buildings, building characteristics and the information mined from points of interesting (POIs). In order to extract the context information, each building was regarded as a graph node. Building characteristics and corresponding POIs information were embedded to mine the building function by the graph convolutional neural network. When training the model, several node labels in the graph were masked, and then these labels were predicted by the trained model so that this work could take full advantage of the node label and the feature information of all nodes in both the training and prediction stages. Quasi-experiments proved that the proposed method for building function classification using multi-source data enables the model to capture more meaningful information with limited labels, and it achieves better function classification results.

Keywords: graph neural network; semi-supervised learning; building function classification; POI

1. Introduction

With the development of cities, the demand for urban functional classification in various industries and applications is increasing. Buildings are a fundamental component of a city, and they shape the urban structure and morphology [1,2]. Sun Wenhua et al. confirmed that there is a major relationship between urban space and building functions [3]. Consequently, building function recognition and analysis are of great significance to urban structure optimization and rational allocation of development planning [4,5]. Throughout this paper, the term "building function" refers to the use function of buildings, such as residence, shopping, etc. Accordingly, buildings can be divided into school buildings, residential buildings, commercial buildings and communal facility buildings. Building function recognition contributes to mapping navigation because buildings convey significant cognitive information and structural knowledge [3,6]. Therefore, obtaining this knowledge is also an important prerequisite for digital mapping. We can retain and enhance this information as much as possible through the data enrichment technology on the map.

A growing body of researchers gives insight into the urban functional areas [7–9]. Most studies in the field of building function or urban function classification method include unsupervised methods [10], semi-supervised methods [11] and supervised methods [12], which is

generally based on multi-source data in different regions [13]. Sun Wenhua et al. introduced time utilization activity analysis with the existing domestic building function standards to address this issue [3]. Remote sensing images and the existing urban land use/cover data with a Bayesian network were used to classify buildings [14]. These classification methods are based on deep learning, which extracts feature information (individual and overall spatial feature information) as much as possible from buildings and building groups, as well as other multi-source data for auxiliary analysis. In the era of big data, deep learning plays an important role in scientific research within different disciplines [15].

Some recent studies showed the advantages of mining information from different data sources and deep learning methods. However, how to the context information from spatial vector data and how to train an intelligent model with less labeled samples are urgent problems for urban building function recognition. This research introduces a new semi-supervised learning model, called the Unified Message Passing model (UniMP), which can be trained by less labeled samples. Moreover, a Graph Transformer jointly labels embedding to propagate both the feature and label information in the UniMP, and it can extract more context information of node label and node features in the graph. In addition, POIs and building footprints are combined to extract the building geometric and attribute information. One-hot encoding was used to describe the attributes of the building by POIs, and eight geometric features of buildings were defined, including building height, circumscribed circle radius, building orientation, minimum enclosing rectangle area, compactness, complexity, density and shape.

2. Literature Review

With the development of artificial intelligence, deep learning has been used in many geospatial analysis tasks [15–19], for it can learn the non-line relations from the training dataset. Under the impetus of multi-source of geospatial big data in the past few years [20,21], a number of deep learning methods and machine learning methods have been proposed to address the building type classifications problem [22]. In order to recognize the building type (detached building, semi-detached building, terraced building, villa, Wilhelminian-style building, etc.) from very coarse 3D city model data, support vector machines (SVMs) were introduced [23,24]. Random forest (RF) was also introduced to classify the building's roof into a flat roof, gabled roof, hipped roof, mixed form roof, pitched/shed roof and pyramid roof [25]. The above machine learning methods perform well in building roof type classification; however, none of them are suitable for building function classification.

Over a long period in the past, the functional description was characterized by visual features [26] (e.g., spectral, textural and geometrical features). With the extensive application of convolutional neural networks (CNN) in computer vision, a set of neural networks were used to classify the building function from street images [27–30]. A fusion model was proposed by Hoffmann et al. [31] to analyze the building function classification from aerial and street view images. Several scholars demonstrated that CNNs could be used to multi-label building function classification through Google Street View. HierarchyNet, a hierarchical network, was developed to classify global urban buildings into main and subcategories [32]. Moreover, in satellite imagery data, recently, the mining of information hidden in emerging geospatial big data (e.g., POIs, street view images data) was combined to classify scenes function by some CNNs [33,34]. Deep learning methods can achieve building function classification, while the accuracy needs to be improved because the contextual information of the building is not considered. Meanwhile, the buildings along roads are easy to obtain street view images, but not all buildings. Therefore, a new method for building function classification is needed.

"Everything is related to everything else, but near things are more related to each other" is not only Tobler's First Law but also a significant law of geospatial analysis. With the development of big data, technology promotes urban development and scientific progress and, at the same time, brings opportunities and challenges to the research field of artificial intelligence [15]. Graph Convolution Neural Networks (GCN) [35] provide new insight into

land use classification [36] and object detection [37,38] by mining more context information. GCN has shown excellent performance in many fields, such as computer vision [39], speech recognition [40] and natural language processing [41]. Moreover, it performs well on vector data [42] because it takes more neighbor information into account [2,43]. A building function is largely related to its neighbors; for instance, residential buildings are always built together and away from factorial buildings. Therefore, the GCN is suitable for identifying individual building functions. However, large amounts of labeled data are required to train the model by general GCNs, which is a great challenge for researchers to label a large number of buildings manually. Furthermore, at the stage of model training for node classification by general GCNs, some neighbors and topological relation information are not taken full advantage of [44–46].

Additionally, it is difficult to classify the building function in a city merely by applying a single data source [47,48]. Under the background of the surging development of information technology, a large amount of text data, including spatial location information of buildings, can be obtained, which provides a data source for the identification and analysis of urban building functions [49]. The effective combination of POIs, remote sensing images, street view data and GIS technology can analyze city information in detail, which not only realizes the classification of building functions but also improves the accuracy and reliability of classification [27,34,50]. Recent research has shown the effectiveness of urban function recognition using multiple data sources [13,51]. However, how to use the multiple geographic data and less labeled samples to train the deep learning model, which can learn more context information, is still a gap in recognizing the building classification.

3. Methodology

In this study, a novel framework was introduced for building function recognition (Figure 1). First, to recognize the function of buildings, the POIs and building footprints were matched by coordination so that the POIs corresponding to a building can be used to extract the building attributes. In this research, we introduced one-hot embedding to express the feature information of building POIs. Second, eight geometric features of building footprints were calculated, including building height, circumscribed circle radius, building orientation, minimum enclosing rectangle area, compactness, complexity, density and shape. Third, to represent spatial relationships between buildings, we conducted a graph through Delaunay Triangulation (DT) with the center point of the buildings' footprints. Therefore, every building was treated as a node of the built graph, and geometric features and attributes extracted from POIs were combined and used as the features of graph nodes. Fourth, we proposed building a multi-layer GCN architecture with two convolutional layers and one softmax layer to tackle the building function recognition problem. Then, the building function could be recognized by the trained model (Figure 1).

