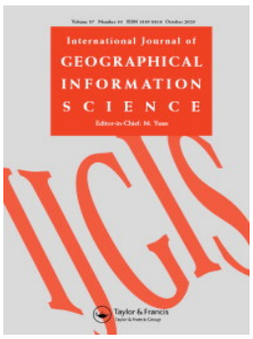

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Unsupervised land-use change detection using multi-temporal POI embedding

Yao Yao, Qia Zhu, Zijin Guo, Weiming Huang, Yatao Zhang, Xiaoqin Yan, Anning Dong, Zhangwei Jiang, Hong Liu & Qingfeng Guan

To cite this article: Yao Yao, Qia Zhu, Zijin Guo, Weiming Huang, Yatao Zhang, Xiaoqin Yan, Anning Dong, Zhangwei Jiang, Hong Liu & Qingfeng Guan (26 Sep 2023): Unsupervised land-use change detection using multi-temporal POI embedding, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2023.2257262

To link to this article: https://doi.org/10.1080/13658816.2023.2257262

Published online: 26 Sep 2023.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2023.2257262

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Unsupervised land-use change detection using multi-temporal POI embedding

Yao Yao[a,b,c], Qia Zhu[a], Zijin Guo[a], Weiming Huang[d], Yatao Zhang[e], Xiaoqin Yan[f], Anning Dong[a], Zhangwei Jiang[g], Hong Liu[g] and Qingfeng Guan[a]

[a]School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei Province, China; [b]Center for Spatial Information Science, The University of Tokyo, Kashiwa-shi, Japan; [c]Guangdong – Hong Kong – Macau Joint Laboratory for Smart Cities, Shenzhen, China; [d]School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore; [e]ETH Zurich, Future Resilient Systems, Singapore-ETH Centre, Singapore, Singapore; [f]Institute of Remote Sensing and Geographical Information Systems, School of Earth and Space Sciences, Peking University, Beijing, China; [g]Alibaba Group, Hangzhou, Zhejiang Province, China

ABSTRACT

Rapid land-use change detection (LUCD) is pivotal for refined urban planning and management. In this paper, we investigate LUCD through learning embeddings of points of interest (POIs) from multiple temporalities. There are several prominent challenges: (1) the co-occurrence problem of multi-temporal POIs, (2) the heterogeneity of POI categorization, and (3) The lack of human-crafted labels. Therefore, multi-temporal POIs need to be aligned in the embedding space for effective LUCD. This study proposes a multi-temporal POI embedding (MT-POI2Vec) technique for LUCD in a fully unsupervised manner. In MT-POI2Vec, we first utilize random walks in POI networks to capture their single-period co-occurrence patterns; then, we leverage manifold learning to capture (1) single-period categorical semantics of POIs to enforce semantically similar POI embedding to be close and (2) cross-period categorical semantics to align multi-temporal POI embedding in a unified embedding space. We conducted experiments in Shenzhen, China, which demonstrates that the proposed method is effective. Compared with several baseline models, MT-POI2Vec can better align multi-temporal POIs and thus achieve higher performance in LUCD. In addition, our model can effectively identify areas with unchanged land use and land use changes in residential and industrial areas at a fine scale.

ARTICLE HISTORY
Received 18 May 2023
Accepted 5 September 2023

KEYWORDS
Land-use change; embedding space alignment; points of interest; POI embedding

1. Introduction

Urban land-use change detection identifies differences in land-use distribution and composition over time (Viana *et al.* 2019), which is a crucial indicator for urban management and socioeconomic assessment. Currently, urban land use is changing in many parts of the world, and the magnitude of the change is influenced by a variety

CONTACT Weiming Huang weiming.huang@ntu.edu.sg

of socioeconomic and biophysical factors, including deforestation, urban expansion, and agricultural intensification (Islam *et al.* 2018). With the continuous development of China's economy and the accelerated pace of urbanization, a large amount of land has been continuously developed and expropriated, resulting in significant changes in the original land cover in a relatively short period of time (Aderele *et al.* 2020). Highly dynamic changes in urban land-use pose challenges to policy-makers and urban planners in managing emerging urban communities. Therefore, effectively identifying urban land-use changes is essential for optimizing urban land-use patterns and promoting sustainable urban development (Wu *et al.* 2021).

Remote sensing images have been proven to be one of the important data sources for monitoring and analyzing land use change over time (Asokan and Anitha 2019). Detecting land use change using remote sensing images consists of the post-classification change detection method and the direct change detection method. The post-classification change detection method first classifies multi-temporal remote sensing images, and then the classification results of different periods are compared and analyzed to derive the final results (Zhu *et al.* 2022). Classification methods include maximum likelihood classification (Samanta and Paul 2016, Halimi *et al.* 2018, Prabu and Dar 2018, Pramanik and Punia 2020), support vector machines (SVMs) (Zewdie and Csaplovies 2015, Nandam and Patel 2022), and deep neural networks (Henry *et al.* 2019, Alhassan *et al.* 2020, Feizizadeh *et al.* 2021), etc. The post-classification change detection method is simple and intuitive, but the results of change identification are strongly influenced by the classification accuracy. In addition, the post-classification change detection method based on remotely sensed images relies on supervised learning; thus, it requires a large number of manually produced labels, which in many cases are expensive or even impractical. In contrast, the methods of direct change detection include the spectral-based comprehensive change detection method (CCDM) (Jin *et al.* 2013), change vector analysis method (CVA) (Jung and Chang 2015, Singh *et al.* 2021), and object-based backtracking method (Yu *et al.* 2016). The direct change detection method alleviates this problem, as it avoids classification and cumulative errors (Xiaolu and Bo 2011).

However, remote sensing images can mainly reflect the external physical attributes of the study area but fail to reflect the socioeconomic characteristics and human activities of urban areas (Yao *et al.* 2017). The emergence of big data makes it possible to identify refined land-use types by introducing socioeconomic attributes, such as cell phone positioning data, social media data, cab track data, street view images, and geotagged photos (Cao *et al.* 2020). Many studies have explored how to utilize the socioeconomic attributes of big data to identify urban land-use. For example, Pei *et al.* (2014) provided new insights into aggregating time-series cell phone data for land-use classification. Xing and Meng (2018) integrated landscape metrics with socioeconomic characteristics extracted from crowdsourced data to classify functional urban areas. Hu *et al.* (2021) used a large quantity of cab track data to classify urban functions at the road segment level. Zhang *et al.* (2021) quantified the trajectories of residents as high-dimensional semantic vectors to model the relations between mixed urban land-use. In addition, fusing remote sensing images with the data mentioned above can better recognize urban land use from both the 'top-down' and 'bottom-up' perspectives (Yao *et al.* 2020). For example, Zhang *et al.* (2019) introduced interlinking

mechanisms in identifying urban land use using multisource data to achieve a natural correspondence between remote sensing images, POIs, and real-time social media user data. Cao *et al.* (2020) proposed an end-to-end deep multimodal fusion network to fuse remote sensing images and socially sensed feature data for the asynchronous problem of multimodal multisource data. Bai *et al.* (2023) utilized contrastive learning for fusing remote sensing data and POIs in an unsupervised manner, and the learned embeddings can be used for multiple applications like sensing land use.

Among various big data sources depicting urban land-use, POIs are able to reflect socioeconomic attributes and are easier to obtain than other data sources like remote sensing images and human trajectories (Yao *et al.* 2017, Huang *et al.* 2023). A growing number of studies show that the spatial contextual relations of POIs can be used to mine deep semantic land-use information (Yao *et al.* 2017, Zhai *et al.* 2019, Huang *et al.* 2022). Initially, the frequency characteristics of POIs in the region were used to explore land-use classification (Tian and Shen 2011, Jiang *et al.* 2015). However, such methods ignore the spatial interplay among POIs, resulting in compromised effectiveness in mining land-use information. To solve this problem, Yao *et al.* (2017) utilized the Word2Vec model (Mikolov *et al.* 2013) into POI-based land-use classification and established an innovative framework to detect urban land-use distribution at the scale of traffic analysis zones (TAZs). This approach yields a good representation of features but ignores the influence of the first law of geography, which emphasizes that POIs are more associated with points geographically closed (Zhai *et al.* 2019). Yan *et al.* (2017) then proposed a Place2Vec model to consider the first law of geography in constructing the POI corpus. Zhai *et al.* (2019) used the Place2Vec model for the extraction and identification of urban functional areas and achieved good accuracy. These two methods mainly focus on capturing the spatial co-occurrence relations of POIs but fail to capture the intrinsic categorical semantic information of POIs. To tackle this problem, Huang *et al.* (2022) proposed a new method to estimate the proportional distribution of functional types in cities by learning the embeddings of POIs that preserve the semantic relatedness between POI categories according to the category hierarchy.

The above methods based on POI embedding justify that using deep-representing methods to mine POIs is effective for identifying urban land-use. Such an observation raises a natural question, namely whether we could utilize POIs – a single modality data - from multiple temporalities and learn their embeddings for LUCD, which is a particularly interesting avenue in light of the availability of such data. Considering the heterogeneity of POI categories in different years (implying that direct comparison is not possible), a straightforward way of using POI embeddings for LUCD is to separately embed POIs in different years, and such embeddings can then be aggregated to region (land parcel) level to produce region embeddings in different years, which can subsequently be used for LUCD. However, such a straightforward manner omits a technical pitfall, ie the separately learned POI embeddings in different years will end up in isolated embeddings spaces, making the unsupervised land use change detection infeasible. On the other hand, the embedding spaces in different years could be unified if we consider cross-year POI co-occurrence relations, whereas we are reluctant to incorporate such unreliable relations, eg we cannot regard a restaurant in 2013 and a hotel in 2020 as a co-occurred pair, even though they are rather close geographically. In this context, what matters the

most is finding a glue between the two isolated embedding spaces of different years. In fact, such a problem, to the best of our knowledge, has barely been addressed before, and what can serve as a glue for aligning the embedding spaces of POIs in different years remains a challenge. In addition, through aligning the embedding spaces, we could accomplish unsupervised LUCD by comparing the region embeddings in different years, which is particularly useful, as such ground truth data is rarely available.

