EXPLORING URBAN FUNCTIONAL ZONES BASED ON MULTI-SOURCE SEMANTIC KNOWLEDGE AND CROSS MODAL NETWORK

Jiage Chen [1], Shu Peng[1]*, Hongwei Zhang [1], Shangwei. Lin [1] , Wenzhi Zhao[2]

1 National Geomatics Center of China, 100830 Beijing, China - (jiagechen, pengshu, hwzhang, linshangwei)@ngcc.cn
2 State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, 100875 Beijing, China-(wenzhi.zhao@bnu.edu.cn)

KEY WORDS: Multi-source semantic knowledge, deep learning, Refinement of urban functional areas

ABSTRACT:

The refined identification of urban functional zones can provide important basic data and decision-making basis for the formulation of urban spatial development planning, effective relaxation of urban spatial development planning, effective relaxation of urban functions, and optimal allocation of resource space. The multi-source spatiotemporal data represented by multi-source geographic data, social perception data, and thematic data have been widely used in various fields, providing new data sources for the refined identification of urban functional areas. However, there are significant differences in the data generation sources, collection methods, and storage organization formats of multi-source data. In this paper, we propose a method for exploring urban functional zones based on Multi-source Semantic knowledge and deep Coupling Model (MSCM). Our approach integrates information from multiple sources and incorporates the semantics of urban functional zones into a knowledge graph, enabling effective fusion and mining of multi-source data.This method can improve the credibility and precision of the results, providing a richer research perspective for refined urban functional zoning. The results of this paper have important theoretical value and practical significance for the construction of identification, labelling, and monitoring tools for engineering smart cities.

1. INTRODUCTION

Urban functional zones refer to areas within a city that serve specific purposes, such as residential, commercial, industrial, or recreational areas. The identification of functional zones are crucial for urban development planning, resource allocation and citizen welfare (Zhao et al., 2019). In the past, urban functional identification methods were mainly based on field survey statistics and remote sensing image recognition. The former was are time-consuming and costly. Although remote sensing as a feasible solution to consecutively monitor the physical urban functional has been widely studied. It had a single data source that could not meet the needs of refined urban functional area identification. Accurately identify urban objects from remote sensing is one of the fundamental challenges for urban environment management. For example, remote sensing images can only reflect low-level features such as the spectrum, texture, and shape of ground objects, lacking semantic information about urban functional areas (Ma et al., 2019).

With the continuous acquisition of new data, especially the advent of the Web 2.0 era, geographic "big data" is constantly accumulating. At present, the identification method of urban functional areas mainly uses mobile devices, passenger flow and crowd-sourced geographical data to establish the relationship between urban Semantic information and data characteristics through data mining, cluster analysis and other technologies, so as to obtain the urban spatial structure and Semantic information (Rabat and Kumar, 2015; Yao et al., 2017). For example, studies using passenger flow and mobile device data often establish probability models based on residents' travel patterns, calculate the frequency of travel patterns between arrival and departure points, achieve spatial clustering of urban residents' travel characteristics, and conjecture the urban spatial structure and functional zoning in combination with the distribution characteristics of POI data, so as to obtain urban Semantic information (Liao et al., 2017; Liu et al., 2016). In addition, some scholars have established potential LDA and DMR models based on the combination of OSM road network data, taxi track data and POI data, using spatiotemporal semantic mining methods, and obtained urban functional areas through spatial clustering methods. Finally, combined with questionnaires and POI, urban spatial structure and Semantic information are extracted (Barlacchi et al., 2017). However, obtaining data on passenger flow and mobile devices is very difficult, as it involves personal privacy and social security, and data sharing is difficult, resulting in significant limitations on related research.

With the onset of the “big data”, multi-source spatiotemporal data represented by multi-source geographic data, social perception data, and thematic data have been widely used in various fields, providing new data sources for the refined identification of urban functional areas. However, there are significant differences in the data generation sources, collection methods, and storage organization formats of multi-source data. The coupling of multi-source data poses challenges such as inconsistent data attributes and incompatible semantics. If traditional spatial coupling relationships are mechanically used to fuse multi-source data, it may amplify the negative impact of data bias.

