SEMANTIC KNOWLEDGE EMBEDDING DEEP LEARNING NETWORK FOR LAND COVER CLASSIFICATION

Jiage Chen [1], Xiao Du[1], Junhui. Zhang[1]*, Yongtao Wan[1], Wenzhi Zhao[2]

[1] National Geomatics Center of China, 100830 Beijing, China - (jiagechen, duxiao, zhangjunhui, wanyt)@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: Semantic Knowledge, Deep Learning Network, Land Cover Classification.

ABSTRACT:

Land cover classification is essential basic information and key parameters for environmental change research, geographical and national monitoring, and sustainable development planning. Deep learning can automatically and multi-level extract the features of complex features, which has been proven to be an effective method for information extraction. However, one of the major challenges of deep learning is its poor interpret-ability, which makes it difficult to understand and explain the reasoning behind its classification results. This paper proposes a deep cross-modal coupling model (CMCM) for integrating semantic features and visual features. The representation of knowledge map is indicatively introduced into remote sensing image classification. Compared to previous studies, the proposed method provides accurate descriptions of the complex semantic objects within a complex land cover environment. The results showed that the integration of semantic knowledge improved the accuracy and interpret-ability of land cover classification.

1. INTRODUCTION

Land cover classification is critical as it has significant implications for a wide range of applications such as global climate change, natural resource management, food security and environmental monitoring. Remote sensing imagery has been widely used for land cover classification due to its large spatial coverage and temporal resolution. After evolving for several decades, many global or regional land-cover products have been successfully generated using satellite remote sensing image, and is currently experiencing a transformation from Coarse (300-10km) and moderate scales (30m) to much finer scales (10m) to provide more precise knowledge on land. These products include GlobeLand30, FROM_GLC, GLC_FCS30, Esri_Landcover_2020, ESA_WorldCover2020, and FROM_GLC10. These land cover products provide rich spatial information, thus offering greater flexibility for application in many fields.

However, some third-party researchers have found considerably lower accuracies for some regions when verifying land cover products, especially in areas with very complex spatial heterogeneity (Bojanowski et al., 2017). One possible reason is that most land cover data products use traditional machine learning classification methods, such as support vector machines, random forests, decision trees, and so on. These algorithms learn from a set of labelled data and can then classify new images based on the learned patterns. Moreover, these methods rely on manually defined rules and statistical models to classify land cover types based on spectral, textural, and contextual features. However, these methods are limited in their accuracy due to unstable feature extraction and a lack of spatial context information.

In recent years, deep learning has become an increasingly popular approach for land cover classification due to its ability to leverage large datasets and recent hardware advancements. Compared to traditional machine learning classification methods, deep learning methods can automatically extract high-level feature representations and fully utilize spatial context information (Ma et al., 2019; Marcos et al., 2018).

Common deep learning models include convolutional neural networks, recurrent neural networks, and fully convolutional neural networks for semantic segmentation (Canizo et al., 2019). Especially, convolutional neural networks (CNNs) have been widely used for feature extraction and land cover classification. It has gained significant attention in the land cover community and is outperforming other algorithms by a significant margin in some problems, especially in detecting fine-scale land cover types such as small artificial objects and perglacial landforms (Zhao et al., 2017; Zhao et al., 2019). With the increasing availability of high-resolution satellite images, the application of deep learning in land cover research is highly promising and expected to expand rapidly.

Although deep learning-based method have shown promising results in addressing this challenge by automatically learning features from the images, deep learning is essentially still a data-driven approach that relies heavily on the quantity of training samples, as well as the appropriate selection of model architecture and hyper-parameters. Moreover, one of the major challenges of deep learning is its poor interpretability, which makes it difficult to understand and explain the reasoning behind its classification results. To overcome this challenge, researchers have developed various techniques such as visualization tools and attribution methods to improve the interpretability of deep learning models. However, remote sensing images often contain complex patterns and structures that may be difficult for the model to extract meaningful features from.