In this paper, we come up with a solution to utilize the cross-year first-level categories as the glue, with two merits: (1) a POI dataset usually only has 20–30 first-level categories, making the manual alignment light-weight, viable and less painful, and (2) the semantics of first-level categories is generic enough to ensure the correctness of the manual alignment, while aligning second-level categories would lead to the difficulties in fuzzy and subtle semantic differences. With such a principle, we propose the MT-POI2Vec model and construct a framework for LUCD with a fully unsupervised approach. Adapted from the Semantic model proposed in Huang *et al.* (2022), we utilize random walks to capture within-year POI co-occurrence relations and use a manifold learning method to capture the semantic relations from two perspectives: (1) the single-year semantic relations according to the POI category hierarchy (the embeddings of the categories sharing the same upper-level category should be close), and (2) the cross-year category alignment through manual alignment for the first-level POI categories (if the first-level categories of two POI categories in different years are manually aligned, then they should be close in the embedding space). Subsequently, we average the POI embeddings over the parcels to obtain the parcel representation embeddings, derive the land-use type differences of the parcels in different years and compare them with the ground truth for understanding the effectiveness of the proposed MT-POI2Vec. The parcel change embeddings are clustered to obtain the pattern distribution of land-use changes. This study takes Shenzhen, China as the study area and uses two periods of POIs for the experiment.

Our key contributions are (1) we find the first-level categories as a glue to accomplish embedding space alignment for POIs in different temporalities, which has barely been explored before and (2) we first use POI embeddings to detect land-use changes with a fully unsupervised approach, overcoming the problem of difficulty in acquiring labeled data and avoiding the incremental error caused by multiple classification processes. The rest of the paper is organized as follows. Section 2 introduces the study area and data. In Section 3, we present our MT-POI2vec model and a framework for identifying land-use change based on this model. In Section 4, we obtain the clustering distribution of POI category embeddings in 2D space, the cosine distance distribution at the parcel scale, as well as the clustering distribution of the land use change embeddings and then carry out the field validation by using remote sensing images. In Section 5, we discuss the feasibility of our method and several important issues that can be further improved. Finally, we conclude in Section 6.

2. Study area and dataset

The study area is Shenzhen, China (Figure 1). Shenzhen is one of the most developed city in southern China, with a permanent population of 17,681,600 in 2021. The central

urban areas of Shenzhen include Futian District, Luohu District, Nanshan District, Bao'an District, Yantian District, Longgang District, Longhua District, Pingshan District, Guangming District, and Dapeng District. The first four administrative regions belong to the central areas of Shenzhen. Land-use planning parcels are the basic units of urban cadastral management in China. The data of actual land-use planning parcels in Shenzhen used for the study are obtained from the overall plan for land use of Shenzhen (2010–2020) by the Bureau of Planning and Natural Resources of Shenzhen Municipality (https://pnr.sz.gov.cn/), with a total of 6913 actual parcels and an average area of approximately 0.11 square kilometers.

POI refers to geographic points with attribute labels and locations (Zhang *et al.* 2019). The POI data used in this paper come from Baidu Map, the largest map service provider in China, which is widely used in urban research (Yao *et al.* 2017). Our research group collected the POIs of Shenzhen in 2013 and 2020 respectively using Baidu Map's API (https://lbsyun.baidu.com/), and the numbers of POIs in the two periods are 199,748 and 314,138, respectively. To distinguish the POI categories with the same name in the two phases of data, 'A' and 'B' are added after the same type of POI in 2013 and 2020, respectively.

3. Methodology

Figure 2 shows the study workflow. To summarize, (1) after data cleaning of the two periods of Baidu POI data, the POIs are spatially joined with the parcel data to mark the parcel to which each POI point belongs. (2) The MT-POI2Vec embedding model is constructed, which represents the POIs in different periods as vector embeddings in low-dimensional embedding space, containing embedding of POI spatial co-occurrence relations and embedding of POI categorical semantics. (3) Parameter sensitivity analysis is performed on the MT-POI2Vec model to determine the best parameters. The MT-POI2Vec

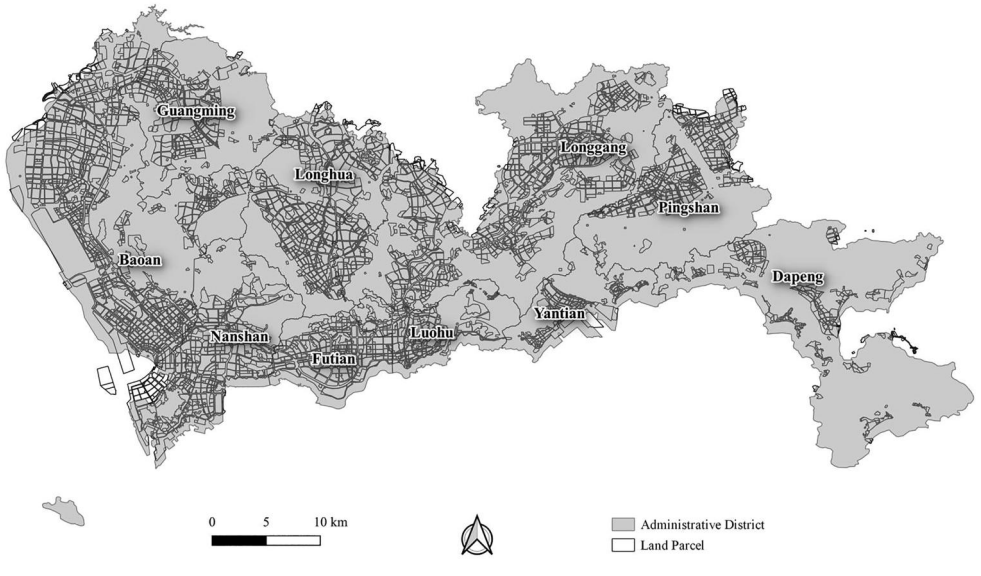

Figure 1. Case study area: Shenzhen, China.

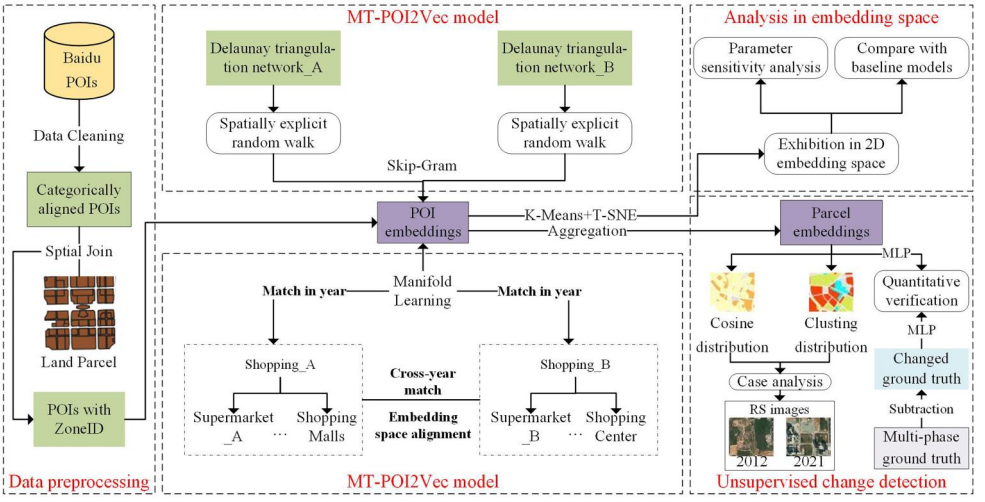

Figure 2. The workflow architecture of the proposed approach.

model is then qualitatively compared with the baseline models to analyze the superiority of the MT-POI2Vec model. (4) The cosine distance between two-period parcel embeddings is mapped onto map space to obtain a cartographic representation of land-use change based on cosine distance and then the effectiveness of the MT-POI2Vec model to represent land-use change is analyzed in comparison with the ground truth. (5) The embedding representing the land-use change in each parcel is obtained by calculating the vector difference between the two-period parcel embeddings and then clustered to analyze the pattern distribution of land-use change. (6) In order to quantitatively evaluate our method, we use the learned embeddings to estimate the LUCD in a supervised manner, against several baseline models.