With the development of artificial intelligence technology, deep learning methods have shown great application value in remote sensing image information extraction due to their strong autonomous feature learning ability and high level of automation, providing a new solution for the refinement

* Corresponding author

research of urban functional areas (Zhang et al., 2016; Zou et al., 2015; Maggiori et al., 2016). However, most existing deep learning methods mainly extract information by mining the depth features of remote sensing images themselves. For example, Convolutional Neural Network (CNN), as one of the most representative deep learning algorithms, has been widely used in remote sensing image feature extraction and classification (Cheng et al., 2018; Zhao and Du, 2016b), and automated hyperspectral image deep feature mining and extraction (Li et al., 2017). So far, there is almost no research on using deep learning methods to fuse multi-source spatiotemporal data for extraction. Although some scholars proposed to use the active learning model to improve the sample selection problem in OSM data, and combined with the deep network framework, the preliminary extraction of urban regional spatial structure and Semantic information was initially realized. But the focus is on extracting urban landmark information, neglecting the spatial distribution and semantic expression characteristics of urban functional areas at different scales (such as blocks, administrative districts, etc.). Different from the extraction of urban Semantic information at the surface feature scale, the research based on urban functional areas realizes the extraction of comprehensive Semantic information at a specific scale by associating geographical data and remote sensing image features, but ignores the distribution information of surface features at a fine scale. Therefore, the current work only considers the extraction and expression of urban surface objects and Semantic information at a single scale, ignoring the changes and uncertainties of urban functional area types brought about by different spatial scales, which will directly affect the extraction and analysis of urban functional areas.

In this paper, we propose a method for exploring urban functional zones based on Multi-source Semantic knowledge and deep Coupling Model (MSCM). First, multi-source spatio-temporal data (such as remote sensing images, multi-source geographic data, etc.) are acquired and pre-processed, and the semantic knowledge base of multi-source data in urban functional areas is constructed through knowledge extraction and knowledge fusion methods. On this basis, a coupled model of knowledge graph and convolutional network is constructed, which can further enhance the interpretability of the deep learning model based on remote sensing image semantic segmentation. This method can improve the credibility and precision of the results, providing a richer research perspective for refined urban functional zoning. The results of this paper have important theoretical value and practical significance for the construction of identification, labelling, and monitoring tools for engineering smart cities.

The major contributions of this paper are summarized as follows: (1) our approach integrates information from multiple sources and incorporates the semantics of urban functional zones into a knowledge graph, enabling effective fusion and mining of multi-source data. (2) In order to solve the difference between knowledge representation and deep learning representation, this paper proposes a deep coupling model for integrating semantic features and visual features.

The remainder of this paper is organized as follows. Section 2 introduces the representation learning of remote sensing knowledge graph and the deep cross-modal coupling model MSCM in detail. Section 3 summarizes the experimental results. Finally, the conclusion is detailed in Section 4.

2. METHODOLOGY

Our method consists of two main components: the multi-source semantic knowledge graph construction and a deep coupling model. Firstly, we construct a knowledge graph that integrates the semantics of urban functional zones from multiple data sources, including crowd-sourced geospatial data (eg. OpenStreetMap), social sensing data, and thematic data. The knowledge graph enables effective fusion and mining of multi-source data and can capture the complex relationships among urban functional zones. The knowledge graph semantically represents and organizes different types of data, enabling interaction and integration between them. Secondly, we embed the knowledge graph vectors into a deep coupling network. The deep coupling model combines multiple modalities, including spatial, temporal, and attribute information, to learn the distribution of different functional zones in the city. The model utilizes a cross-modal attention mechanism to adaptively fuse different modalities and capture the interactions among them. Lastly, for each urban scene, we apply spatial measurements to evaluate the attributes of urban functional in terms of their spatial distributions. Urban functional zones can be effectively classified based on the semantic contents and spatial properties. The flow of processes within the proposed framework is depicted in Figure 1.