3.1. Building Impact Factor Calculation

The difference between buildings cannot be shown only by coordinates and the height of the buildings [52]. Several attempts were made to describe the building more comprehensively. Xu et al. introduced 8 geometric feature factors of buildings to extract the features from building footprint, including building height, circumscribed circle radius, building orientation, minimum enclosing rectangle area, compactness, complexity, density and shape [1,2]:

- Hight, radius and area

Building height was applied to distinguish the functions of buildings as a 3D feature and played a certain role in our research object. For example, office buildings are generally higher than gymnasiums. This paper also introduces the area of minimum bounder rectangle (MBR) as one of the features of a building, which is often used in GIS to give the approximate location of a geographical element. Radius refers to the radius of the minimum circumscribed circle (MCC) of the building footprint;
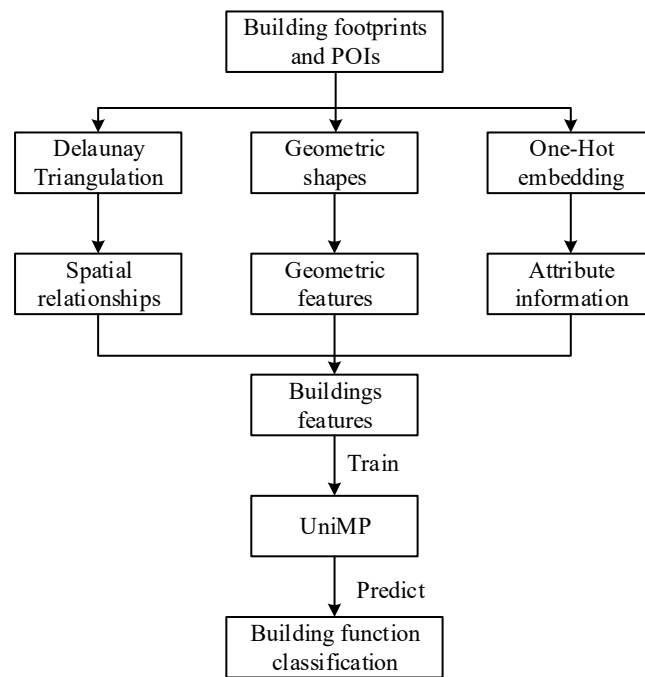
Figure 1. The workflows of building function recognition using semi-supervised classification.

- Complexity

Complexity means the complexity of a building footprint structure, which is calculated by the proportion between skeleton line lengths and the building footprint perimeter. This formula is as follows:

$$Complexity = \sum_{i=0}^{m} Skeletonline_i / Perimeter \tag{1}$$

m refers to the number of skeleton lines in the building footprint;

- Compactness

Compactness means the compact and plump degree of building footprint and is always used for the comparison and analysis of urban morphology [53]. In the case of different shapes with the same area, the compactness of their spatial distribution is different. An equation was formulated based on the moment of inertia [54] and as follows:

$$Compactness = \frac{A^2}{2\pi I_g} \tag{2}$$

In Equation (2), I_g refers to inertia moment, and A represents the area of the footprint. In this work, we suppose that all building footprint consists of several infinitesimal area units da; it can define the inertia moment as follows:

$$I_g = \int z_g^2 da \tag{3}$$

In Equation (3), z_g denotes the distance of centroid of building footprint from da. The value of I_g ranges from 0 to 1. If I_g is equal to 1, it has a circular footprint. On the contrary, when I_g approaches 0, the building footprint is close to a line and is the least compact;

- Orientation

Orientation is defined as the direction of the MBR in footprints and shows the direction consistency of a building with other buildings. Accordingly, orientation is calculated based on the vertexes coordinates of MBR, and the equation was formulated as follows:

$$Ori = \begin{cases} \arctan\frac{y_3 - y_2}{x_3 - x_2} & x_1 \neq x_2, l_2 \geq l_1 \\ \arctan\frac{y_1 - y_2}{x_1 - x_2} & x_1 \neq x_2, l_2 < l_1 \\ \frac{\pi}{2} & x_1 = x_2 \end{cases} \tag{4}$$

where l_1 and l_2 refer to the edge's length of MBR. Correspondingly, (x_1, y_1), (x_2, y_2) and (x_3, y_3) in Equation (4) represent vertex coordinates;

- Density

The term "Density" in this paper refers to the building density within a certain range. Before calculating this index, A buffer with a maximum radius R of buildings was conducted according to the footprint centroid. The density formula derived from the ratio of the sum area of building footprints in the buffer to the buffer area is as follows:

$$dens = \frac{\sum\limits_{i=0}^{n} A_i}{\pi \times R^2} \tag{5}$$

where A_i describes the i-th building's area in the buffer, and n means the building number within the buffer area. The value of density ranges from 0 to 1;

- Shape

Building shape was introduced to represent the span ratio of the building footprint. There are three typical types of building shapes: expansive, circular and compact. The calculation formula is as follows:

$$Shape = \frac{L}{2\sqrt{\pi \times A}} \tag{6}$$

where A represents the area of the building footprint, and L denotes the longer edge of MBR.

3.2. POI Data Processing

Original POI has only the second-level classification of 14 categories without major classification, including scientific research and education, residential area, committee, catering, etc. This work drew on the POI classification of AutoNavi Map and classified it into the following four categories: educational buildings, residential buildings (e.g., residential building, committee, accommodation), commercial buildings (e.g., lottery shop catering, restaurants, shopping malls, shopping finance buildings) and communal facilities (e.g., other facilities, leisure and entertainment, tourism, government).

In this research, the method of one-hot embedding was introduced to describe the POIs by binary vectors. One-hot encoding is also called one-bit effective coding. It has 4 possible classification values for any POI, and it becomes 4 binary features after one-hot embedding. These category values are mutually exclusive, and for each building, only one is activated at a time. For discrete category features, the ordinary coding method cannot be used in machine learning, and the one-hot embedding makes the calculation between features more reasonable. There are serval benefits of using one-hot encoding:

(1) The problem of the classifier being difficult when handling discrete values, such as category data, is solved;
(2) It can make some progress in expanding features, and the discrete data are converted into sparse data.

