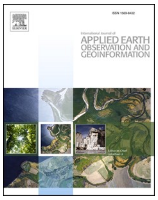
Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



Learning dual context aware POI representations for geographic mapping☆

Quan Qin[a], Tinghua Ai[a,*], Shishuo Xu[b], Yan Zhang[c], Weiming Huang[d], Mingyi Du[b], Songnian Li[e]

[a] *School of Resource and Environmental Sciences, Wuhan University, Wuhan, China*
[b] *School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing, China*
[c] *Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, Hong Kong, China*
[d] *College of Computing and Data Science, Nanyang Technological University, Singapore*
[e] *Department of Civil Engineering, Toronto Metropolitan University, Toronto, Canada*

ARTICLE INFO

Keywords:
Geospatial representation learning
Graph neural network
Geographic mapping
Point of interest
POI embedding

ABSTRACT

Driven by artificial intelligence technologies, geospatial representation learning has become a new trend to better understand urban systems. Points of Interest (POI), as the current mainstream data in urban studies, plays an important role in these methods to discover urban characteristics. Existing studies on POI representation learning focus on spatial and type information, but overlook heterogeneous semantic interaction between POIs as well as hierarchical associations among types. To tackle these two problems, we propose a novel approach, called POI Dual Context Aware Neural Network (DCA) for learning POI representations by jointly embedding both spatial context and type context. For the spatial context of POIs, we introduce a distance decay effect constrained graph attention network as an encoder of DCA, which takes the heterogeneous semantic interaction and spatial proximity of POIs into account. For the type context of POIs, we propose a type hierarchical aggregation neural network architecture for DCA, and design a type infomax optimization objective following contrastive learning mechanism. The superiority of DCA is demonstrated in three geographic mapping tasks, including urban function mapping, region popularity mapping, and housing price mapping. This study provides a new insight to mine deep information from POIs, contributing to a better understanding of urban systems. The source code is released at http://github.com/quan-qin/DCA.

1. Introduction

In recent years, there has been a concerted effort to reveal urban spatial and semantic characteristics to understand complex urban systems. Amid these efforts, geospatial representation learning has emerged as a new way of understanding urban systems from the perspective of GeoAI (Chen et al., 2025; Janowicz et al., 2020; Mai et al., 2024). Geospatial representation learning employs low-dimensional, compact, and informative embeddings to represent the spatial and semantic information of geospatial features, aiming to better characterize urban physical and social spaces and aid in the understanding of urban systems. In practice, geospatial representation learning has informed decision making or catalyzed novel downstream analytics (Liu & Biljecki, 2022; Wang & Biljecki, 2022; Xiao et al., 2024), and helped solve geographic mapping tasks within cities, e.g., urban function, crime prediction, traffic prediction, and housing price mapping (Huang et al., 2023; Zhang et al., 2023c).

Against the backdrop of ubiquitous urban big data, numerous studies have leveraged GeoAI techniques to provide comprehensive insights into urban systems by learning geospatial representations, e.g., trajectories (Zhang et al., 2024b), building footprints (Yan et al., 2019), street views (Cao et al., 2025), and points of interest (POIs) (Huang et al., 2022; Wang et al., 2024). Among them, POIs refer to any geographic entity that can be of interest to people, e.g., schools, parks, and banks. Being a mainstream data source in current urban studies (Andrade et al., 2020), POI data contributes to a better understanding of the intricate interlinkages between people and places (Psyllidis et al., 2022). POI data is commonly utilized both independently for urban studies and as an effective supplementary dataset to complement other data sources. This can be attributed to the meaningful urban representations that can be, to some extent, delineated by the information entailed from POIs (Huang et al., 2023). Consequently, there is a pressing need for a more effective

☆ This article is part of a special issue entitled: ‘Foundation models EO’ published in International Journal of Applied Earth Observation and Geoinformation.

* Corresponding author.

E-mail address: tinghuaai@whu.edu.cn (T. Ai).

https://doi.org/10.1016/j.jag.2025.104683
Received 4 March 2025; Received in revised form 24 May 2025; Accepted 15 June 2025
Available online 20 June 2025

representation learning approach specifically focus on POIs. Beyond feature engineering and traditional rule-based methods, e.g., using frequency density or kernel density of POI types (Su et al., 2021), POI representation learning can effectively mine potential spatial and semantic information of POIs and express them with more meaningful distributed embedding vectors, and further contributes to computing and reasoning spatial phenomena and principles.