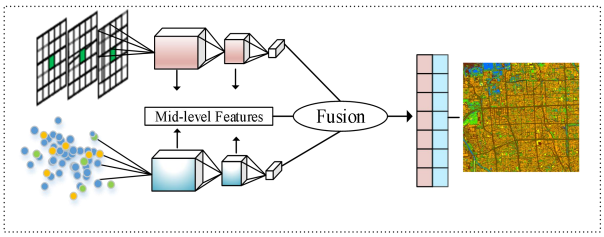


Figure 1. Flowchart of the proposed method.

2.1 Multi source semantic knowledge extraction and fusion

The refinement research of urban functional areas based on multi-source data still faces challenges due to the diversity and complexity of multi-source data. In this paper, the domain knowledge network was constructed through knowledge extraction, spatial or non-spatial association. At the same time, based on the entity linking method, a knowledge graph of multi-source data was constructed to form a semantic knowledge base of multi-source data. This knowledge base semantically represents and organizes different types of data, achieving interaction and integration between data, and providing a data foundation for subsequent model training and inference.

The multi-source data involved in urban functional areas mainly include remote sensing images, POI (point of interest) multi-source geographic data, land use data, other statistical data, text data, etc. This paper divides structured data, vector data and other data according to the storage type, and uses different knowledge extraction methods according to different data types. For structured data, the mapping relationship between concepts in the database and ontology in the knowledge map and rule-based reasoning are established to automatically extract geographical entities, attributes and their relationships from the database. For semi-structured data, domain knowledge is extracted from unstructured text through natural language processing technology. Training on vector data using existing relationships between knowledge graph entities and OSM nodes can effectively capture the similarity between semantic nodes, discover the relationships between entities and OSM nodes, and achieve knowledge

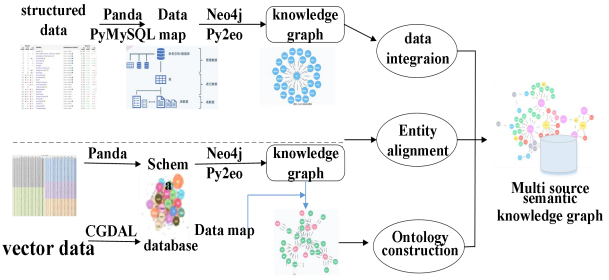


Figure 2. Multi source semantic knowledge fusion

2.2 Deep Learning Model Construction

In order to obtain better matching results between different modalities, this paper adopts a deep cross modal model, which maps both visual and semantic features to the hidden space. While achieving matching between visual features and semantic representations, it also enhances the coupling relationship between visual features and semantic representations. Firstly, the cross modal framework of deep convolutional neural networks is constructed. Secondly, complete the learning and adjustment of deep network model parameters based on the training samples. This model adopts a multi-layer convolutional neural network for automatic extraction and learning of data features. At the same time, the model also introduces a multi-source data fusion mechanism for interaction and collaboration between different data sources, improving the recognition accuracy and robustness of the model. Suppose, there are N semantic objects S_m, that have the spatial distribution of D which is within scene B_q, then the spatial distribution is calculated by measuring the shortest distances between the centroids of any two objects. Therefore, the distribution of spatial distances can be represented as follow:

$$D_{dist} = \min_{i,j} \left(dist\left(S_i, S_j\right) \right), where\ (i,j) \epsilon N$$

Therefore, the D_{dist} indicates the distribution of semantic objects in terms of spatial distances. To minimize the match error between the reference scene and the target one, the greedy iteration procedure is applied. Finally, the land use type of the whole scene B_q can be determined with the minimized match error.