3.1. Random walks based on the Delaunay triangle network

The Semantic model proposed by Huang *et al.* (2022) is a POI embedding model representing both the POI co-occurrence patterns and the categorical semantics POIs. The Semantic model consists of two main components: a random walk to construct POI sequences based on the Delaunay triangle network, and the embedding of categorical semantic information in the training process. In fact, this method of embedding POIs mimics techniques in natural language processing (NLP). In NLP, the idea is mainly that the words often occurring closely in eg sentences should be more relevant than the words that barely co-occur. This idea can be naturally extended to POI categories, meaning that the POI categories that often appear closely are highly related, and should be embedded closely in the embedding space. For example, shopping mall and restaurant usually co-occur, and they, together, reflect certain land use types (commercial); however, natural reserve and kindergarten usually do not co-occur, and their embeddings should be set apart to boost the capacity of discerning their corresponding land use types. Based on this idea, we need to capture the co-occurrence relationship between POIs first. Based on the Semantic model, this study first constructs a POI network using the Delaunay triangulation (DT) network for the POIs in

each year. Previous studies have shown that Delaunay triangulation is suitable for representation learning of spatial vector data (Yan *et al.* 2019). Specifically, to generate DT, the POIs are represented as vertices in the networks, and the line connecting the two POIs is regarded as an edge. Since the two POIs are in different spatiotemporal locations and have no spatial co-occurrence relations, Delaunay triangular networks are constructed separately for each period (year) to be applied in the random walk process. Each node is set to have 3 random walking paths, and the length of each random walking path is 5. Multiple traversals are performed starting from each node, and then multiple POI sequences can be obtained. To make the POI corpus richer but not overly complex, we use the second-level category to represent each POI in the network, and each POI category can be considered a word in the corpus.

Then, the POI sequences were trained using the skip-gram model (Mikolov *et al.* 2013). We optimize the training process using a negative sampling technique, which essentially entails minimizing the objective function:

$$\mathcal{L}_{co-occurrence} = \sum_{t=1}^{H} \sum_{-w \le j \le w} -\left(\log\left(\sigma\left(\boldsymbol{v}_t^T \boldsymbol{v}_{t+j}\right)\right) - \sum_{i=1}^{Q} \log\left(\sigma\left(\boldsymbol{v}_t^T \boldsymbol{v}_N\right)\right) \right) \tag{1}$$

where H are all the categories in the sequence captured by random walk, w is the window size of the central word for locating the surrounding words, σ is the sigmoid function, $\boldsymbol{v}_t$ represents the t-th vector embedding of the central word in the sequence, and v_{t+j} represents the $t+1$-th vector embedding of the context word of the central word. Q is the number of negatively sampled words for each central word, and $\boldsymbol{v}_N$ is the N-th vector embedding of each negatively sampled word.

3.2. Embedding cross- and single-year categorical semantics based on manifold learning

Since the two periods of POIs are in different temporalities, the POIs from the two years are trained separately in skip-gram, which means that we do not incorporate cross-period POI co-occurrence. In this context, the POI embeddings of different years are not in the same embedding space. Hence, the semantic relations between POI categories in different years can barely be reflected by the embeddings. For example, we might have two embeddings for restaurants in 2013 and 2020; the two embeddings are likely to be separate due to the separated training processes, whereas in principle, they should reside closely to capture their semantic similarity for change detection.

Therefore, it is necessary to pull the POI vector embeddings of different years into the same embedding space so that cross-year POI embeddings with similar categorical semantics are close to each other, which is similar to the local invariance assumption in the manifold learning theory (Guo *et al.* 2015). Therefore, we borrow the technical solution from Huang *et al.* (2022), and utilize a manifold learning algorithm, ie the eg Laplacian Eigenmaps (Belkin and Niyogi 2001) to help us to achieve this embedding goal. Specifically, we align the embedding spaces of the two temporalities by making the embeddings of POI categories from different years but entail similar semantics close in a unified embedding space. As we are training the embeddings for the second-level POI categories, we need to specify every single pair of cross-year POI

second-level categories that share similar semantics. However, this is a laborious process, as the POIs in each year belong to hundreds of second-level categories, making manual alignment impractical.

To solve this problem, we propose to utilize the first-level POI categories as the bridge between the two temporalities, ie we manually align the first-level categories in the two different years, and all the second-level categories under the aligned first-level categories can be paired with each other (only cross-year second-level categories are paired in this step). In this way, the manual alignment process becomes more light-weight and less painful, as there are usually only around 20 first-level categories in each year, the semantics of first-level categories are generic enough to be matched with high confidence, and the names of first-level categories are more static across years compared to second-level ones.

In practice, there are also unusual cases where only aligning first-level categories results in semantic discrepancy, which is mainly due to the changes made to the POI categories over the years. Therefore, after the first-level manual alignment, we refine the resulted second-level pairs through a further round of manual inspection. For example, 'Leisure & Recreation_A' and 'Leisure & Recreation_B' are two first-level categories in 2013 and 2020 with the same name. 'Fitness Centre_A' and 'Stadiums and Gym_A' subsumed by 'Leisure & Recreation_A' are semantically similar to 'Fitness Centre_B' and 'Stadiums and Gym_B' subsumed by 'Sports & Fitness', which is a different first-level categories than 'Leisure & Recreation_B'. We define these semantically similar second-level categories in such unaligned first-level as special cross-year second-level category pairs, which are also aligned manually as a refinement process.

Specifically, the cross-year second-level POI category pairs used in the manifold learning process for embedding space alignment are constructed in the following steps: (1) manually align first-level categories with similar semantics from two periods; (2) manually align the special second-level categories separately; (3) the categories belonging to the pairing results of the special cross-year second-level categories are removed from the paired first-level categories; (4) the remaining cross-year second-level categories are paired with each other according to the pairing results of the first-level categories. The matching principle of cross-year POI second-level categories is that they have similar functional semantics, even if the first-level categories they belong to do not match each other. By this method, we paired a total of 16 cross-year first-level category pairs and 26 special cross-year second-level category pairs. Finally, all the cross-year second-level category pairs consist of the matching results of the second-level categories respectively included in the matched first-level categories from two periods and the special cross-year second-level category pairs.

We utilize the cross-year pairs to manipulate the embeddings spaces of the two years, thereby making them aligned. At the meantime, we also construct within-year second-level category pairs to preserve the semantics entailed from the category hierarchy of POIs in each year. This is in line with the method of semantic preservation for POI category embeddings in Huang *et al.* (2022), ie if two second-level categories share the same first-level category (eg the second-level categories Chinese restaurant and Western restaurant both belong to the first-level category Food), we also utilize Laplacian Eigenmaps to make them close in the embedding space.

With the constructed cross-year and single-year category pairs, we can then utilize Laplacian Eigenmaps for both the embedding space alignment between two temporalities, and the preservation of categorical semantics within each year.

For each year, the embeddings of POI categories sharing the same upper-level categories are semantically smoothed to capture the semantics carried by POI categorical hierarchies. The calculation is based on the following Equation (2):

$$\mathcal{L}_{single-period-smoothing} = \frac{1}{2}\left(\sum_{i=1}^{m}\sum_{j=1}^{m}||\boldsymbol{v}_{A_i} - \boldsymbol{v}_{A_j}||_2^2 + \sum_{i=1}^{n}\sum_{j=1}^{n}||\boldsymbol{v}_{B_i} - \boldsymbol{v}_{B_j}||_2^2\right)S_{ij} \tag{2}$$

where $\boldsymbol{v}_{Ai}$, $\boldsymbol{v}_{A_j}$ and $\boldsymbol{v}_{Bi}$, $\boldsymbol{v}_{B_j}$ are the i-th and j-th vector embeddings of the second-level categories in two phases, respectively, m, n is the total number of second-level categories of POIs from two periods respectively, and S_{ij} is used to measure the semantic smoothness of the embedding space. In each year of POIs, if two second-level categories belong to the same first-level category, the value of S_{ij} is 1; otherwise, it is 0.

For cross-year embedding space alignment, the embeddings of POI categories belonging to the matched (manually aligned) cross-year first-level POI categories or belonging to special cross-period second-level matching pairs are semantically smoothed to capture the POI cross-period categorical semantics. The calculation is based on the following Equation (3):

$$\mathcal{L}_{cross-period-alignment} = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}|\boldsymbol{v}_{A_i} - \boldsymbol{v}_{B_j}|_2^2 C_{ij} \tag{3}$$

Where if $\boldsymbol{v}_{A_i}$ and $\boldsymbol{v}_{B_j}$ belong to the manually aligned first-level categories, the value of C_{ij} is 1; otherwise, it is 0. As part of the objective function, we minimize Formula 3 during the training process, then the difference of $\boldsymbol{v}_{A_i}$ and $\boldsymbol{v}_{B_j}$ is getting smaller and smaller, so that we establish a connection between POIs from different temporalities, and make the cross-year POI category pairs with similar semantics close in the embedding space. In a mathematical sense, the embedding space (a manifold) is smoothed according to the cross-year category pairs. To distinguish between the impact of second-level category pairs of within-year POIs and cross-year POIs on the vector embeddings, different weights are assigned to the within-year and cross-year categorical semantics. At this point, all the objective functions of the MT-POI2Vec model for optimizing category embeddings of POIs have been manifested. $\mathcal{L}_{co-occurrence}$ is used to encode spatial co − occurrence information, and $\mathcal{L}_{single-period-smoothing}$ and $\mathcal{L}_{cross-period-alignment}$ represent the within-year categorical semantic objective function constructed with within-year second-level category pairs and the cross-year categorical semantic objective function constructed with cross-year second-level category pairs, respectively, for encoding category semantics. Combining the three objective functions, the spatial co-occurrence relations of POIs are coupled with the categorical semantic information to form the overall objective function:

$$\mathcal{L}_{all} = \mathcal{L}_{co-occurrence} + \lambda_1\mathcal{L}_{single-period-smoothing} + \lambda_2\mathcal{L}_{cross-period-alignment} \tag{4}$$

where λ_1 and λ_2 are the hyper parameters controlling the impact of within-year second-level category pairs and cross-year second-level category pairs on categorical

semantics. Finally, the overall objective function is optimized by the stochastic gradient descent method.