In this research, all buildings were considered research objects. If a building contained a POI of the corresponding type, the location was set to 1, and vice versa is 0. At the same

time, results were predicted through the subsequent unsupervised model according to the characteristic matrix of the building. In order to match the buildings with POIs, a 10 m buffer for every building was built. Then, a one-hot encoding matrix for subsequent experiments was carried out to the training quantity of the sample data set.

3.3. Building Function Classification

Semi-supervised classification model UniMP was used for building function classification. UniMP can aggregate the features and label information, which is helpful for graph node classification. UniMP is also a multi-layer Graph Transformer jointly using label embedding to transform nodes labels into the same vector space as nodes features, which can benefit in extracting the context information of buildings and improving the recognition ability of the model. Moreover, multi-head attentions were used as the transition matrix for propagating feature vectors so that each node could aggregate more information from its neighbors. The framework of the proposed method mainly included two parts (Figure 2): (1) representation of building graph network structure; (2) reasonable application of semi-supervised message passing network model to realize the recognition and classification of a single function of buildings in building groups. This model used multi-data and a semi-supervised model to mask the label of the building in order to avoid the leakage of node labels in the iterative process of the experimental model, resulting in inaccurate or even wrong prediction results. It can not only predict node labels according to node characteristics but also further integrate other node label information to predict the current node label.
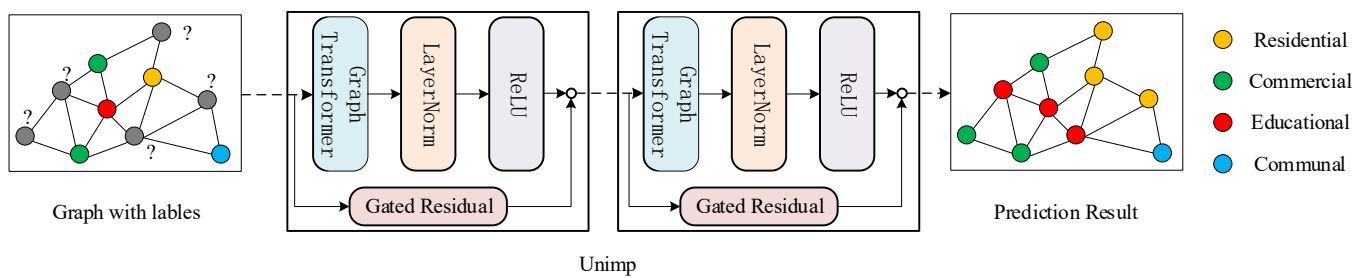

Figure 2. The framework of building function classification by UniMP.

When recognizing the building function by the UniMP, the graph structure is crucial to represent the building. Additionally, it is also the key to describing the spatial distribution of buildings. By considering the properties of the edges, an undirected connected graph $G = (V, E)$ was used in this paper, where $V = (v_1, v_2, \ldots, v_n)$ and $E \in n * n$ is a finite set of $|V| = n$ nodes and edges.

The UniMP model applies the Graph Transformer and label embedding to learn the node features and node labels from the graph. In the Graph Transformer modular, graph multi-head attention (Figure 3) was used to extract more features from multiple dimensions. For given nodes features $H^{(l)} = \left\{ h_1^{(l)}, h_2^{(l)}, h_3^{(l)} \ldots, h_n^{(l)} \right\}$, multi-head attention for each edge from j to i can be calculated as follows:

$$
\begin{aligned}
q_{c,i}^{(l)} &= W_{c,q}^{(l)} h_i^{(l)} + b_{c,q}^{(l)} \\
k_{c,j}^{(l)} &= W_{c,k}^{(l)} h_j^{(l)} + b_{c,k}^{(l)} \\
e_{c,ij} &= W_{c,e} e_{ij} + b_{c,e} \\
\alpha_{c,ij}^{(l)} &= \frac{\langle q_{c,i}^{(l)}, k_{c,j}^{(l)} + e_{c,ij} \rangle}{\sum_{u \in \mathcal{N}(i)} \langle q_{c,i}^{(l)}, k_{c,u}^{(l)} + e_{c,iu} \rangle}
\end{aligned}
\tag{7}
$$

where the $\langle q, k \rangle = \exp(\frac{q^T k}{\sqrt{d}})$ represents the exponential scale dot-product function and d is the hidden size of each head. $h_i^{(l)}$ is the source feature and $h_j^{(l)}$ is the distant

feature; $q_{c,i}^{(l)} \in \mathbb{R}^d$ is the query vector and $k_{c,j}^{(l)} \in \mathbb{R}^d$ is the key vector for the c-th head attention; $W_{c,q}^{(l)}$, $W_{c,k}^{(l)}$, $b_{c,q}^{(l)}$ and $b_{c,k}^{(l)}$ are the trainable parameters; $e_{c,ij}$ is the edge features of the graph, which are encoded and added into the key vector as additional information for each layer.
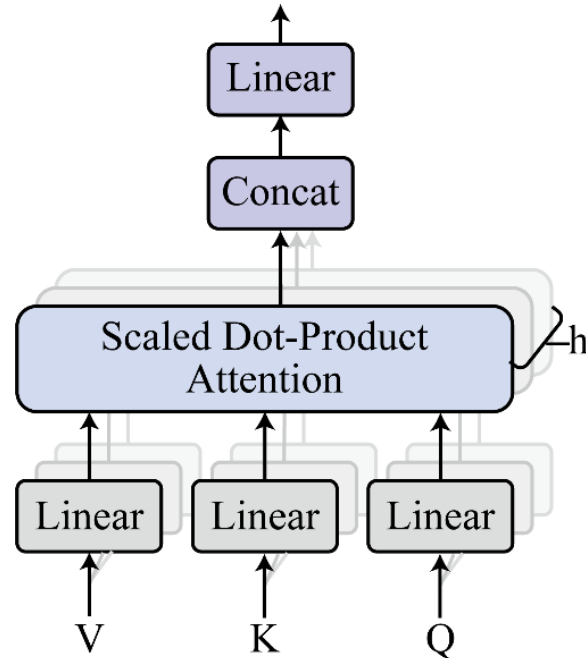

Figure 3. Graph multi-head attention.