3. EXPERIMENT

3.1 Study areas and data preparation

In this section, we test the capability of the proposed MSCM-based scene recognition method to classify urban functional areas. To achieve this, a high-resolution dataset was acquired by Worldview-2 in 2010 with the ground spatial resolution of 0.5m. In order to verify the effectiveness of the proposed urban scene classification method, an experiment was conducted on a research area of approximately 56 km^2 in a certain area of Beijing. As the capital of China, Beijing is the representative of land cover and land use in China. Buildings with different styles and purposes can be observed in the image. It is difficult to accurately predict semantic objects with heterogeneous image patterns, especially in complex urban areas. The high-resolution images used in this experiment were captured on November 10, 2010. There are 8 spectral bands with a spatial resolution of 2.0 m and one full color band with a resolution of 0.5 m. The pan sharpening strategy is used to improve the spatial resolution of multispectral bands, including selecting blue, green, and near-infrared bands for image analysis.

Figure 3. Study areas located in Beijing, China.

In the CNN based classification stage, only five categories with significant visual differences were considered, namely vegetation, shadows, roof spacing, buildings, and roads. The size of the selected research area is 14264 pixels high × Width 12844 pixels. In order to divide the entire image into meaningful sub scenes, road lines were used to crop the image into rectangular blocks. In this study, 56 commercial scenes, 4 entertainment scenes, 18 public service scenes, 18 educational scenes, and 120 residential scenes were considered for semantic interpretation and classification. The reference map is a website obtained from the OpenStreetMap, which was released in June 2016. There are a total of 1742 independent buildings in the research area, with a total area of 2.8 km^2. In order to minimize the impact of temporal differences between OSM and remote sensing images, we visually examined land cover and made necessary edits in the OSM data. According to these land cover features, training samples can be randomly extracted from the corresponding images to improve the performance of depth learning. In addition, in order to obtain semantic labels at the object level, this study included 5145 POIs (provided by Baidu Maps in 2017). Based on the attributes of POI, four different semantic terms have been identified, namely commercial venues (such as shopping centers, restaurants), entertainment facilities (such as gyms, spas), public services (such as government agencies), and educational venues (such as universities).

3.2 Experimental designs

In this study, we employed a well-studied 5-layer configuration MSCM model. In order to automatically generate rich training samples, the land cover vector was recognized from OSM data (such as buildings and roads) to guide the extraction of samples from remote sensing images. We divided all available samples into training and testing samples in a 4:1 ratio. Set the size of the input sample to 28 ×28× 3. For the first convolutional layer, obtain 12×12×100 outputs. Then, the second convolutional layer contains 200 filters, each with a size of 3 × 3. Use 300 filters for convolution to generate 3 × 3× 300 feature maps. Finally, a fully connected dense layer with 5 hidden units is used to decompose it into 5 different labels in the softmax layer. In this step, only objects with significant visual differences, such as vegetation, shadows, roof slopes, buildings, and roads, are considered as the basic classification rule set. During the

training phase, normalization is used to set the learning rate to 0.0001 and the mini batch size to 500. In order to avoid over fitting, a dropout strategy was also implemented in the CNN framework. In the prediction stage, the depth features of each sample patch were extracted by MSCN and dissolved with the fully connected layer, and then classified as semantic labels. In order to demonstrate the robustness of the proposed method, a comparison was made between MSCM and Recurrent Neural Network (CNN).

Due to the limitations of Semantic information extraction, POI dataset was introduced to enrich the overall semantic content of high-resolution remote sensing images. In the last step, the high-resolution image was divided into several main types of land cover, such as buildings and roads. The POIs were directly integrated with the classification map, which supplements the Semantic information at the land use level. To achieve this, pixel level classification maps were converted into object based predictions. In order to better represent geographical objects, the image objects obtained through image segmentation were merged with the predicted pixel labels. In the merging step, the majority voting strategy was used to determine the Semantic information of the image object. After determining the geographical objects, introducing POIs further increased the semantic richness of the classification map. Specifically, we stacked POIs and image objects directly based on their geographic location. Considering the characteristics of POI and classified high-resolution images, we only considered POIs that define the purpose or purpose of the building. In this experiment, we considered five different types of buildings with different purposes and purposes, namely residential buildings, commercial buildings, entertainment buildings, public service buildings, and educational buildings.