3.3. Evaluating the embedding space alignment

To facilitate the evaluation of the effect of embedding space alignment, we utilize the proxy of whether the paired second-level POI categories are indeed located closely in a unified embedding space. To this end, a quantitative evaluation method, ie proportional accuracy assessment, is used.

Proportional accuracy refers to the proportion of the cross-year POI pairs that belong to the same cluster in the embedding space. Let all embeddings cluster into k classes denoted $C_0, C_1, C_2 \ldots C_{k-1}$, and among all the POI second-level categories, there are s cross-year POI category pairs denoted $\{(P_{A_0}, P_{B_0}), ((P_{A_1}, P_{B_1}) \ldots ((P_{A_{s-1}}, P_{B_{s-1}})\}$. The number of cross-year POI category pairs, whose embeddings from the two periods belong to the same cluster C_i is c_i; then, the proportional accuracy p is calculated as follows:

$$p = \frac{\sum_i^{k-1} c_i}{s} \tag{5}$$

3.4. Accuracy evaluation for LUCD

To manifest the advantages of the proposed MT-POI2Vec model, we compare it our proposed model against several other popular POI embedding methods in a quantitative experiments for LUCD, we conduct supervised training for evaluation using MLP. Specifically, there may be multiple types of land use in each parcel, so the quantitative accuracy verification of LUCD can be considered as a multi-label distribution learning problem (Geng 2016).

Let $\mathcal{F} = \{f_0,\ f_1 \ldots\ f_{t-1}\}$ be the set of t labels representing the land use types of each urban area (land parcel). For each parcel $r_i \in R = \{r_0,\ r_1 \ldots r_{z-1}\}$, let $\boldsymbol{v}_{r_i}$ be the average embedding of all POI embeddings contained in the i-th parcel representing the parcel embedding, and $\boldsymbol{v}_{B_{r_i}} - \boldsymbol{v}_{A_{r_i}}$ is the land use change embedding of the i-th parcel. Let $y_i^{f_j}$ denote the proportion of the j-th land use type for the i-th parcel, where $y_i^{f_j} \in [0,\ 1]$ and $\sum_{j=0}^{t} y_i^{f_j} = 1$. Then the real proportion distribution of the land use change for this parcel in the period of 2013–2020 is $l_{r_i} = [y_{B_i}^{f_0} - y_{A_i}^{f_0},\ y_{B_i}^{f_1} - y_{A_i}^{f_1} \ldots y_{B_i}^{f_{t-1}} - y_{A_i}^{f_{t-1}}]$, where $\sum_{j=0}^{t} y_{B_i}^{f_j} - y_{A_i}^{f_j} = 0$. Then we send $\boldsymbol{v}_{B_{r_i}} - \boldsymbol{v}_{A_{r_i}}$ as feature and l_{r_i} as ground truth label into the MLP for supervised training. Between the input and output layers, there can be one or several hidden layers with nonlinear activation functions. In short, the MLP takes the difference of the embeddings in different years as input, and outputs the changes of each land use type over the years. For training the MLP, we use the L1 distance (Cha 2007) as the loss function to optimize the MLP. The specific calculation method is as follows:

$$\mathcal{L}_{MLP} = \frac{1}{z} \sum\nolimits_{i=1}^{z} \sum_{j=1}^{t} \left| \widehat{y_{B_i - A_i}}^{f_j} - (y_{B_i}^{f_j} - y_{A_i}^{f_j}) \right| \tag{6}$$

Where $(y_{B_i}^{f_j} - y_{A_i}^{f_j})$ is the changed proportion of the ground truth corresponding to the j-th land use type for the i-th parcel. By minimizing the objective function $\mathcal{L}_{MLP}$,

we can learn the relationship between the change embedding $\boldsymbol{v}_{B_{r_i}} - \boldsymbol{v}_{A_{r_i}}$ representing the land use change of the i-th parcel and the corresponding proportion change distribution of the ground truth. Then we also choose the L1 distance as an accuracy evaluation indicator to assess the degree of match between our predicted proportion distribution of land use change and the corresponding ground truth.

4. Results

4.1. Parameter sensitivity analysis

To determine the optimal hyper parameters of the MT-POI2Vec model to control the effect of categorical semantics on embedding, we set $\lambda_1 \in \{10^{-11},\ 10^{-10}, 10^{-9}, 10^{-8}\}$, $\lambda_2 \in \{10^{-8},\ 10^{-7},\ 10^{-6}\}$ for grid search in the MT-POI2Vec model. Parametric sensitivity analysis was performed according to the abovementioned criteria. K-means clustering was first performed on the embeddings, and the number of clusters was set to 40. The clustered POI embeddings were then projected onto the two-dimensional plane using the T-SNE (Van der Maaten and Hinton 2008), as shown in Figure 3 below. The values of λ_1 and λ_2 are annotated in the Figure 3.

When $\lambda_2 \in \{10^{-7}, 10^{-8}\}$, the clustering results show a clear separation between clusters and a certain differentiation within each cluster. Apart from the parameters in this range, other clustering results encountered the problem that the embeddings within the same clusters are too crowded, such as Figure 3(a,d,g,j), which shows a lack of clear differentiation between the POI category embeddings within each cluster. To more intuitively distinguish the two periods of POI data in embedding space in Figure 3(b,e,h,k), the POI embeddings from the two periods are represented by circles and triangles, respectively. In Figure 3(b,e,h,k), it is evident that each cluster contains both circles and triangles, indicating that the MT-POI2Vec model pulls the two periods of POI categories into the same embedding space. The proportional accuracy was further evaluated to determine the best parameters, and the comparison between the proportional accuracy p of the eight groups of candidate parameters is shown in Table 1 below.

In Table 1, we observe that the clustering result of $\lambda_1{=}10^{-10}$, $\lambda_2{=}10^{-7}$ contains the highest percentage of cross-year POI pairs clustered in the same class, which is 99.2% among all cross-year POI pairs, so the optimal parameters of $\lambda_1{=}10^{-10}$, $\lambda_2{=}10^{-7}$ are selected to perform the subsequent analysis.

4.2. Comparison of the effects between different embedding models

4.2.1. Comparison in the embedding space

To further validate the performance of MT-POI2Vec, Word2Vec (Yao *et al.* 2017), Place2Vec (simplified) (Yan *et al.* 2017), and Semantic (Huang *et al.* 2022) models were selected as the baseline models for comparison. For Word2Vec and Place2Vec models, the corpus is constructed separately for the two data periods before training. For the Semantic model, the Delaunay triangle networks are constructed separately for POIs from two periods before random walking, and finally, the categorical semantic embedding is completed, but only the within-year POI categories are paired. Similar to

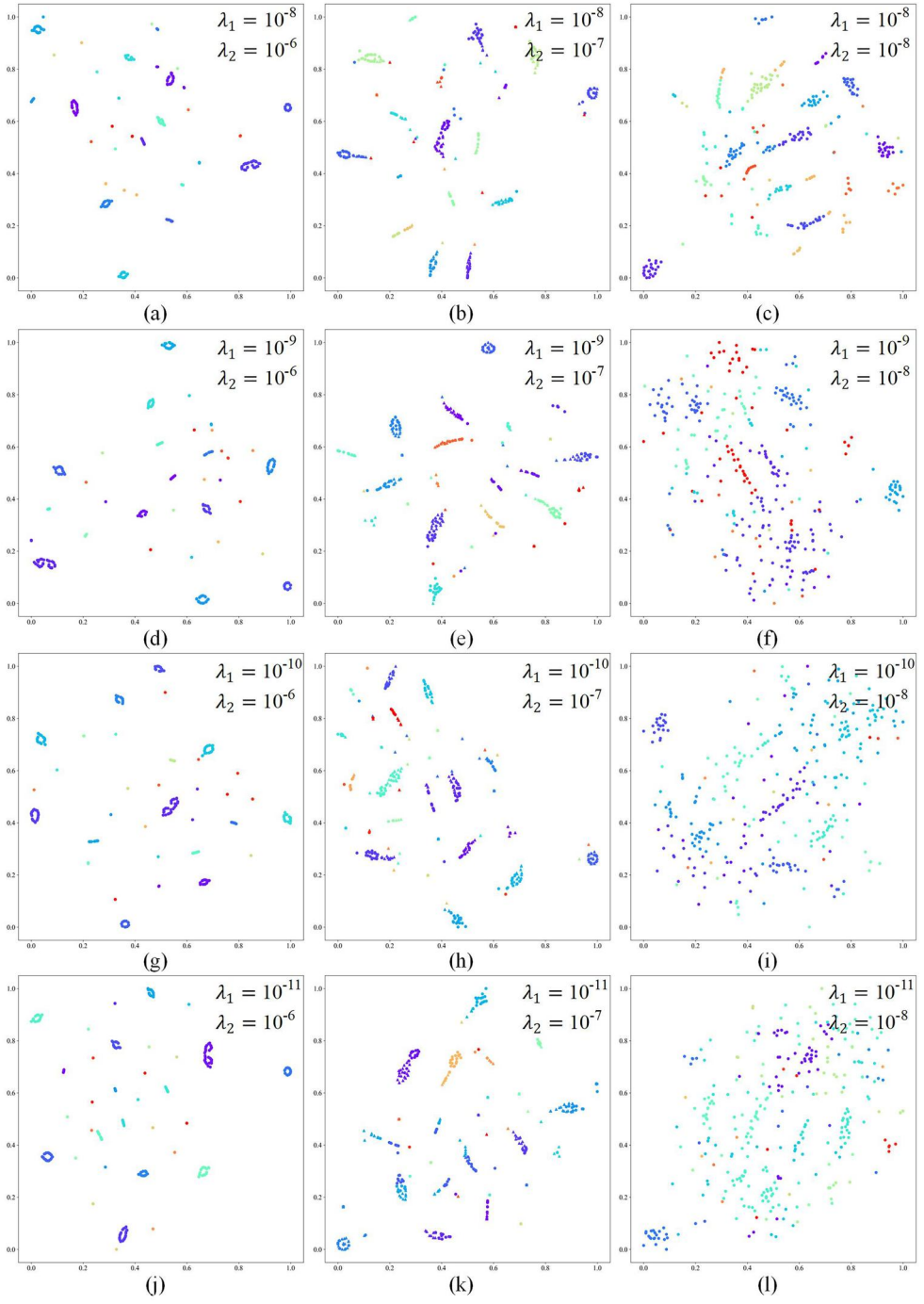

Figure 3. The clustering performance of POI category embeddings in 2D planes for the MT-POI2Vec model based on 12 sets of parameters.