For each head attention, the source feature $h_i^{(l)}$ and distant feature $h_j^{(l)}$ are transformed into query vector $q_{c,i}^{(l)}$ and key vector $k_{c,j}^{(l)}$, respectively. In this study, we transformed and propagated the features of each layer of nodes to obtain the node feature information data of the next layer. After the graph multi-head attention, the information aggregation from l-th layer to $(l + 1)$-th layer is as follows:

$$
\begin{aligned}
v_{c,j}^{(l)} &= W_{c,v}^{(l)} h_j^{(l)} + b_{c,v}^{(l)} \\
\hat{h}_i^{(l)} &= \overset{C}{\underset{c=1}{\|}} \left[\sum_{j \in \mathcal{N}(i)} \alpha_{c,ij}^{(l)} \left(v_{c,j}^{(l)} + e_{c,ij} \right) \right] \\
r_i^{(l)} &= W_r^{(l)} h_i^{(l)} + b_r^{(l)} \\
\beta_i^{(l)} &= sigmoid(W_g^{(l)} [\hat{h}_i^{(l)}; r_i^{(l)}; \hat{h}_i^{(l)} - r_i^{(l)}]) \\
h_i^{(l+1)} &= ReLU(LayerNorm((1 - \beta_i^{(l)})\hat{h}_i^{(l)} + \beta_i^{(l)} r_i^{(l)}))
\end{aligned}
\tag{8}
$$

where the $\|$ represents the concatenation operation for head attention C. For the output layer, the averaging for multi-head is used as the output as follows:

$$
\begin{aligned}
\hat{h}_i^{(l)} &= \frac{1}{C} \sum_{c=1}^{C} \left(\left[\sum_{j \in \mathcal{N}(i)} \alpha_{c,ij}^{(l)} \left(v_{c,j}^{(l)} + e_{c,ij}^{(l)} \right) \right] \right) \\
h_i^{(l+1)} &= (1 - \beta_i^{(l)}) \hat{h}_i^{(l)} + \beta_i^{(l)} r_i^{(l)}
\end{aligned}
\tag{9}
$$

The label embedding was used in UniMP to propagate the obtained label information. First, the labeled node was represented by a one-hot matrix, and the node without a label was represented by a zero vector. Second, the adjacency matrix $D^{-1}A$ after regularization was used to represent $\hat{Y}^l$ propagation and to obtain the label information representation of

the $(l + 1)$-th layer. Third, the fused node feature information and label information was further propagated, as shown below:

$$H^0 = X + \hat{Y}W_c$$
$$H^{l+1} = \sigma\left(((1 - \beta)A^* + \beta I)H^l W^l\right) \tag{10}$$

This paper also took full advantage of the node's label information, further enhancing the characteristic information of the node so that the model can obtain more necessary information. The UniMP framework model predicts the node by the neighbor labeled node, node feature and topological information as follows:

$$\text{argmax } \log p_\theta\left(\overline{Y}\Big|X, \widetilde{Y}, A\right) = \sum_{i=1}^{\overline{V}} \log p_\theta\left(\left(\overline{y_i}\Big|X, \widetilde{Y}, A\right)\right) \tag{11}$$

4. Quasi-Experiment and Analysis

4.1. Study Area

Nanjing is situated in the middle and lower reaches of the Yangtze River in eastern China. As the capital of Jiangsu province, it is one of the first batches of national historical and cultural cities and an important birthplace of Chinese civilization. The city covers an area of over six thousand square kilometers, and it governs 11 districts. The specific geographical location is shown in Figure 4. This study took Nanjing as the research area, and the core functional area consists of commercial buildings, residential buildings, educational buildings and communal buildings. In this research, building footprints were obtained from the software of Shuijingzhu, and the POIs were downloaded from Gaode. Both of them are geospatial data providers in China. In order to ensure data reliability and validity for this study, the building functions were labeled by at least three participants according to Google Maps or Baidu Maps. If the results marked by the three participants were inconsistent, the area was remarked or discarded, which are manually marked and verified to ensure the authenticity, reliability and accuracy of the data.
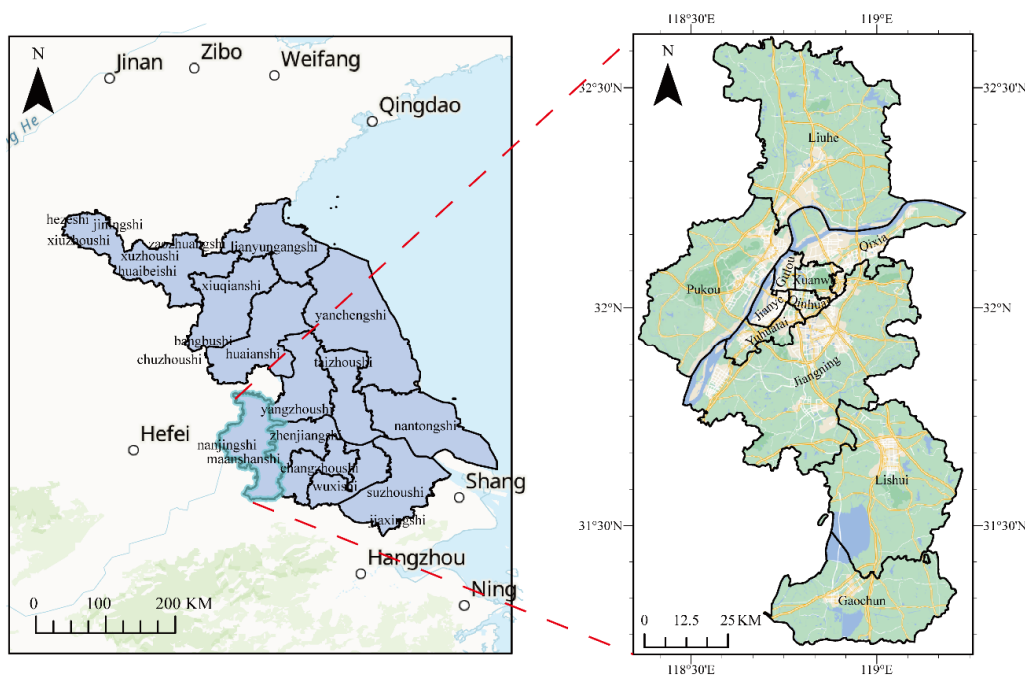

Figure 4. Study area, Nanjing.

4.2. Experimental Process and Network Setting

In order to match the POIs with buildings, we performed a buffer analysis with a radius of 10 m of each POI, thereby finding the building closest to the POI point data within the set radius and finally generating a one-to-one matching attribute table between POI and build. Therefore, the attributes of POIs can be used as the features of buildings to recognize their function. Four core fields of POI were reserved: name, address, coordinate and category.