Figure 4. POI spatial distribution

As mentioned above, although the CNN based image classification algorithm is effective for building detection, it ignores the functional design of buildings, so it still does not use the Semantic information of land cover. In order to enrich the semantic content of images, POI is introduced to increase the Semantic information of image classification. POI is represented by a series of spatial points with rich Semantic information. However, appropriate semantic labels should be used to accurately classify buildings as urban spatial units in order to obtain the semantic content of high-resolution images. First, POI contains rich Semantic information that can be used for image analysis. Secondly, the CNN based classification algorithm has accurately depicted complex buildings and other land cover from remote sensing images. In order to achieve semantic classification, it is necessary to integrate POI with the detected buildings. By utilizing semantic mapping, semantic content in urban scenes can be well studied. Each type of building represents a specific semantic content. Therefore, for a given urban scene, the number of semantic buildings in the scene is a direct indicator of classification.

3.3 Experimental Results

In this experiment, the classification based on MSCM considered five semantic elements (vegetation, water, roof slope, buildings, and roads). Then, using object based post-processing algorithms, pixel labels are converted into object based geographic objects using image fragments. To achieve semantic mapping, POIs are integrated with the detected object. Finally, land cover objects are further classified according to the functional design of commercial buildings, educational buildings, etc. In this process, the category of the scene is equal to the label of the most common land cover object, which conforms to a certain spatial distribution pattern. For example, if there are 56.8% commercial buildings in a city scene, with residential buildings accounting for only 21.3% of the total area and public service facilities accounting for 21.9%, then the city block will be mainly considered as a commercial district. There are a total of 10135 buildings in the research area, which can be divided into 216 urban blocks. Five different types of urban scenes were considered in the experiment, namely commercial areas, entertainment areas, public service areas, education areas, and residential areas.

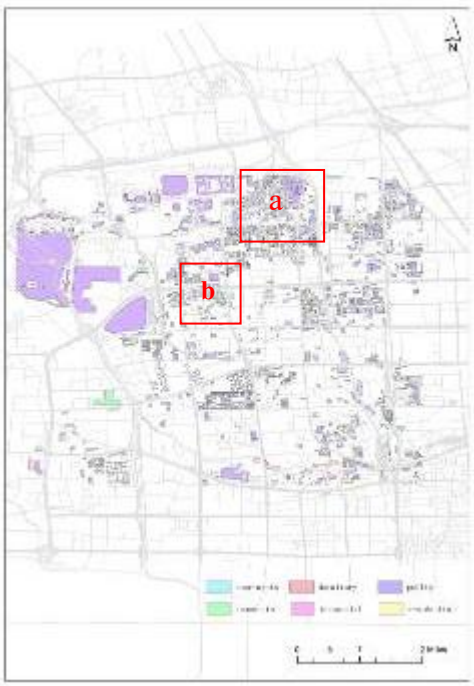
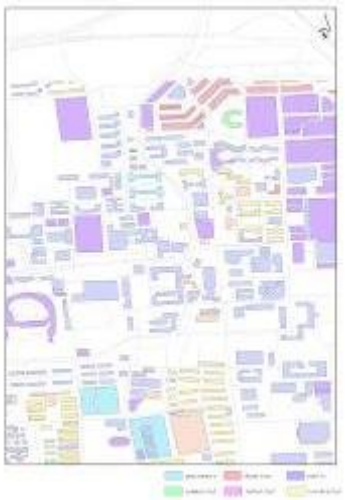


Figure 5. Classifications results of urban scenes of the Beijing dataset.

The classification results are illustrated in Figure 5 and Figure 6. The detailed information about classification accuracies is shown in Table 1. As reported in Table 1, the proposed method produces the highest accuracy in terms of urban functional classification. The proposed the MSCM method is very capable of carrying out urban functional classification, despite the heterogeneous images patterns may contain. For example, the classification accuracies of complex public services area are quite low for CNN (76%). The MSCM method shows a remarkable improvement in recognizing urban functional with classification accuracy as high as 80%. Moreover, the results showed that our method improved the accuracy and interpretability of urban functional classification, especially for fine-grained classification tasks such as urban functional zone recognition. The study demonstrated the potential of the proposed method, which could lead to improved accuracy and interpretability in various applications such as urban planning, environmental monitoring, and land resource management.