With the development of deep learning and semantic web technologies, constructing a remote sensing knowledge has become a promising approach to improve the interpretability and applicability of remote sensing data (Li et al., 2021; Schonfeld et al., 2019). Remote sensing knowledge contains rich semantic relationship information and powerful reasoning ability, which can further enhance the interpretability of deep

---

* Corresponding author

learning models for remote sensing image interpretation, improve the reliability and precision of the results (Chen et al., 2020; Tao et al., 2019b). A knowledge graph is a structured representation of knowledge in which entities are represented as nodes and relationships between them are represented as edges (Dettmers et al., 2018). By incorporating remote sensing-specific knowledge, such as spectral signatures, spatial patterns, and contextual information, into a knowledge graph, it is possible to create a powerful tool for land cover classification (Devlin et al., 2018). Overall, it deserves much more exploration to promote the integration of remote sensing semantic knowledge and deep learning for land cover classification.

With the aforementioned considerations, this paper mainly focuses on exploiting the application of coupling knowledge map and deep learning network to land cover classification. The quality of remote sensing semantic knowledge plays an important role in land cover classification (Li et al., 2017a, b,c). To generate the high-quality semantic knowledge of remote sensing image, this paper constructs a relatively complete remote sensing knowledge map based on the prior knowledge of domain experts, where knowledge map fully considers the complex relationship between the main types of remote sensing images (He et al., 2020; Inglada et al., 2017).

Therefore, this paper, for the first time, the semantic representation of remote sensing knowledge map is carried out to solve the difference in representation between knowledge map and deep learning, so that it can be embedded in the deep learning network. Then, this paper proposes a new deep cross-modal coupling model (CMCM), which can well integrate the semantic features of the knowledge map and the visual features of the depth network. Experimental results show that our proposed CMCM is superior to traditional classification methods. The major contributions of this paper are summarized as follows.

1) In this paper, the representation of knowledge map is innovatively introduced into remote sensing image classification. Based on this, remote sensing knowledge map is used as a priori knowledge or rule to improve the accuracy of land cover classification. Extensive experiments verify its superiority compared with traditional classification methods.

2) In order to solve the difference between knowledge representation and deep learning representation, this paper proposes a deep cross-modal coupling model (CMCM) for integrating semantic features and visual features. The results showed that the integration of semantic knowledge improved the accuracy and interpretability of land cover classification.

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 CMCM 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: remote sensing knowledge graph construction and a deep cross-modal coupling model. We first construct a knowledge graph based on domain-specific knowledge and semantic relationships. The knowledge graph consists of nodes and edges, where each node represents a concept or entity, and each edge represents a semantic relationship between two nodes. We then use an embedding algorithm to map the nodes and edges of the knowledge graph to a low-dimensional vector space, where each node and edge is represented as a vector. This embedding process preserves the semantic relationships between nodes and edges in the

knowledge graph. We then embed the knowledge graph vectors into a deep neural network, which consists of multiple layers of artificial neurons. The neural network takes input data and learns to map it to the output labels. We modify the neural network architecture to include the knowledge graph embedding as an additional input to the network. The knowledge graph embedding guides the learning process of the neural network by providing additional semantic information and constraining the feature space of the model. The flow of processes within the proposed framework is depicted in Figure 1.
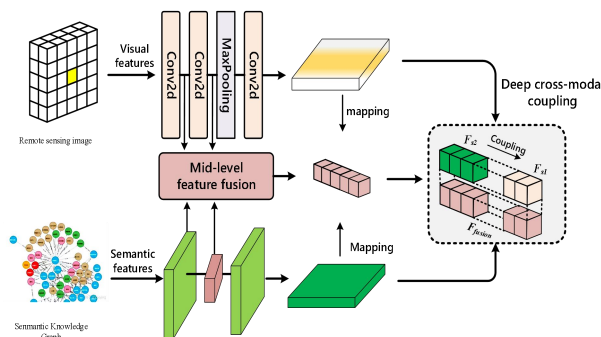

Figure 1. Flowchart of the proposed method.