In this experiment, computer configuration and experimental environments were Intel i7-10700k eight-core CPU and Ubuntu 15.5.0 Linux operating system, 64 GB memory and two NVIDIA RTX 2080Ti GPU graphics cards. The methods involved in this paper were implemented in the Pytorch platform implemented by Python. Some geometric features of building footprints were calculated by ArcGIS, such as the area, perimeter, etc. Building footprints were obtained from the software of Shuijingzhu. We set the learning rate to 0.001 and used the Adam optimization algorithm to realize the iterative update of the weights in the model. The maximum number of iterations of the model was 2000, and the network model performance was tested once every 10 iterations of the model.

4.3. Evaluation Indexes

The provided metrics of F_1 score and the accuracy of each building function type are used to assess the quantitative performance. F_1 score is a representation of the harmonic mean of precision and recall, and it can be calculated as follows:

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{12}$$

where

$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN} \tag{13}$$

Here, *TP* is the number of true positives for each building function type, *FP* and *FN* represent false positives and false negatives, respectively. These metrics were computed using the sample-based confusion matrices for the dataset [55].

4.4. Results and Analysis

Through the description of the study area, we can understand that the characteristics of the building group are closely related to the local climate, green landscape and human factors. A Building group is usually regarded as the basic unit of a city block. The good results of the experiment cannot be achieved by only using building data. We tried to improve the accuracy of recognizing building functions using building footprints and POI data because they can reflect the real situation of buildings and express urban building units objectively and specifically. Information in this model, including class name, geographical location and POI category (primary and secondary classification), provides a powerful basis for the analysis of urban basic unit buildings in this paper.

From the results (Figure 5), we can see before 800 iterations of the model, the loss value of the model decreases at a fast and stable rate. On the contrary, the training accuracy increases at a faster rate. The test accuracy and verification accuracy basically increased at a relatively rapid rate before 200 iterations of the model. At this time, the loss value of the model is still declining, and the test accuracy cannot be determined before the loss becomes stable. Finally, when the loss reached about 1600, it basically stabilized at a small fluctuation around 0.1. At this time, the training accuracy has reached a stable peak, and finally, the model verification and testing accuracy has reached the most stable range. The stability and convergence rate of the model was greatly improved, and the most concerned test accuracy was achieved at over 81%.

Figure 6 displays the classification results of UniMP for the building function classification on the validated dataset. From the results, we can see that most buildings can be classified into the right category, especially residential buildings and educational buildings. The accuracies for these two types were over 80%, and the accuracies for educational

buildings and residential buildings were 81% and 89%, respectively. Such results are due to these features for these two categories of buildings being obvious compared to other categories of buildings. For example, most of the residential buildings and educational buildings are regular. Moreover, the POIs for educational buildings are obvious, such as "teaching building" and "faculty", etc. In the research area, some educational buildings were misclassified into residential buildings. The reason for the misclassification may be because of the widespread existence of residential faculties and family residential areas within Chinese university areas. However, the performance of the designed model was a little poor in identifying communal facilities. First of all, there are relatively few training samples in this type of buildings compared with other types. Secondly, communal facilities are usually included or adjacent to commercial and residential buildings within the research area. From the confusion matrix, it can be found that communal facilities are partially recognized as residential buildings and commercial buildings.
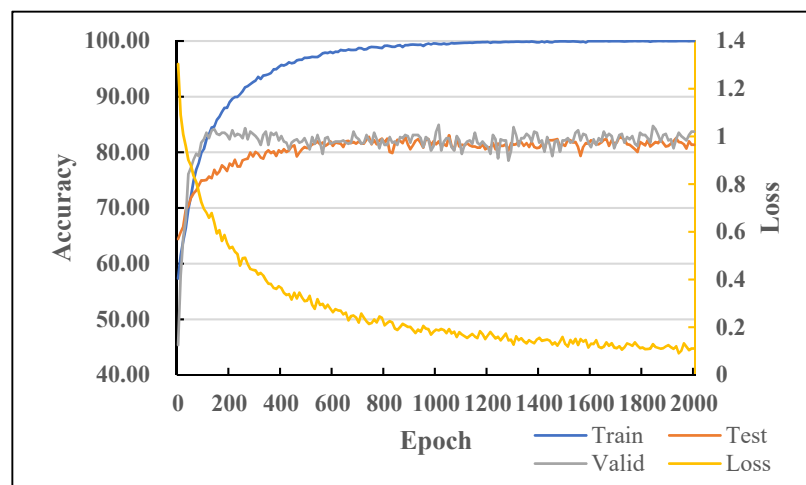

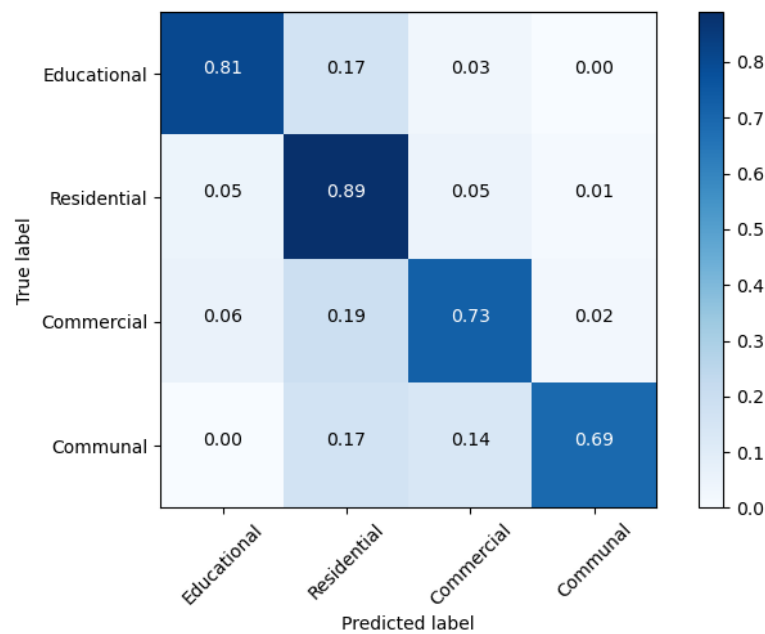**Figure 5.** The results of combined POI and building.

Figure 6. The confusion matrix for the classification results of UniMP on the classification of the building function.

Based on the building feature and POI attributes, the semi-supervised message passing model can predict the building function according to the characteristics of the node.