(a)

(b)

Figure 6. Classifications results of (a) and (b) urban scenes

Table 1. Classification results of urban functional

class	Methods/ F-1	
	Our method	CNN
Commercial area	88%	80%
Entertainment area	84%	79%
Public services area	80%	76%
Educational area	92%	87%
Residential area	93%	77%
Overall Accuracy	88%	79%
Kappa	0.85	0.78

4. CONCLUSION

We evaluated our proposed method on a High resolution remote sense dataset of urban functional zones. The experimental results show that our proposed method outperforms traditional methods and achieves high accuracy in identifying functional zones in a city. In addition, we conducted a sensitivity analysis to evaluate the impact of different factors on the performance of our method. The results show that our method is robust to changes in the input data and the network architecture.

The proposed method has several advantages over traditional methods. First, it is automated and does not require manual surveys, which can save time and cost. Second, it can provide a more comprehensive and accurate understanding of the spatial distribution of functional zones in a city. In future work, we plan to explore the use of additional data sources, such as weather data and air quality data, to further improve the accuracy of our method. Overall, our proposed method provides a promising approach for exploring urban functional zones based on multi-source semantic knowledge and cross-modal networks.

ACKNOWLEDGEMENTS

This research was supported by the National Key Research and Development Program of China (Grant No. 2022YFC3802903 and Grant No. 2022YFC3802904).

Dr. Jiage Chen conceived the idea of MSCM for urban functional classification. Mr. shu Peng constructed the experimental framework and gathered data sets of the study area; Dr. Shang Wei helped with the experiments and results analysis; Prof. Wenzhi Zhao offered help in paper revision and language proof-reading.

REFERENCES

Ma, L., Liu, Y., Zhang, X.L., et al., 2019. Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review. ISPRS Journal of Photogrammetry and Remote Sensing, 152:166-177.

Barlacchi, G., Rossi, A., Lepri, B., Moschitti, A., 2017. Structural Semantic Models for Automatic Analysis of Urban Areas, Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, 279-291.

Cheng, G., Yang, C., Yao, X., Guo, L., Han, J., 2018. When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs. IEEE transactions on geoscience and remote sensing 56, 2811-2821.

Li, Y., Zhang, H., Shen, Q., 2017. Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. Remote Sensing 9, 67.

Liao, F., Arentze, T., Molin, E., Bothe, W., Timmermans, H., 2017. Effects of land-use transport scenarios on travel patterns: a multi-state supernetwork application. Transportation 44, 1-25.

Liu, X., He, J., Yao, Y., Zhang, J., Liang, H., Wang, H., Hong, Y., 2017. Classifying urban land use by integrating remote sensing and social media data. International Journal of Geographical Information Science 31, 1675-1696.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2016. Fully convolutional neural networks for remote sensing image classification, Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International. IEEE, 5071-5074.

Rawat, J., Kumar, M., 2015. Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. The Egyptian Journal of Remote Sensing and Space Science 18, 77-84.

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., Mai, K., 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. International Journal of Geographical Information Science 31, 825-848.

Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and Remote Sensing Magazine 4, 22-40.

Zhao, W., Du, S., 2016b. Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing 54, 4544-4554.

Zou, Q., Ni, L., Zhang, T., Wang, Q., 2015. Deep Learning Based Feature Selection for Remote Sensing Scene Classification. IEEE Geosci. Remote Sensing Lett. 12, 2321-2325.

Zhao, W.Z., Bo, Y.C., Chen, J.G., et al., 2019. Exploring semantic elements for urban scene recognition: Deep integration of high-resolution imagery and OpenStreetMap. ISPRS Journal of Photogrammetry and Remote Sensing, 151:237-250.