Learning on POI representations tries to capture the spatial distribution pattern or semantic association characteristics to derive POI representations. The previous studies on POI representations focused on the spatial distribution characteristics of POIs, and usually constructed POI type sequences according to the spatial structure of POIs, and combined with natural language processing technologies to model the textual context of type sequence, i.e., spatial context of POIs. For instance, Yao et al. (2017) introduced Word2Vec model to learn the representation of POI types, which captures the spatial co-occurrence pattern of POI type sequences constructed by greedy algorithm. Niu and Silva (2021) used K-nearest-neighbors method to sample POI sequences and combined with Doc2Vec model to learn the representation of POI types, and Qin et al. (2022a) enhanced the spatial information in POI representations by capturing multi-spatial distribution patterns of POIs. However, the sequential spatial relationship modeling method has inherent limitations, that is, the complex spatial context of POIs does not strictly adhere to the 1-D linear structure as the context in natural language, leading to the loss of a significant amount of POI spatial context information. As an improvement, the construction of 1-D linear structure spatial context can be extended to 2-D considering multi-connection in the planar context, which is consistent with the actual spatial relationship of POIs. Consequently, a graph-based method for modeling POI spatial relationships has emerged. Xu et al. (2022) modeled the POI spatial relationship into a graph based on Delaunay Triangulation (DT) network, and introduced graph convolutional network (GCN) to encode POI graph. The above studies rely on the inherent encoding capacity of deep learning models to learn spatial features. Some studies have further attempted to capture more abstract and higher-level spatial features tied to POIs. In practice, Bai et al. (2023) captured spatial dependencies by contrasting POIs with remote sensing imageries. Zhao et al. (2023) performed soft graph editing on POI graphs followed by graph similarity contrastive learning to capture region-level spatial distribution patterns of POIs. Li et al. (2023) employed contrastive learning to capture the spatial proximity of POIs at regional scale. Huang et al. (2023) proposed HGI to capture the hierarchical spatial relationships among POI-region-city, leading to more comprehensive spatial information.

Distinct from the above that focus on spatial information of POIs, some studies emphasize type information for a more nuanced characterization of POI representations. Huang et al. (2022) employed the Laplacian eigenmaps (LE) algorithm to preserve hierarchical type information by reducing the semantic distance between POI type embeddings on first- and second-level type, and Yao et al. (2023) similarly applied LE algorithm to enhance the POI type embeddings by aligning multi-temporal semantics. Some other studies have explored combining POI type information with the powerful capabilities of large language models (LLMs) to derive POI embeddings. For instance, pre-trained BERT models were employed to generate embeddings for POI types (Zhang et al., 2021) and POI names (Zhang et al., 2023a). While using pre-trained LLM can conveniently provide embeddings for POI representations, the covariate shift issue unavoidably results in a semantic gap between geographic semantics and natural language semantics.

Recent studies on POI representation learning endeavor to derive POI embeddings based on type and spatial information, but they face two major challenges. (1) Regarding the spatial information, existing studies typically employ graph convolution-based message passing for POIs, assuming uniform importance across POIs within the spatial context, while overlooking the heterogeneous semantic influences among POIs. (2) Regarding the type information, most existing studies focuses solely on a fixed type level, always using second-level type information for POI representations due to the scarcity of information on first-level and the relative superfluous nature of third-level type information (Hu et al., 2020; Xu et al., 2022). Consequently, valuable information from POI internal type hierarchies is lost, which serves as a primary motivation for this study.

Given the shortcomings of the existing studies, in this study, we focus on POI representation learning and propose a novel POI representation learning approach called POI Dual Context Aware Neural Network (DCA) to tackle the aforementioned challenges. DCA employs the Graph Attention Network (GAT) (Veličković et al., 2018) to sense the spatial structure of POIs (i.e., spatial context) and simultaneously leverages a contrastive learning mechanism to capture the type hierarchy of POIs (i. e., the type context defined in this study). DCA regards the POI type context as self-supervised signals to guide the network to delicately sculpt on the POI embedding shaped by spatial context information. Through this design, DCA sophisticatedly shapes informative and discriminative POI representation embeddings by jointly embedding both spatial context and type context. The proposed DCA can be further extended to various downstream geographic mapping tasks within cities. Overall, the contributions of this work are two-fold:

- For the spatial context, we introduce a distance decay effect constrained GAT, which takes the heterogeneous semantic interaction and spatial proximity of POIs into account.
- For the type context, we propose a type hierarchical aggregation neural network architecture and design a type infomax optimization objective following contrastive learning mechanism.

The remainder of this paper is organized as follows. Section 2 elaborates on the overall architecture and design details of DCA, and introduces its application in a typical geographic task. Section 3 describes the study area and data, as well as experimental results and relevant analyses. In Section 4, we conduct additional exploratory analysis on POIs, and discuss both the limitations and potential of the DCA approach. Finally, the paper ends with conclusion and future work in Section 5.