2.1 Construction of Remote Sensing Knowledge Graph

A knowledge graph is a type of graph that represents knowledge as a set of nodes and edges. The nodes represent entities, such as concepts, objects, and events, while the edges represent relationships between the entities. In the context of remote sensing imagery, the entities may include image features, land cover types, atmospheric conditions, and data sources, while the relationships may include attribute relations, spatial relations between entities. The attribute relationship can be regarded as a child-parent relationship, for example, the relationship between the primary class and the secondary class, and the relationship between the secondary class and the tertiary class in the remote sensing classification system. In addition, the attribute relationship also includes the texture, shape, colour, width, height and other characteristics of the land class. Spatial relationships mainly include positional relation and topological relationships. The positional relationship mainly includes marked on, stop at, on; Topological relationships include surrounded by, intersecting, cross, meet, cover.

According to the above two relationships, in this paper, we use experts and scholars who are familiar with remote sensing image interpretation to construct a remote sensing knowledge map. The first step of knowledge atlas construction needs to complete the design and construction of ontology, which is crucial to the construction of knowledge atlas. In this paper, the artificial experience of experts is used as the basis of ontology construction to obtain entity categories, the relationship between categories, and the definition of attributes contained in entities.

Table 1. Classification results of main land cover types with

Different strategies.

head	relation	tail
farmland	airport	around
shrubwood	bare_land	in
weed	bare_land	is_part_of
grass	baseball_field	is_part_of
weed	cemetery	is_part_of

tree	christmas_tree_farm	in
circular_farmland	circular	shape
circular_farmland	green	color
cloud	white	color
cloud	irregular	shape
commercial_area	buildings	have
commercial_area	high-rise	height
commercial_area	dense	distribution
commercial_area	road	near
tree	convention_center	around
tree	dense_residential	in
weed	dense_residential	is_part_of
dense_residential	buildings	have
dense_residential	low-rise	height
dense_residential	dense	distribution
shrubwood	desert_chaparral	in
grass	golf_course	is_part_of

2.2 Deep Learning Model Construction

In this paper, we propose a method for coupling knowledge graphs with deep learning for model construction. Our method aims to leverage the semantic information and reasoning ability provided by knowledge graphs, while harnessing the power of deep learning to learn complex and abstract features from data. Specifically, we use a knowledge graph to represent domain-specific knowledge and semantic relationships, and then embed the knowledge graph into a deep neural network. The embedded knowledge graph can guide the learning process of the neural network and enhance the interpretability and reliability of the model.

We then embed the knowledge graph vectors into a deep neural network, which consists of multiple layers of artificial neurons. The neural network takes input data and learns to map it to the output labels. We modify the neural network architecture to include the knowledge graph embedding as an additional input to the network. The knowledge graph embedding guides the learning process of the neural network by providing additional semantic information and constraining the feature space of the model. Our method aims to leverage the semantic information and reasoning ability provided by knowledge graphs, while harnessing the power of deep learning to learn complex and abstract features from data. Specifically, we use a knowledge graph to represent domain-specific knowledge and semantic relationships, and then embed the knowledge graph into a deep neural network. The embedded knowledge graph can guide the learning process of the neural network and enhance the interpretability and reliability of the model.

Therefore, in order to couple multimodal features, we follow the architecture of VAE networks (Kingma and Welling, 2013) to learn a reconstruction model for visual features and semantic representations, which projects visual features and semantic representations into potential spaces. The loss function can be defined by an Eq.(1).

$$
\begin{aligned}
L_v = {} & E_{q_{\emptyset^{(v)}}}\big(z^{(v)}\big|x\big)\big[\log p_{\theta^{(v)}}\big(x\big|z^{(v)}\big)\big] \\
& - D_{KL}\big(q_{\emptyset^{(v)}}\big(z^{(v)}\big|x\big)\|p_{\theta^{(v)}}\big(z^{(v)}\big)\big) \\
& + E_{q_{\emptyset^{(a)}}}\big(z^{(a)}\big|c\big)\big[\log p_{\theta^{(a)}}\big(c\big|z^{(a)}\big)\big] \\
& - D_{kl}\big(q_{\emptyset^{(a)}}\big(z^{(a)}\big|c\big)\|p_{\theta^{(a)}}\big(z^{(a)}\big)\big)
\end{aligned}
$$

$$(1)$$

Where x represents the original image, $q_{\emptyset^{(v)}}$ and $p_{\theta^{(v)}}$ corresponds to the encoder and decoder of visual features respectively, c represents the semantic representation, $q_{\emptyset^{(a)}}$ and $p_{\theta^{(a)}}$ corresponds to the encoder and decoder of semantic representations respectively.