Section 4.1, the embeddings trained by each model were clustered and projected onto a two-dimensional plane. Four sets of cross-year POI category pairs were labeled in the figure, namely, (Art Gallery_A, Art Gallery_B), (Chinese restaurant_A, Chinese

restaurant_B), (Office Tower, Office Building), and (Bank_A, Bank_B), as shown in Figure 4 below.

In Figure 4(a–c), the separate POI embeddings in all four sets of cross-year POI category pairs show a bipolar distribution and do not belong to the same cluster because the POIs in the same cross-year POI category pairs are located in different spatiotemporal spaces. Therefore, the embeddings trained by Word2Vec, Place2Vec, and

Table 1. Comparison of proportional accuracy based on 8 sets of candidate parameters.

λ_1/λ_2	10^{-8}	10^{-9}	10^{-10}	10^{-11}
10^{-7}	0.839	0.991	**0.992**	0.917
10^{-8}	0.386	0.303	0.298	0.266

The best value of p among the 8 sets of candidate parameters is presented in bold.

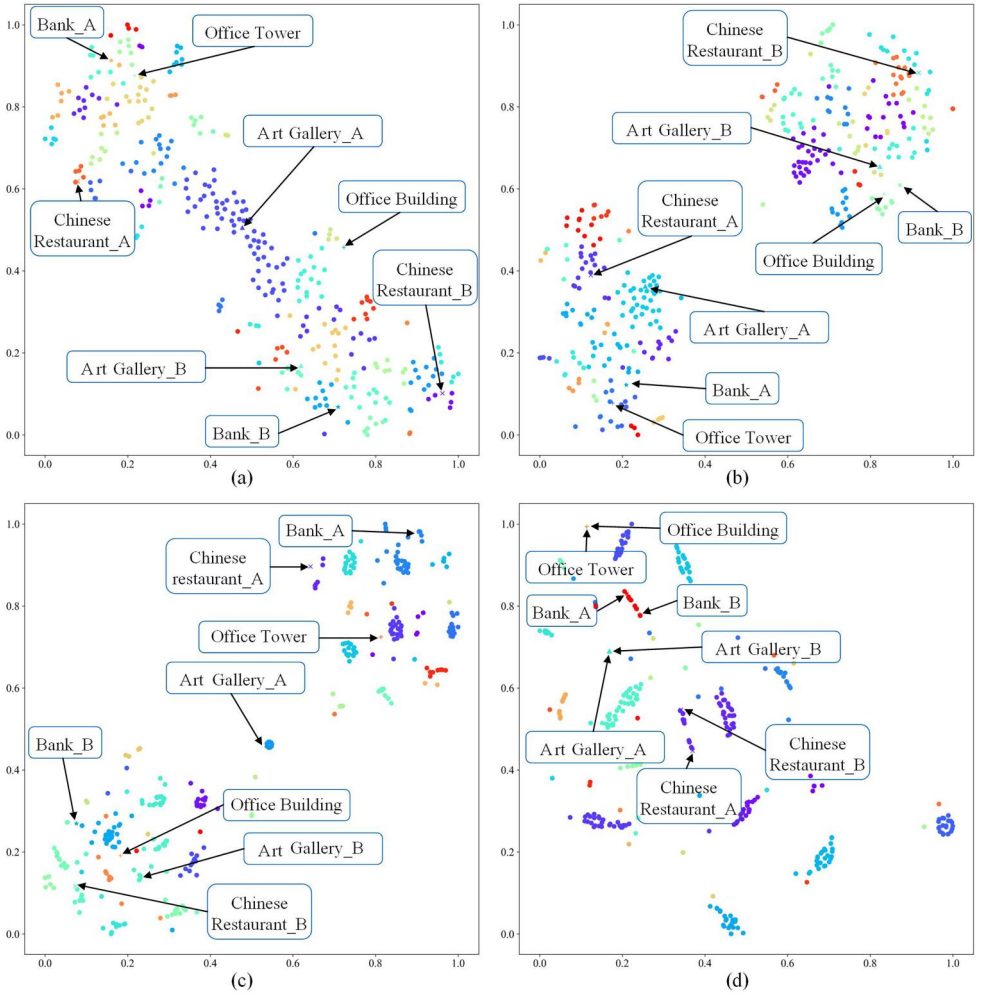

Figure 4. The clustering comparison of POI vector embeddings based on different models. (a) Word2Vec model, (b) Place2Vec model, (c) Semantic model, (d) MT-POI2Vec model.

Semantic models are very far apart when displayed in the same embedding space and do not have any connection. In Figure 4(d), the two vector embeddings of any set of cross-year POI category pairs are very close to each other and belong to the same cluster. This is because the MT-POI2Vec model introduces embedding space alignment, which establishes a link between two POIs of a cross-year POI category pair. In addition, compared with Figure 4(a–c), the separation between clusters and the difference between embeddings in the same cluster are more obvious in Figure 4(d). This indicates that the MT-POI2Vec model can pull the multi-temporal POI vector embeddings under a unified space and increase the co-occurrence probabilities of functionally similar cross-year POI embedding pairs in the same clustering cluster so that the clustering effect can be optimized.

4.2.2. Comparison in the map space

To further analyze the land-use change in Shenzhen, all the POI embeddings of each parcel for each year are first averaged to obtain the embedding expression of each parcel. Then, the cosine distance between the two-period parcel embeddings is calculated to represent the land-use change for each parcel. Finally, the cosine distance is mapped to the map space to obtain the land-use change map representation based on the cosine distance, as shown in Figure 5 below. The lighter color of the parcels represents less variation, and a darker color, more variation. To facilitate further detailed comparison, six typical areas and their remote sensing images were selected for field verification. Due to the limitation in the data, the remote sensing images of 2012 and 2021 were used instead of the images of 2013 and 2020, respectively.

The perception of the overall land use change is not obvious in the Place2Vec model, and the Semantic model is too sensitive to the overall land use change. Only the MT-POI2Vec model can reasonably perceive the land use change. For example, in Figure 5, #1 is located at the airport, which is an important transportation hub and will always perform its transportation function in the long term. #2 is surrounded by the core area of the city government, with a well-developed living support infrastructure, highly concentrated population density, and developed economic conditions. The urbanization at #2 is very mature and will not change much over a long period. The land use types in #1 and #2 did not change from 2012 to 2021. #3 was originally an industrial area, but from 2012 to 2021, some industrial parks were demolished and a new residential area and its supporting commercial facilities were built. During this period, two industrial parks were demolished and a new commercial and residential mixed land with high-end office buildings, businesses and apartments was built in #4. The Place2Vec model can correctly perceive no land-use change in #1 and #2, but does not capture the change in #3 and #4, while the Semantic model shows exactly the opposite result. Only the MT-POI2Vec model can correctly perceive the land use changes in both cases.

Further combined with Figure 4, it is found that compared with the Place2Vec model, the Semantic model can cluster the embeddings of single-period POIs more compactly, but this will result in the multi-period embeddings characterizing the same POI category being farther apart in embedding space. Therefore the cosine distance between them will be larger, which will lead to the Semantic model being too

sensitive to the overall land use change. On the basis of the Semantic model, the MT-POI2Vec model uses the embedding space alignment method to pull the cross-year POI embeddings with similar categorical semantics into the same embedding space, so that the difference between all POI embeddings with similar categorical semantics is smaller. Therefore, the MT-POI2Vec model can perceive land use change more reasonably and accurately.