2. Methodology

2.1. Overview

This study focuses on developing a general-purpose POI representation learning model toward geographic mapping. Fig. 1 presents the framework overview. For the first stage, we propose a POI representation learning model built upon the theories of contrastive learning and graph theory, which shapes the POI representations by spatial context and type context. The second stage is to utilize POI representations for the further three geographic mapping tasks, namely urban function mapping, housing price mapping, and region popularity mapping.

2.2. POI representation learning with DCA

The intuition behind DCA is to jointly embed the external spatial structure and internal type hierarchy of POIs, enabling the awareness of spatial and type contexts for POI representations. An overview of DCA is shown in the Fig. 2. The architecture consists of three main modules: (1) a shared graph encoder, implemented as a distance decay-constrained GAT, which senses spatial context by encoding a spatial context preserved POI graph; (2) a parameter-free type aggregator, which preserves type context by gradually aggregating POI representations along the type hierarchy; (3) a contrastive learning module, which senses type context by maximizing the mutual information (MI) between the POI embeddings across type hierarchies. The encoder and aggregator together form the type hierarchical aggregation neural network responsible for forward propagation, while the contrastive learning module serves to guide backpropagation. In this way, POI embeddings

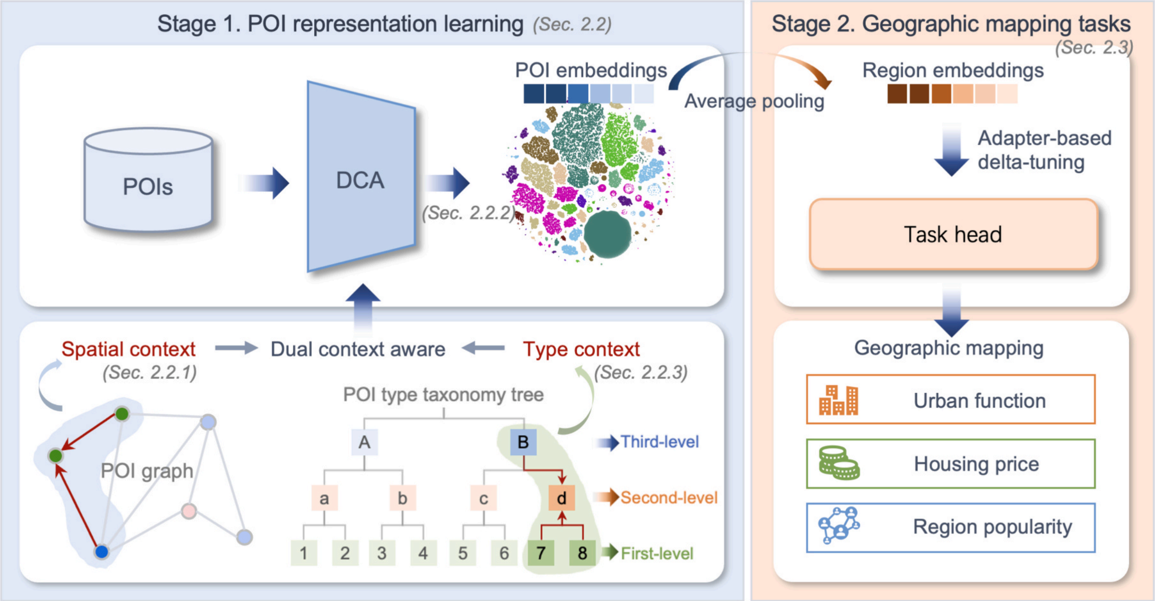


Fig. 1. Framework overview.