Here, visual features and semantic representations are cross input to the encoder corresponding to another mode, and the loss function of cross modal feature reconstruction can be defined by the Eq.(2).

$$
L_{oss} = \sum_{n=1}^{N}\Big|x_n - p_{\theta^{(v)}}\big(q_{\emptyset^{(a)}}(C_n)\big)\Big| + \Big|C_n - p_{\theta^{(a)}}\big(q_{\emptyset^{(v)}}(x_n)\big)\Big|
$$

$$(2)$$

Where N represents the number of training samples, x_n and C_n represent visual features and semantic representations of the same category.

3. EXPERIMENT

3.1 Study areas and data preparation

To illustrate the classification ability of the proposed method, the experiments are conducted on two data sets for algorithm evaluation. Specifically, the selected two data sets are season-varying images, and they have different spatial heterogeneity characteristics. Fig.1 shows the study area (1 and 2) with a spatial extent of approximately 95,300 km^2.

The first study area is located in Wuhan, China. Wuhan is located in the middle and lower reaches of the Yangtze River Plain, in the eastern part of the Jianghan Plain. It is a national regional central city (Central China), a sub-provincial city and the capital of Hubei Province. Located at 113°41′-115°05′ east longitude, 29°58′-31°22′ north latitude. In China's economic geography circle, Wuhan is in a superior central position, just like Tianyuan on the Go board, and is known as the "heart" of China's economic geography. Wuhan is a humid subtropical monsoon climate zone, characterized by abundant rainfall and sunshine, four distinct seasons, high temperature in summer, concentrated precipitation, and cool and humid winter.
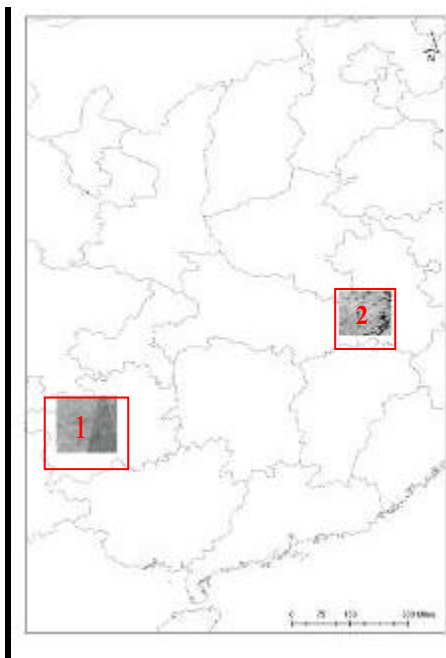
Figure 2. Study areas located in Guizhou (1) and Wuhan (2), China. The marked red subareas (1 and 2) were used for evaluating the fusion network performance. Subareas 1 is with a spatial extent of 250 km × 230 km. Subareas 2 is with a spatial extent of 210 km × 180 km.

The second area is located in Guizhou, China. The landform of Guizhou belongs to the plateau mountainous area in the south-western part of China. The terrain in the territory is high in the west and low in the east. It slopes from the central part to the north, east and south, with an average altitude of about 1100 meters. Most of the Guizhou plateau is mountainous, known as "eight mountains, one water and one field". The climate in Guizhou is warm and humid, belonging to the subtropical humid monsoon climate. The temperature changes little, the winter is warm and the summer is cool, and the climate is pleasant.