Moreover, the model can learn the information of adjacent nodes to predict the label of the current node. Because multi-source data were used, it was easier for the model to capture more meaningful feature information. The comparison between predicted results and the actual situation is shown in Figure 7. We can see that the proposed model can recognize the building function of some complex situations, such as school and residential buildings.
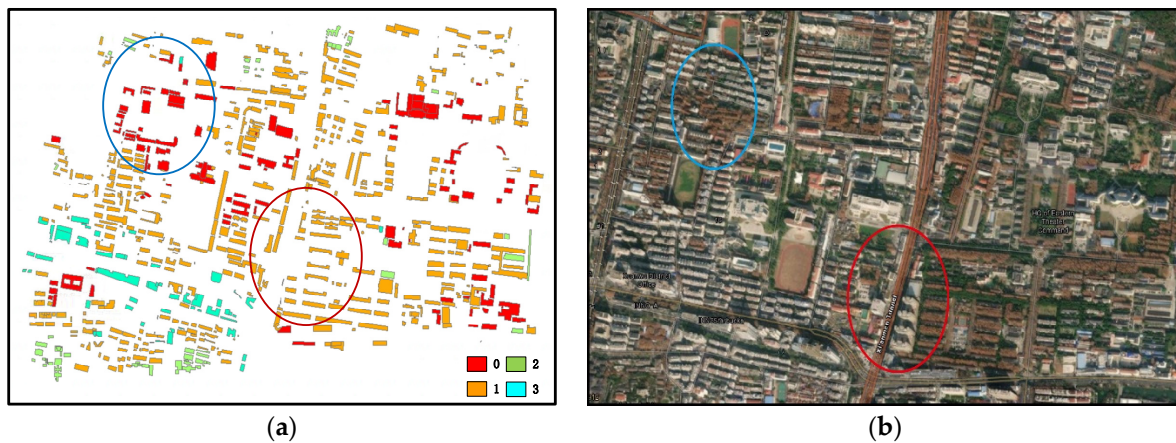

(**a**)   (**b**)

Figure 7. The comparison between (**a**) predicted results and (**b**) actual situation, where 0 represents school buildings, 1 represents residential buildings, 2 represents commercial buildings and 3 represents communal facilities.

4.5. Ablation Study

In order to analyze the advantages of combining the POIs and building footprints data, this research conducted several ablation experiments. From Figure 8, we can see that after adding the POI characteristic information, the model loss convergence speed was significantly improved, and the average convergence rate of the loss value in the model iteration was also faster. By using both POIs attributes and building footprint features, the loss value was stable at around 0.1. This is an important improvement compared with the results of single POI or single build footprints data. The results proved that combining the POIs and building footprints is useful for building function recognition by semi-supervised methods.
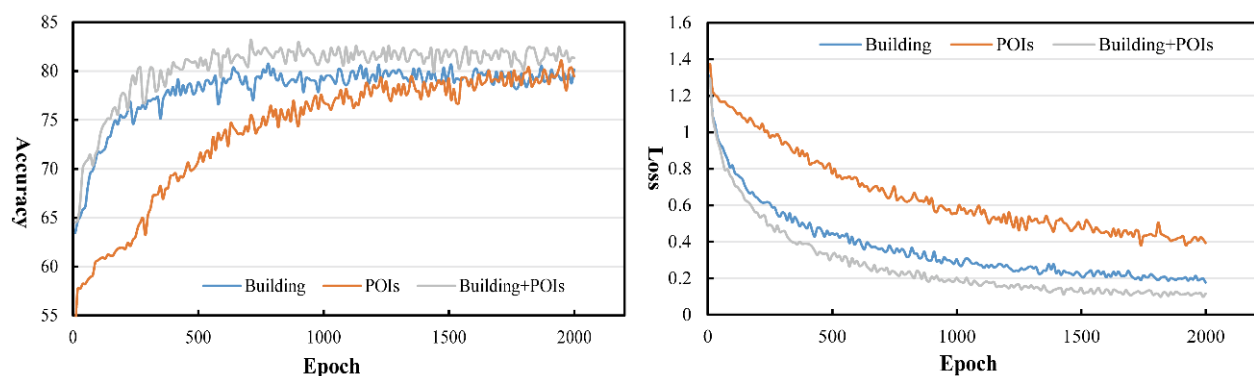

Figure 8. The accuracy and loss with respect to epoch of different datasets.

4.6. Comparison to Other Methods

This paper selected several mainstream classification methods to make a contrast to the method proposed in this work, including Support Machine (SVM) and Random Forest (RF). During the experiment, we used the same datasets in the above three models, and indexes of accuracy, recall and F1 were used to evaluate the performance of the three models statically. The results presented in Table 1 demonstrated that the proposed model performs superior to other models because it effectively benefits from context features learning.

The UniMP algorithm extracts the features and label information of surrounding nodes both in the training and prediction stages, which is more convenient to extract building information according to surrounding buildings. On the other hand, machine learning models such as SVM and RF cannot learn the deep features and surrounding buildings' information; therefore, it is not suitable for building function recognition. From the results, we can see that UniMP has the best performance for all the evaluator indicators. In detail, the F1 score of UniMP reached 80.56%, which improved 23.25% and 12.51% for SVM and RF, respectively. The accuracy was improved maximum by 11.22% compared with RF, and the recall was improved maximum by 9.58% compared with SVM. Therefore, the proposed novel framework for building function recognition can learn more context information. Moreover, the deep learning model can be trained well with less labeled samples.

Table 1. Results of UniMP, SVM and RF.

Model	Accuracy	Recall	F1
UniMP	0.8106	0.8007	0.8056
SVM	0.7767	0.7049	0.5731
RF	0.6984	0.7304	0.6805

4.7. Discussion

From the perspective of contextual, geometric and attribute information of buildings, the aggregation of contextual information, geometric information and attribute information is meaningful. In this paper, we proposed a novel semi-supervised framework based on UniMP for building function recognition, and POIs and building footprints were combined to mine the features of building functions. In order to extract the context information of buildings and improve the recognition ability, a multi-layer graph transformer was used to transform node labels into the same vector space as node features in UniMP.

With the development of information technology, there are a large amount of spatial location information of buildings can be obtained; these data sources are helpful for the identification and analysis of urban building functions. However, how to use multiple data sources to train a deep learning model, especially for heterogeneous data, is still a gap in recognizing the building classification. This paper proposes encoding the POIs with one-hot; eight geometric features of buildings were designed, including building height, circumscribed circle radius, building orientation, minimum enclosing rectangle area, compactness, complexity, density and shape. The experiment proves that combing the POIs and building footprints can improve the accuracy of recognition and speed of model convergence.