In addition, compared with the other two models, the MT-POI2Vec model also has significant advantages in reflecting the fineness of land use change. In Figure 5, the cosine distance color of the Place2Vec model, Semantic model and MT-POI2Vec model gradually deepens in #5 and gradually becomes lighter in #6. Combined with remote

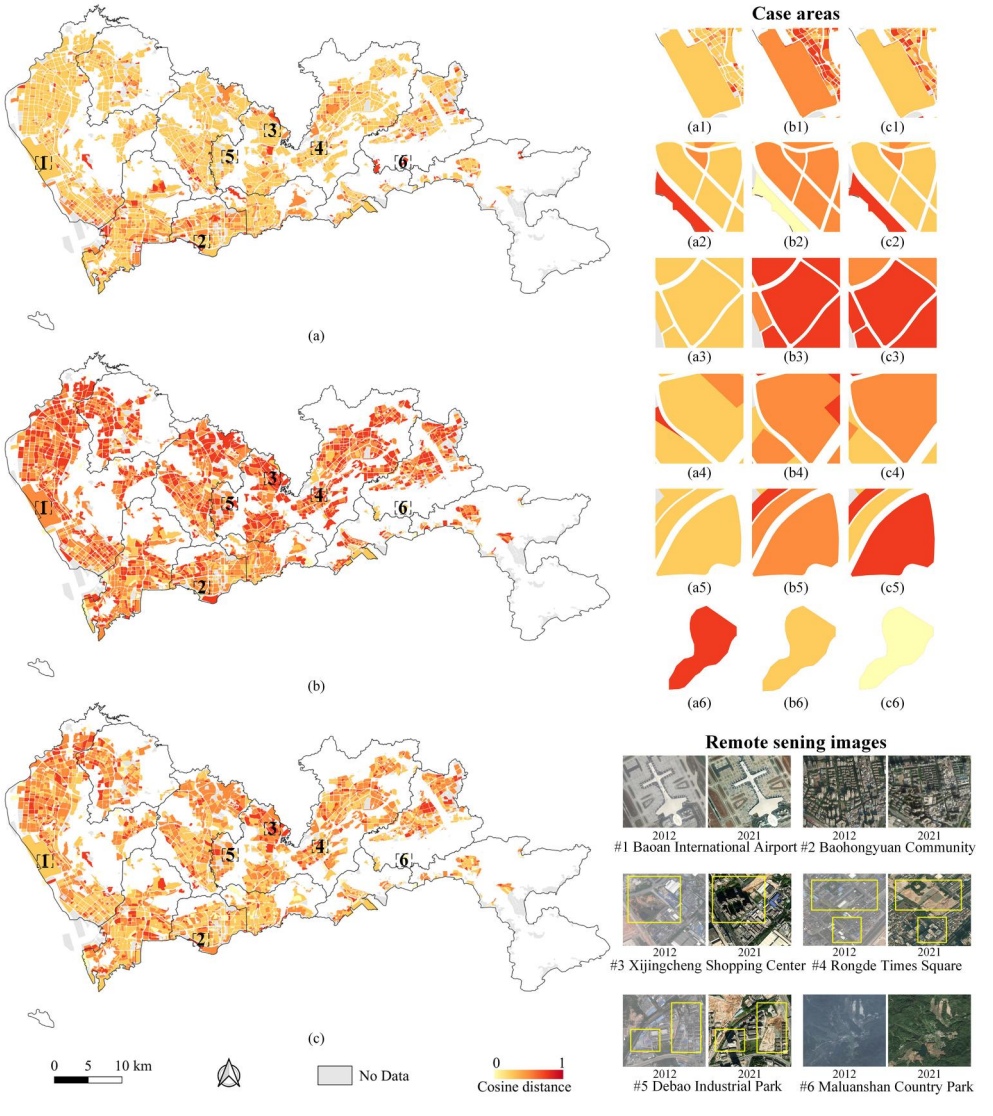

Figure 5. Cosine distance distribution comparison based on two-period parcel embeddings between baseline models and our approach: (a) Place2Vec model; (b) Semantic model; (c) MT-POI2Vec model.

sensing images, it is found that the #5 was originally an industrial land, but between 2012 and 2021, many industrial parks were demolished into bare lands, and a large commercial office building and some apartment houses were built. The proportion of land use types has changed greatly. #6 is a tourist attraction with no land-use change from 2012 to 2021. And the MT-POI2Vec model has the darkest color in #5 and the lightest color in #6, which is consistent with the change of corresponding remote sensing images. Therefore, the MT-POI2Vec model can more accurately perceive the intensity of land use changes.

4.2.3. Quantitative accuracy verification of LUCD

In order to evaluate the accuracy of the model, we need to use parcels that contain both periods of POI data. For this purpose, we selected some parcels in the central urban area of Shenzhen, which are relatively highly built-up, for quantitative validation. The selected area covers all or part of Nanshan District, Futian District, Luohu District and Baoan District of Shenzhen, with a total of 878 parcels. We use the ground truth data from the Bureau of Planning and Natural Resources of Shenzhen Municipality in 2013 and 2020, which both have five land use categories: (1) industrial, (2) residential, (3) commercial, (4) public management and public service, and (5) green space.

In the process of quantitative verification, we split the dataset into a training set (60%), a verification set (20%) and a test set (20%) and the number of training epochs is 100. MLP has two 256-dimensional hidden layer. For reliability, the entire dataset (878 regions) are shuffled 10 times at random to repeat the above-mentioned training, validation and testing processes. We evaluated our models and baseline models by using the test set. The average L1 distance and its standard deviation between the proportion change distribution of land use types predicted by each model and the corresponding proportion change distribution of the ground truth are shown in Table 2 below.

It can be seen that the average L1 distance of the proposed MT-POI2Vec model is the smallest, indicating that our model has the smallest average estimation error for LUCD, which is closer to the changed ground truth and has the best performance. This is because our proposed MT-POI2Vec model pulls POIs with similar semantics in different years into the same embedding space, making the operation more accurate.

4.3. Category analysis of land use change using clustering method

To further analyze the transformation relations of land-use change types for each parcel, the clustering results of the land-use change embeddings are mapped to the Shenzhen city map, and K represents the number of clusters. Figure 6 shows the

Table 2. Comparison of average L1 distance and standard deviation based on the baseline and our proposed model.

Model	L1	Standard deviation
Word2Vec	0.502	0.021
Place2vec	0.495	0.030
Semantic	0.485	0.031
MT-POI2Vec	0.448	0.019

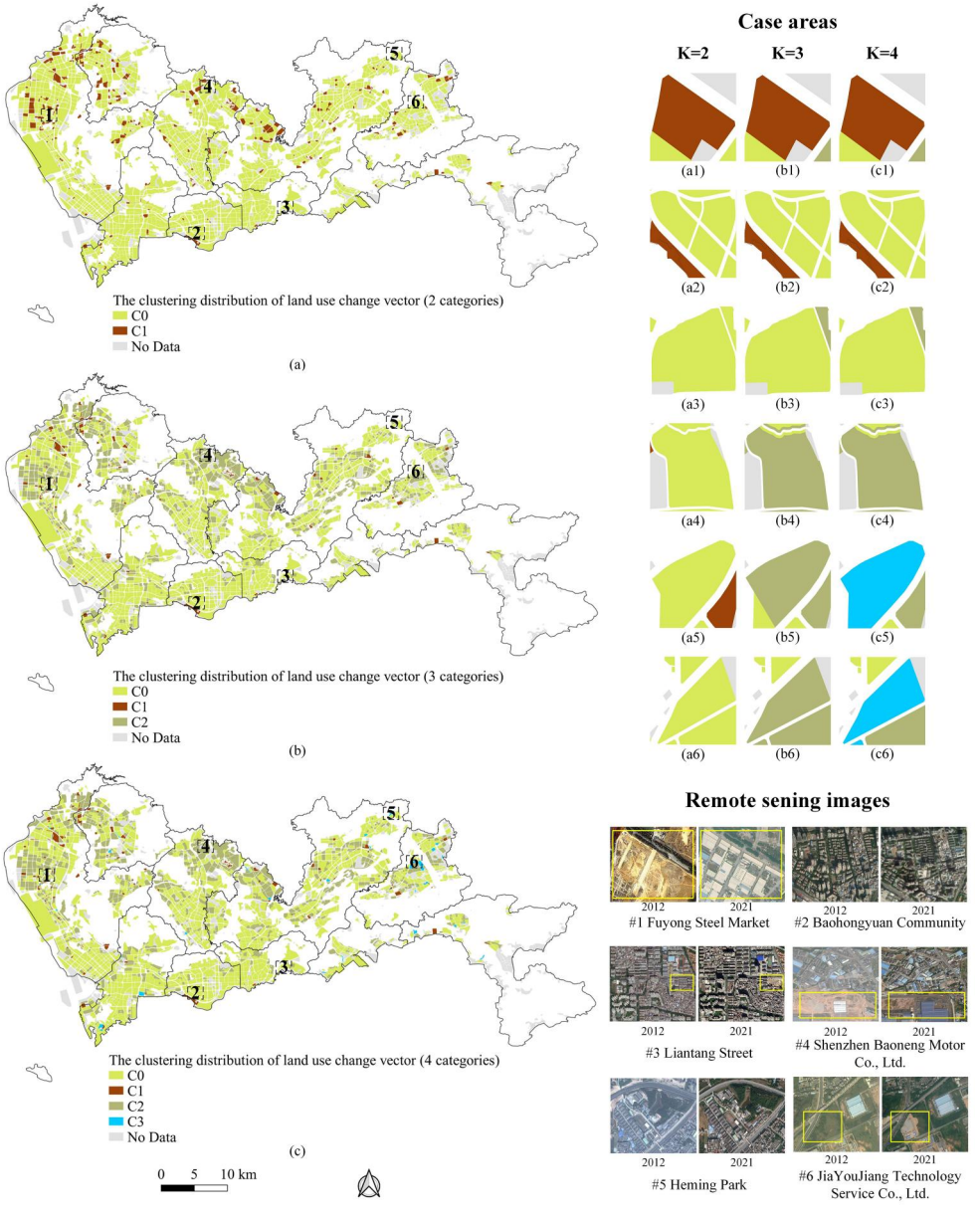

Figure 6. Clustering distribution comparison of the land use change vector embeddings under different clustering categories based on the MT-POI2Vec model: (a) 2 categories; (b) 3 categories; (c) 4 categories.

mapping distributions with the number of clusters from 2 to 4, where C0 and C1 denote two kinds of clusters when the number of clusters is 2, and the rest of the clusters are named similarly. To facilitate further detailed comparison and validation, similar to the analysis in Section 4.2.2, six typical regions and their remote sensing images are also selected for field validation, where #2 is the same as region #2 in Figure 5.