For the optical data set, we downloaded a total of 3 Sentinel-2 images that span from January 1, 2022 to December 27, 2022, (including Sentinel-2a and Sentinel-2b). At the same time, the preprocessing was performed to extract optical image features, including radiation calibration, cloud mask, atmospheric correction, and calculation of NDVI. Meanwhile, in order to reduce the impact of noise, only cloudless or partially clouded images were included in the Sentinel-2 data set.

3.2 Experimental Results

In this section, the experiment details are explained. First, a 4-layer CMCM framework was trained based on the Sentinel-2 dataset. Then, the well-trained CMCM model was applied to predict the labels for the entire image. In order to keep a balanced training samples, we selected 150 samples for each class for the model training. After the training process, we applied the well-trained model to extract deep features and obtain the initial classification results. Also, the segmentation scale was set to 30, in order to get accurate information about shapes and edges of geographical objects. With the integration of image segments and the model-based classification results, the classification accuracy can be further improved. The proposed framework was evaluated using 10m-resolution

remote sensing images and compared with traditional Convolutional neural network (CNN). Moreover, the CNN method was included to demonstrate the effectiveness of the proposed method. In this experiment, four types of land cover (including Cropland, Water bodies, Tree cover and Built-up) were considered during the CMCM-based classification. The classification results are illustrated in Figure 3 and Figure 4. The detailed information about classification accuracies is shown in Table 2 and Table 3.

As reported in Table 2 and Table 3, the proposed method produces the highest accuracy in terms of land cover classification. The proposed the CMCM method is very capable of carrying out land cover classification, despite the heterogeneous images patterns may contain. For example, the classification accuracies of complex cropland and tree cover are quite low for CNN (73% and 70%). The CMCM method shows a remarkable improvement in recognizing cropland with classification accuracy as high as 86%. Moreover, the results showed that our method improved the accuracy and interpretability of land cover 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)

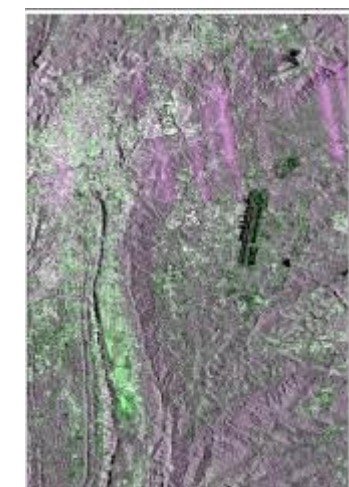**Table 2**. Wuhan: Classification results of main land cover types with different strategies.

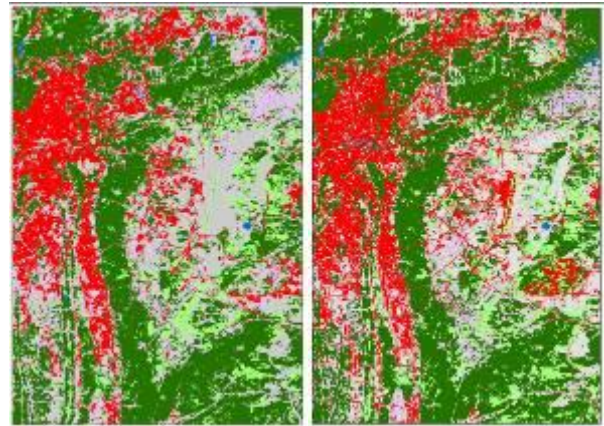| class	Methods/ F-1	
	CMCM	CNN
Cropland	86%	82%
Water bodies	85%	83%
Tree cover	80%	79%
Bulit-up	92%	70%
Overall Accuracy	83%	78%
Kappa	0.89	0.81

(b)

(c)

Figure 3. Classification maps of the Sentinel-2 with different strategies in Wuhan : (a) NDVI image, (a) CMCM, (b) CNN

(a)

(b)   (c)

Figure 4. Classification maps of the Sentinel-2 with different strategies in Guizhou: (a) SAR image, (b) CMCM, (c) CNN

Table 3. Guizhou: Classification results of main land cover types with different strategies.