Contextual information is important for building function recognition based on the first law of geography. Moreover, how to train an intelligent model with fewer samples is also a difficult problem for building function recognition. In order to solve these issues, this paper introduced the semi-supervised classification, UniMP. These buildings were organized into a graph by the Delaunay triangulation, and the geometric and attribute information of buildings were treated as the node features of graph. Experiential results show that the UniMP can extract contextual information effectively. Moreover, semi-supervised learning models can obtain a good performance compared with other machine learning methods, such as SVM and RF.

5. Conclusions

Buildings are the core elements of people living in the city. In order to recognize the function types of city buildings, the semi-supervised classification model, UniMP, was introduced to speed up cognition and understanding of buildings. In order to mine more information for building function recognition, the buildings' footprints and POI information were combined. For the POIs, the one-hot encoding was used to extract the attributes of buildings; while for buildings footprints, eight geometric features were defined to describe the different functional buildings, including building height, circumscribed circle radius,

building orientation, minimum enclosing rectangle area, compactness, complexity, density and shape. Due to the use of multi-source data, the model captured more meaningful feature information of relevant nodes, then achieved better results in building function recognition. Moreover, this research treated every building as a node of the graph during the training model. Thus, we could not only predict the label according to the node characteristics but also integrate the label information of other neighbor nodes. Compared with the results of single POI (79.14%) or a single building's footprint data (79.97%), the accuracy for using both POIs attributes and buildings' footprint features is stable at around >81%. The results proved that combining the POIs and buildings' footprints is useful for building function recognition by semi-supervised methods.

In order to extract more context information of node labels and node features in the graph, the semi-supervised classification model was introduced in this work, which can achieve good performance with limited labeled samples. Although this study effectively classified urban building functions, it had certain limitations because all the buildings are Chinese architectural styles. In the future, we will try to apply some other types of data, including social media data such as microblog check-in data, taxi track information, street view information, etc. It is expected that the model can achieve better results than the current experimental results in the subsequent experimental process.

References

1. Guo, M.; Liu, H.; Xu, Y.; Huang, Y. Building extraction based on U-Net with an attention block and multiple losses. *Remote Sens.* **2020**, *12*, 1400. [CrossRef]
2. Xu, Y.; He, Z.; Xie, X.; Xie, Z.; Luo, J.; Xie, H. Building function classification in Nanjing, China, using deep learning. *Trans. GIS* **2022**, *26*, 2145–2165. [CrossRef]
3. Sun, W.; Chen, J. Time Utilization Activity Based Classification of Architectural Functions. *Hous. Sci.* **2016**. [CrossRef]
4. Wurm, M.; Schmitt, A.; Taubenböck, H. Building types' classification using shape-based features and linear discriminant functions. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *9*, 1901–1912. [CrossRef]
5. Steiniger, S.; Lange, T.; Burghardt, D.; Weibel, R. An approach for the classification of urban building structures based on discriminant analysis techniques. *Trans. GIS* **2008**, *12*, 31–59. [CrossRef]
6. He, Z.; Deng, M.; Cai, J.; Xie, Z.; Guan, Q.; Yang, C. Mining spatiotemporal association patterns from complex geographic phenomena. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 1162–1187. [CrossRef]
7. Hu, S.; Gao, S.; Wu, L.; Xu, Y.; Zhang, Z.; Cui, H.; Gong, X. Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. *Comput. Environ. Urban Syst.* **2021**, *87*, 101619. [CrossRef]
8. Yao, Y.; Li, X.; Liu, X.; Liu, P.; Liang, Z.; Zhang, J.; Mai, K. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 825–848. [CrossRef]
9. Gao, S.; Janowicz, K.; Couclelis, H. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Trans. GIS* **2017**, *21*, 446–467. [CrossRef]
10. Frias-Martinez, V.; Frias-Martinez, E. Spectral clustering for sensing urban land use using Twitter activity. *Eng. Appl. Artif. Intell.* **2014**, *35*, 237–245. [CrossRef]
11. Pei, T.; Sobolevsky, S.; Ratti, C.; Shaw, S.; Li, T.; Zhou, C. A new insight into land use classification based on aggregated mobile phone data. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1988–2007. [CrossRef]
12. Zhai, W.; Bai, X.; Shi, Y.; Han, Y.; Peng, Z.; Gu, C. Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Comput. Environ. Urban Syst.* **2019**, *74*, 1–12. [CrossRef]
13. Chen, S.; Zhang, H.; Yang, H. Urban Functional Zone Recognition Integrating Multisource Geographic Data. *Remote Sens.* **2021**, *13*, 4732. [CrossRef]

14. Li, M.; Stein, A.; Bijker, W.; Zhan, Q. Urban land use extraction from Very High Resolution remote sensing imagery using a Bayesian network. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 192–205. [CrossRef]

15. Xu, Y.; Chen, Z.; Xie, Z.; Wu, L. Quality assessment of building footprint data using a deep autoencoder network. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1929–1951. [CrossRef]

16. Xu, Y.; Xie, Z.; Feng, Y.; Chen, Z. Road extraction from high-resolution remote sensing imagery using deep learning. *Remote Sens.* **2018**, *10*, 1461. [CrossRef]

17. Hu, B.; Wan, B.; Xu, Y.; Tao, L.; Wu, X.; Qiu, Q.; Wu, Y.; Deng, H. Mapping hydrothermally altered minerals with AST_07XT, AST_05 and Hyperion datasets using a voting-based extreme learning machine algorithm. *Ore Geol. Rev.* **2019**, *114*, 103116. [CrossRef]

18. Hu, B.; Xu, Y.; Wan, B.; Wu, X.; Yi, G. Hydrothermally altered mineral mapping using synthetic application of Sentinel-2A MSI, ASTER and Hyperion data in the Duolong area, Tibetan Plateau, China. *Ore Geol. Rev.* **2018**, *101*, 384–397. [CrossRef]

19. Hu, B.; Xu, Y.; Huang, X.; Cheng, Q.; Ding, Q.; Bai, L.; Li, Y. Improving Urban Land Cover Classification with Combined Use of Sentinel-2 and Sentinel-1 Imagery. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 533. [CrossRef]

20. Lee, J.-G.; Kang, M. Geospatial big data: Challenges and opportunities. *Big Data Res.* **2015**, *2*, 74–81. [CrossRef]

21. Cai, J.; Huang, B.; Song, Y. Using multi-source geospatial big data to identify the structure of polycentric cities. *Remote Sens. Environ.* **2017**, *202*, 210–221. [CrossRef]

22. Shirowzhan, S.; Trinder, J. Building classification from lidar data for spatio-temporal assessment of 3D urban developments. *Procedia Eng.* **2017**, *180*, 1453–1461. [CrossRef]

23. Römer, C.; Plümer, L. Identifying architectural style in 3d city models with support vector machines. *PFG Photogramm. Fernerkund. Geoinf.* **2010**, *2010*, 371–384. [CrossRef] [PubMed]

24. Henn, A.; Römer, C.; Gröger, G.; Plümer, L. Automatic classification of building types in 3D city models. *GeoInformatica* **2012**, *16*, 281–306. [CrossRef]

25. Biljecki, F.; Dehbi, Y. Raise the roof: Towards generating LOD2 models without aerial surveys using machine learning. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *4*, 27–34. [CrossRef]

26. Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Land use classification in remote sensing images by convolutional neural networks. *arXiv* **2015**, arXiv:1508.00092.