$K = 2$:

Areas with relatively smaller land use change [$K = 2$, C0] and areas with relatively larger land use change [$K = 2$, C1]: Compared with Figure 5(c), it is found the cluster region [$K = 2$, C0] basically overlaps with the light-colored parcels in the cosine distance distribution map based on two-period parcel embeddings, and the cosine distance in this region is smaller than the cluster region [$K = 2$, C1]. The cluster region [$K = 2$, C1] is the opposite. Therefore, it can be considered that the cluster region [$K = 2$, C0] represents the areas with relatively smaller land use change, and the cluster region [$K = 2$, C1] represents the areas with relatively larger land use change. For example, combined with remote sensing images, it can be found that #2 is residential land with no land-use change, while #1 is completely changed from bare land to commercial land.

$K = 3$ and $K = 4$:

Areas dominated by land use change in residential areas [$K = 3$, C0] and [$K = 4$, C0], and areas dominated by land use change in industrial areas [$K = 3$, C2] and [$K = 4$, C2]: Compared with the results of $K = 2$, this study can find these two clusters are mainly stripped from the cluster region [$K = 2$, C0] with relatively small land use changes. We find the residential areas of cluster regions [$K = 3$, C0] and [$K = 4$, C0] account for the highest proportion before and after land use change. Further combined with remote sensing images, it is found this cluster is mainly residential areas with no land-use change or with certain land use changes but no change in regional functions, such as #2 and #3. #3 is an unchanged residential area from 2012 to 2021, but the eastern part of the block is demolished to become bare land. Similarly, we can find the industrial areas of cluster regions [$K = 3$, C2] and [$K = 4$, C2] account for the highest proportion before and after land use change. This cluster is mainly industrial areas with no land-use change or certain land use change but no change in regional functions, such as #4, #5 and #6. #4 is a dense industrial area with no change from 2012 to 2021, and only a large industrial park has been built on the bare land in the south of the parcel. Both #5 and #6 are industrial areas with unchanged regional functions, and part of the green space in #6 is changed into an industrial park.

Areas with large land use change [$K = 3$, C1] and [$K = 4$, C1]: Compared with the results of $K = 2$, we can find the distribution of parcels in this cluster region is basically consistent with the distribution of [$K = 2$, C1], and the land use change in this cluster region is large. For example, the regional functions in #1 have completely changed.

Areas containing mainly land use changes in green space and industrial areas [$K = 4$, C3]: Compared with the results of $K = 3$, this study can find this cluster is mainly stripped from the cluster region [$K = 3$, C2]. We find the land use change in this cluster mainly occurred in industrial areas with a high proportion of green space, such as #5 and #6. A large area of park green space is included in #5, and part of the green space in #6 is changed into an industrial park, which contains a higher proportion of green space.

5. Discussion

Utilizing POI embeddings from multiple temporalities to accomplish unsupervised LUCD is useful in the case of insufficient ground truth. However, the separately

learned POI embeddings in different years will end up in isolated embeddings spaces, making the unsupervised land use change detection infeasible. To solve these problems, this study proposes the MT-POI2Vec embedding model, which maps POIs to the real-valued embeddings in a unified low-dimensional mathematical space. We propose the MT-POI2Vec model to construct a framework for identifying land-use changes by fully unsupervised methods and explore in depth the effectiveness and capability of representing land-use changes using POI embeddings.

The MT-POI2Vec model is the first model that uses POI embeddings for unsupervised LUCD, highlighting the temporal heterogeneity of land use change, effectively solving the problem that two periods of POI embeddings are not in the same space-time, resulting in the inability to interact directly. We introduce embedding space alignment based on the Semantic model proposed by Huang *et al.* (2022). The POIs from two periods are first manually aligned, ie POIs with similar categorical semantics are matched into cross-year second-level category pairs. Then, we combine the cross-year second-level category pairs with the co-occurrence relations of POIs to achieve the vector embedding. In the mathematical space, the MT-POI2Vec model can pull multi-temporal POI vector embeddings under a unified space and establish connections and increase the co-occurrence of functionally similar cross-year POI embedding pairs in the same cluster to make the clustering effect optimal. In addition, a quantitative experiment shows that the proposed MT-POI2Vec model has the highest accuracy in identifying land use change. This is because our method can gather cross-year POI category pairs closer together in the embedding space to improve the accuracy of land-use change identification in downstream tasks.

By taking the cosine of the multi-temporal parcel embeddings and projecting it to map space, we find land-use changes are greater in areas more closely associated with human activities, such as industrial and residential areas of megacities. In comparison, land-use changes are smaller in areas with public service facilities, such as airports. The MT-POI2Vec model has a more advanced effect than the baseline models in representing land-use changes and can more finely identify areas with unchanged land use during urbanization and land-use changes in residential and industrial areas. In addition, by clustering the parcel change embeddings and mapping them to the map space, the land-use classification is refined step by step and the accuracy of identification gradually improves as the number of clusters changes from 2 to 4. The results of the cluster analysis reflect the ability of the MT-POI2Vec embedding method to represent the land-use change in a hierarchical way. Our method directly uses parcel-scale change embeddings for LUCD and skillfully avoids gradually accumulating errors in the classification. In contrast, the traditional methods of first-classify-then-detect can only rigidly identify the transformation from one land-use type to another land-use type and cannot discover the hierarchical characteristics of land-use change. Moreover, in this case, the final change identification results are also subject to the influence of classification accuracy (Tewabe and Fentahun 2020). In addition, the discovered hierarchical characteristics of land-use change patterns can guide the exploration of subsequent researches on the drivers of land-use changes.

However, there are still some limitations in this paper. First, the limitations of the single-source POI data could add some errors to the study results. Due to the various

categories and large number of POIs, as well as differences in statistical methods and time, even as the largest map service provider in China, the POIs collected by Baidu Maps will inevitably have missing and inconsistent information on scale and rank. Moreover, the sparsity of data in some areas can affect the results of LUCD. Subsequent studies can consider combining POI data from multiple sources, such as Gaode POI, to supplement the missing POI information with each other to form a higher-quality POI dataset. Second, the MT-POI2Vec model is greatly affected by the manual alignment of cross-year POI categories. Specifically, the names of first-level categories between multi-period POI data with a large time span are not exactly the same. In the case where the first-level categories cannot be directly matched, it may also need to be judged in combination with specific second-level categories. Therefore, the manual process of aligning categories could introduce errors to some extent, resulting in the inaccurate matching of certain cross-year POI category pairs, which in turn affects the accuracy of the model. Therefore, high-quality POI data with more consistent or similar categories can reduce the impact of manual alignment on the model. Furthermore, future research can consider developing a program or neural network that can match specific POI names for automatic and accurate category matching, which could likely improve the accuracy of matching and increase efficiency. Finally, this study uses the same parcel data as the basic unit to group POIs and analyze land use changes in both time periods, while the parcel boundaries in different periods may change. In this case, the parcels need to be divided into smaller units to unify the analysis units. However, too small a scale will lead to the absence of POI data in some units, thus making the LUCD infeasible. Therefore, in future research, what is the optimal analytical scale for LUCD using our approach is also worth exploring.

6. Conclusion

In this paper, we propose the MT-POI2Vec model to identify land use change in an unsupervised manner, which achieves embedding space alignment and pulls multi-period spatiotemporally non-cooccurring POI vector embeddings into the unified embedding space. Compared with the baseline models such as Word2Vec, Place2Vec and Semantic, the MT-POI2Vec model can increase the co-occurrence probabilities of functionally similar cross-year POI embedding pairs in the same cluster with the best performance in quantitative accuracy verification of LUCD. The results of the megacity study show that land-use types have greater variability in areas that are more closely associated with human activities, such as industrial and residential areas. In contrast, land-use change in areas with public service facilities such as airports is smaller due to the continuous urbanization process in recent years. In addition, by comparing the cluster mapping distribution of the land-use change vector embeddings under three kinds of cluster numbers, we find the MT-POI2Vec model can effectively reflect the hierarchical nature of land-use changes as the number of clusters increases. In future studies, the hierarchical characteristics of land-use change can be used to explore the potential drivers of land-use change more deeply. This study can provide a reference for the rapid urban land-use change monitoring.

Acknowledgements

We would like to acknowledge the comments and insights from the editors and three anonymous reviewers that helped lift the quality of the article.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Key Research and Development Program of China [2019YFB2102903], the National Natural Science Foundation of China [41801306, 42101421 and 42171466]; the "CUG Scholar" Scientific Research Funds at China University of Geosciences (Wuhan) [2022034], a grant from Alibaba Innovative Research Project [20228670], a Guangdong-Hong Kong-Macau Joint Laboratory Program [2020B1212030009], and a grant from State Key Laboratory of Resources and Environmental Information System. W.H. acknowledges the financial support from the Knut and Alice Wallenberg Foundation.