class	Methods/ F-1	
	CMCM	CNN
Corpland	76%	73%
Water bodies	90%	83%
Tree cover	78%	70%
Bulit-up	81%	77%
Overall Accuracy	81%	76%
Kappa	0.85	0.78

4. CONCLUSION

We propose a strategy to classify complex land cover by integrating remote sensing images semantic knowledge with deep learning model. Unlike most previous studies, we have designed a deep cross-modal coupling model (CMCM) for integrating semantic features and visual features. Based on the sensing knowledge graph, we further extracted semantic information from the remote sensing images. In this way, the CMCM method is the first algorithm to classify land cover using the semantic knowledge and cross-modal model. Compared to previous studies, the proposed method provides accurate descriptions of the complex semantic objects within a complex land cover environment. In conclusion, the construction and utilization of knowledge graphs for remote sensing applications is an emerging and promising research direction. It has the potential to enhance the accuracy and interpretability of remote sensing analysis and facilitate decision-making in various fields. Future research could focus on further improving the performance and efficiency of knowledge graph-based methods and exploring their applications in other areas of remote sensing analysis. Although, the deep integration of the semantic knowledge and deep learning model has greatly improved classification accuracy in terms of land cover classification. There are still some problems that need further improvements. For instance, it is difficult to construct a semantic knowledge graph based on remote sensing images due to the lack of fixed rules or prior knowledge. The construction of semantic knowledge graph is a huge and complex project, making it difficult to construct a complete and

accurate knowledge graph. It involves ecological geographical characteristics, social environment, and expert prior knowledge.

REFERENCES

Bojanowski, P., Grave, E., Joulin, A., Mikolov, T., 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics 5, 135–146.

Canizo, M., Triguero, I., Conde, A., Onieva, E., 2019. Multi-head CNN‐RNN for multi-time series anomaly detection: an industrial case study. Neurocomputing. 363, 246‐260.

Chen, T., Kornblith, S., Norouzi, M., et al., 2020. A simple framework for contrastive learning of visual representations. International conference on machine learning. PMLR 1597–1607.

Dettmers, T., Minervini, P., Stenetorp, P., Riedel, S., 2018. Convolutional 2d knowledge graph embeddings. In: Proceedings of the AAAI Conference on Artificial Intelligence.

Devlin, J., Chang, M., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv: 1810.04805v2.

He, K., Fan, H., Wu, Y., 2020. Momentum contrast for unsupervised visual representation learning. In: In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 9729–9738.

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., Rodes, I., 2017. Operational high resolution land cover map production at the country scale using satellite image time series. Remote Sens. 9, 95.

Li, Y., Kong, D., Zhang Y., et al., 2021. Robust Deep Alignment Network with Remote Sensing Knowledge Graph for Zero-Shot and Generalized Zero-Shot Remote Sensing Image Scene Classification. ISPRS Journal of Photogrammetry and Remote Sensing, 179:145-158.

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.

Marcos, D., Volpi, M., Kellenberger, B., Tuia, D., 2018. Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models. ISPRS Journal of Photogrammetry and Remote Sensing 145, 96–107.

Schonfeld, E., Ebrahimi, S., Sinha, S., Darrell, T., Akata, Z., 2019. Generalized zero- and few-shot learning via aligned variational autoencoders. In: In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8247–8255.

Tao, Chao, Qi, Ji, Li, Yansheng, Wang, Hao, Li, Haifeng, 2019b. Spatial information inference net: road extraction using road-specific contextual information. ISPRS Journal of Photogrammetry and Remote Sensing 158, 155–166.

Zhao, W., Du, S., Wang, Q., Emery, W.J., 2017. Contextually guided very-high resolution imagery classification with semantic segments. ISPRS-J. Photogramm. Remote Sens.132, 48‐60.

Zhao, W., Bo, Y., Chen, J., Tiede, D., Thomas, B., Emery, W.J., 2019. Exploring smantic elements for urban scene recognition: deep integration of high-resolution imagery and OpenStreetMap (OSM). ISPRS-J. Photogramm. Remote Sens. 151, 237‐250.