27. Xu, Y.; Jin, S.; Chen, Z.; Xie, X.; Hu, S.; Xie, Z. Application of a graph convolutional network with visual and semantic features to classify urban scenes. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 2009–2034. [CrossRef]

28. Cao, R.; Zhu, J.; Tu, W.; Li, Q.; Cao, J.; Liu, B.; Zhang, Q.; Qiu, G. Integrating aerial and street view images for urban land use classification. *Remote Sens.* **2018**, *10*, 1553. [CrossRef]

29. Fang, F.; Yu, Y.; Li, S.; Zuo, Z.; Liu, Y.; Wan, B.; Luo, Z. Synthesizing location semantics from street view images to improve urban land-use classification. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 1802–1825. [CrossRef]

30. Zhanjun, H.; Wang, Z.; Xie, Z.; Wu, L.; Chen, Z. Multiscale analysis of the influence of street built environment on crime occurrence using street-view images. *Comput. Environ. Urban Syst.* **2022**, *97*, 101865.

31. Hoffmann, E.J.; Wang, Y.; Werner, M.; Kang, J.; Zhu, X. Model fusion for building type classification from aerial and street view images. *Remote Sens.* **2019**, *11*, 1259. [CrossRef]

32. Taoufiq, S.; Nagy, B.; Benedek, C. HierarchyNet: Hierarchical CNN-Based Urban Building Classification. *Remote Sens.* **2020**, *12*, 3794. [CrossRef]

33. Lu, W.; Tao, C.; Li, H.; Qi, J.; Li, Y. A unified deep learning framework for urban functional zone extraction based on multi-source heterogeneous data. *Remote Sens. Environ.* **2022**, *270*, 112830. [CrossRef]

34. Bao, H.; Ming, D.; Guo, Y.; Zhang, K.; Zhou, K.; Du, S. DFCNN-based semantic recognition of urban functional zones by integrating remote sensing data and POI data. *Remote Sens.* **2020**, *12*, 1088. [CrossRef]

35. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.

36. Li, M.; Stein, A.; de Beurs, K.M. A Bayesian characterization of urban land use configurations from VHR remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102175. [CrossRef]

37. Wang, L.; Wang, C.; Zhang, X.; Lan, T.; Li, J. S-AT GCN: Spatial-Attention Graph Convolution Network based Feature Enhancement for 3D Object Detection. *arXiv* **2021**, arXiv:2103.08439.

38. Xu, M.; Fu, P.; Liu, B.; Li, J. Multi-stream attention-aware graph convolution network for video salient object detection. *IEEE Trans. Image Process.* **2021**, *30*, 4183–4197. [CrossRef]

39. Cheng, K.; Zhang, Y.; He, X.; Chen, W.; Cheng, J.; Lu, H. Skeleton-based action recognition with shift graph convolutional network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 183–192.

40. Liu, H.; Chen, Z.; Yang, B. Lip Graph Assisted Audio-Visual Speech Recognition Using Bidirectional Synchronous Fusion. In Proceedings of the Interspeech 2020, Shanghai, China, 25–29 October 2020.

41. Alhogail, A.; Alsabih, A. Applying machine learning and natural language processing to detect phishing email. *Comput. Secur.* **2021**, *110*, 102414. [CrossRef]

42. Yan, X.; Ai, T.; Yang, M.; Yin, H. A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 259–273. [CrossRef]

43. Xu, Y.; Zhou, B.; Jin, S.; Xie, X.; Chen, Z.; Hu, S.; He, N. A framework for urban land use classification by integrating the spatial context of points of interest and graph convolutional neural network method. *Comput. Environ. Urban Syst.* **2022**, *95*, 101807. [CrossRef]

44. Bruna, J.; Zaremba, W.; Szlam, A.; Le, C. Spectral networks and locally connected networks on graphs. *arXiv* **2013**, arXiv:1312.6203.

45. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4–24. [CrossRef] [PubMed]

46. He, Z.; Deng, M.; Xie, Z.; Wu, L.; Chen, Z.; Pei, T. Discovering the joint influence of urban facilities on crime occurrence using spatial co-location pattern mining. *Cities* **2020**, *99*, 102612. [CrossRef]

47. Lin, A.; Sun, X.; Wu, H.; Luo, W.; Wang, D.; Zhong, D.; Wang, Z.; Zhao, L.; Zhu, J. Identifying urban building function by integrating remote sensing imagery and POI data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8864–8875. [CrossRef]

48. Fonte, C.; Minghini, M.; Anyoniou, V.; Patriarca, J. Classification of Building Function using available sources of VGI. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*, 209–215. [CrossRef]

49. Ai, T.; Yun, S.; Li, J. A spatial query based on shape similarity cognition. *Acta Geod. Et Cartogr. Sin.* **2009**, *38*, 356–362.

50. Xu, S.; Qing, L.; Han, L.; Liu, M.; Peng, Y.; Shen, L. A new remote sensing images and point-of-interest fused (RPF) model for sensing urban functional regions. *Remote Sens.* **2020**, *12*, 1032. [CrossRef]

51. Zhang, Y.; Li, Q. Functional urban land use recognition integrating multi-source geospatial data and cross-correlations. *Comput. Environ. Urban Syst.* **2019**, *78*, 101374. [CrossRef]

52. Wang, Z.; Liu, Q. The Cluster of City Buildings Based on the SOM Neural Network. *IOP Conf. Ser. Earth Environ. Sci.* **2017**, *57*, 012047. [CrossRef]

53. Gillman, R. Geometry and Gerrymandering. *Math Horiz.* **2002**, *10*, 10–12. [CrossRef]

54. Li, W.; Goodchild, M.; Church, R. An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 1227–1250. [CrossRef]

55. Xu, Y.; Wu, L.; Xie, Z.; Chen, Z. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sens.* **2018**, *10*, 144. [CrossRef]