Notes on contributors

Yao Yao is a professor at China University of Geosciences (Wuhan), a researcher from the Center for Spatial Information Science at the University of Tokyo, and a visiting scholar at Alibaba Group. His research interests are geospatial big data mining, analysis, and computational urban science.

Qia Zhu is a graduate student at China University of Geosciences (Wuhan). His research interests are spatial representation learning and urban land use change detection.

Zijin Guo is a graduate student at China University of Geosciences (Wuhan). His research interests are trajectory data mining and complex network analysis.

Weiming Huang received his PhD in Geographical Information Science at Lund University, Sweden in 2020. He is a Wallenberg-NTU Postdoctoral Fellow at Nanyang Technological University, Singapore. His research interests mainly include spatial data mining and geospatial knowledge graphs.

Yatao Zhang is a doctoral student at the Mobility Information Engineering lab at ETH Zurich and the Future Resilient Systems at the Singapore-ETH centre. His research interests lie in context-based spatiotemporal analysis, geospatial big data mining, and traffic forecasting.

Xiaoqin Yan is currently a Ph.D. student in GIScience at the Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing. His research interests are spatiotemporal big data computing and social perception.

Anning Dong is a graduate student at China University of Geosciences (Wuhan). His research interests are spatiotemporal big data mining and crime geography.

Zhangwei Jiang is a staff algorithm engineer at Alibaba Group. His research interests are LBS data mining and research&recommendation algorithm.

Hong Liu is a senior staff algorithm engineer at Alibaba Group. His research interests are data mining and research&recommendation algorithm.

Qingfeng Guan is a professor at China University of Geosciences (Wuhan). His research interests are high-performance spatial intelligence computation and urban computing.

ORCID

Yao Yao http://orcid.org/0000-0002-2830-0377
Qia Zhu http://orcid.org/0000-0001-6580-2213
Zijin Guo http://orcid.org/0009-0005-4121-4341
Weiming Huang http://orcid.org/0000-0002-3208-4208
Yatao Zhang http://orcid.org/0000-0001-5701-2836
Xiaoqin Yan http://orcid.org/0009-0000-3664-1465
Anning Dong http://orcid.org/0009-0002-4160-5148
Qingfeng Guan http://orcid.org/0000-0002-7392-3709

Data and codes availability statement

We share the codes and the sub-sampled data of the study at https://doi.org/10.6084/m9.figshare.24081699.

References

Aderele, M.O., Bola, T.S., and Oke, D.O., 2020. Land use/land cover changes of Ago-Owu Forest Reserve, Osun State, Nigeria using remote sensing techniques. *Open Journal of Forestry*, 10 (04), 401–411.

Alhassan, V., *et al.*, 2020. A deep learning framework for land-use/land-cover mapping and analysis using multispectral satellite imagery. *Neural Computing and Applications*, 32 (12), 8529–8544.

Asokan, A., and Anitha, J., 2019. Change detection techniques for remote sensing applications: a survey. *Earth Science Informatics*, 12 (2), 143–160.

Bai, L., *et al.*, 2023. Geographic mapping with unsupervised multi-modal representation learning from VHR images and POIs. *ISPRS Journal of Photogrammetry and Remote Sensing*, 201, 193–208.

Belkin, M., and Niyogi, P., 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. *NIPS'01: proceedings of the 14th international conference on neural information processing systems: natural and synthetic*, 585–591. ACM.

Cao, R., *et al.*, 2020. Deep learning-based remote and social sensing data fusion for urban region function recognition. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163, 82–97.

Cha, S.-H., 2007. Comprehensive survey on distance/similarity measures between probability density functions. *City*, 1 (2), 1.

Feizizadeh, B., *et al.*, 2021. A comparison of the integrated fuzzy object-based deep learning approach and three machine learning techniques for land use/cover change monitoring and environmental impacts assessment. *GIScience & Remote Sensing*, 58 (8), 1543–1570.

Geng, X., 2016. Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 28 (7), 1734–1748.

Guo, S., *et al.*, 2015. Semantically smooth knowledge graph embedding. ed. *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (Volume 1: Long Papers)*, 84–94.

Halimi, M., Sedighifar, Z., and Mohammadi, C., 2018. Analyzing spatiotemporal land use/cover dynamic using remote sensing imagery and GIS techniques case: Kan basin of Iran. *GeoJournal*, 83 (5), 1067–1077.

Henry, C.J., *et al.*, 2019. Automated LULC map production using deep neural networks. *International Journal of Remote Sensing*, 40 (11), 4416–4440.

Hu, S., *et al.*, 2021. Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. *Computers, Environment and Urban Systems*, 87, 101619.

Huang, W., *et al.*, 2022. Estimating urban functional distributions with semantics preserved POI embedding. *International Journal of Geographical Information Science*, 36 (10), 1905–1930.

Huang, W., *et al.*, 2023. Learning urban region representations with POIs and hierarchical graph infomax. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 134–145.

Islam, K., *et al.*, 2018. Land use classification and change detection by using multi-temporal remotely sensed imagery: The case of Chunati wildlife sanctuary, Bangladesh. *The Egyptian Journal of Remote Sensing and Space Science*, 21 (1), 37–47.

Jiang, S., *et al.*, 2015. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, 53, 36–46.

Jin, S., *et al.*, 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sensing of Environment*, 132, 159–175.

Jung, M., and Chang, E., 2015. NDVI-based land-cover change detection using harmonic analysis. *International Journal of Remote Sensing*, 36 (4), 1097–1113.

Mikolov, T., *et al.*, 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Nandam, V., and Patel, P., 2022. A novel hybrid approach using SVM and spectral indices for enhanced land use land cover mapping of coastal urban plains. *Geocarto International*, 37 (16), 4714–4736.

Pei, T., *et al.*, 2014. A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science*, 28 (9), 1988–2007.

Prabu, P., and Dar, M.A., 2018. Land-use/cover change in Coimbatore urban area (Tamil Nadu, India)—a remote sensing and GIS-based study. *Environmental Monitoring and Assessment*, 190 (8), 445.

Pramanik, S., and Punia, M., 2020. Land use/land cover change and surface urban heat island intensity: source–sink landscape-based study in Delhi, India. *Environment, Development and Sustainability*, 22 (8), 7331–7356.

Samanta, S., and Paul, S.K., 2016. Geospatial analysis of shoreline and land use/land cover changes through remote sensing and GIS techniques. *Modeling Earth Systems and Environment*, 2 (3), 1–8.

Singh, S., *et al.*, 2021. Qualitative and quantitative analysis of topographically derived CVA algorithms using MODIS and Landsat-8 data over Western Himalayas, India. *Quaternary International*, 575-576, 85–95.

Tewabe, D., and Fentahun, T., 2020. Assessing land use and land cover change detection using remote sensing in the Lake Tana Basin, Northwest Ethiopia. *Cogent Environmental Science*, 6 (1), 1778998.

Tian, L., and Shen, T., 2011. Evaluation of plan implementation in the transitional China: A case of Guangzhou city master plan. *Cities*, 28 (1), 11–27.

Van der Maaten, L., and Hinton, G., 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9 (11), 2579–2605.

Viana, C.M., *et al.*, 2019. Land use/land cover change detection and urban sprawl analysis. *In*: H.R. Pourghasemi and C. Gokceoglu, eds. *Spatial modeling in GIS and R for earth and environmental sciences*. Amsterdam: Elsevier, 621–651.

Wu, H., *et al.*, 2021. Identifying core driving factors of urban land use change from global land cover products and POI data using the random forest method. *International Journal of Applied Earth Observation and Geoinformation*, 103, 102475.

Xiaolu, S., and Bo, C., 2011. Change detection using change vector analysis from Landsat TM images in Wuhan. *Procedia Environmental Sciences*, 11, 238–244.

Xing, H., and Meng, Y., 2018. Integrating landscape metrics and socioeconomic features for urban functional region classification. *Computers, Environment and Urban Systems*, 72, 134–145.

Yan, B., *et al.*, 2017. From itdl to place2vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. *Proceedings of the 25th ACM SIGSPATIAL international conference on advances in geographic information systems*, 1–10. ACM.

Yan, X., *et al.*, 2019. A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150, 259–273.

Yao, Y., *et al.*, 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science*, 31 (4), 825–848.

Yao, Y., *et al.*, 2020. Delineating mixed urban "jobs-housing" patterns at a fine scale by using high spatial resolution remote-sensing imagery. *Complexity*, 2020, 1–13.

Yu, W., *et al.*, 2016. A new approach for land cover classification and change analysis: Integrating backdating and an object-based method. *Remote Sensing of Environment*, 177, 37–47.

Zewdie, W., and Csaplovies, E., 2015. Remote sensing based multi-temporal land cover classification and change detection in northwestern Ethiopia. *European Journal of Remote Sensing*, 48 (1), 121–139.

Zhai, W., *et al.*, 2019. Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Computers, Environment and Urban Systems*, 74, 1–12.

Zhang, J., *et al.*, 2021. The Traj2Vec model to quantify residents' spatial trajectories and estimate the proportions of urban land-use types. *International Journal of Geographical Information Science*, 35 (1), 193–211.

Zhang, Y., *et al.*, 2019. Functional urban land use recognition integrating multi-source geospatial data and cross-correlations. *Computers, Environment and Urban Systems*, 78, 101374.

Zhu, Q., *et al.*, 2022. Land-use/land-cover change detection based on a Siamese global learning framework for high spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184, 63–78.