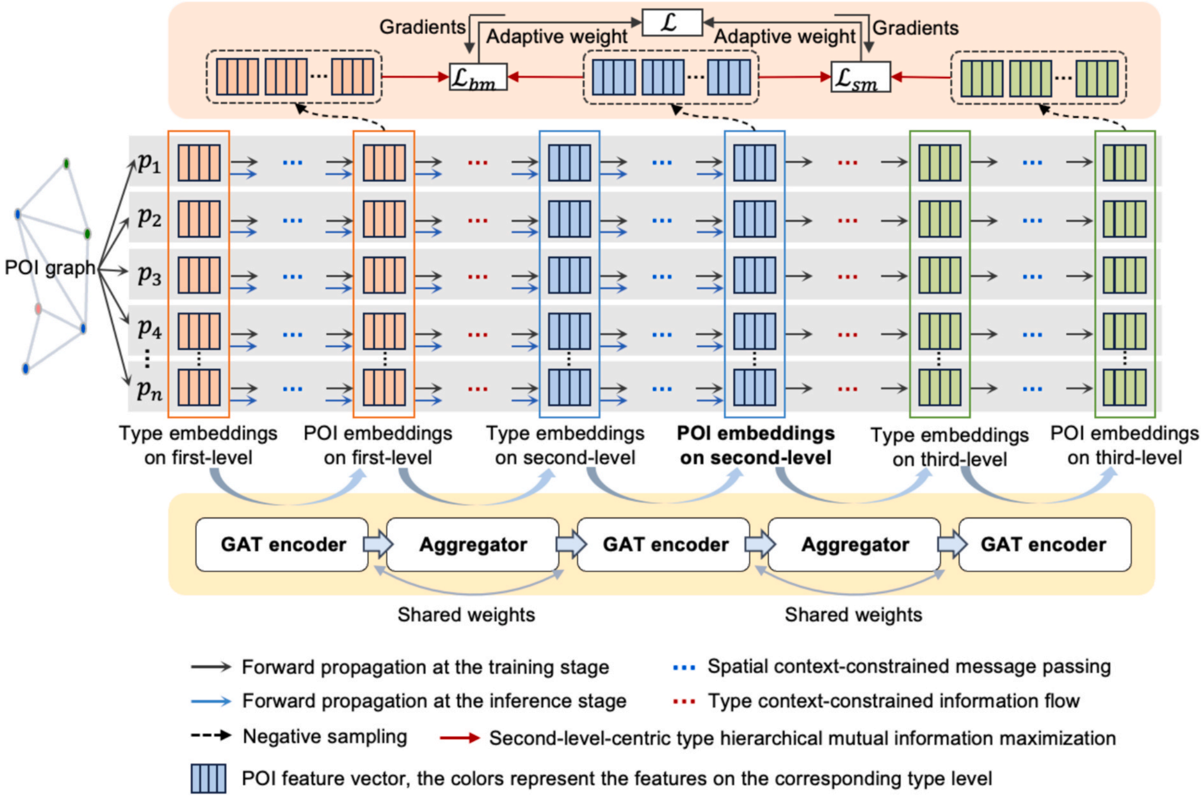

Gradients
Adaptive weight
$\mathcal{L}$
Adaptive weight
Gradients
$\mathcal{L}_{hm}$
$\mathcal{L}_{sm}$
POI graph
p_1
p_2
p_3
p_4
p_n
Type embeddings on first-level
POI embeddings on first-level
Type embeddings on second-level
POI embeddings on second-level
Type embeddings on third-level
POI embeddings on third-level
GAT encoder
Aggregator
GAT encoder
Aggregator
GAT encoder
Shared weights
Shared weights
Forward propagation at the training stage
Spatial context-constrained message passing
Forward propagation at the inference stage
Type context-constrained information flow
Negative sampling
Second-level-centric type hierarchical mutual information maximization
POI feature vector, the colors represent the features on the corresponding type level

Fig. 2. An overview of DCA.

are jointly shaped by both spatial context and type context during training process, leading to informative and discriminative representations.

2.2.1. *Sensing spatial context with GAT*

The spatial context of POIs reflects the spatial autocorrelation under the First Law of Geography, which emphasizes that nearby POIs tend to present similar features, and this is crucial to shape POI representations, especially when considering the semantic nuances of POIs of the same type. As the raw POI data provides geographical location of each POI but lacks the explicit information about spatial interconnections between POIs, our initial step is to model the spatial contextual relationships of POIs. To this end, we model POIs into a spatial relationship explicit graph structure based on their geographical locations, which is defined as a weighted undirected POI graph $G = (\mathscr{V}, \mathscr{E})$. $\mathscr{V} = \{p_i\}_{i=1}^{N_G}$ denotes vertexes (i.e., POI nodes) in the graph with $\boldsymbol{X} = \left\{\overrightarrow{t}_i\right\}_{i=1}^{N_G} \in \mathbb{R}^{N_G \times F}$ being the corresponding node feature matrix for describing POI information on type-wise, $\mathscr{E} = \{e_i\}_{i=1}^{M_G}$ denotes edges (i.e., pair-wise POI spatial context) with $\boldsymbol{E} = \left\{\vec{e}_i\right\}_{i=1}^{M_G} \in \mathbb{R}^{2\times M_G}$ being the edge index matrix, and $\boldsymbol{W}_G = \{w_i\}_{i=1}^{M_G} \in \mathbb{R}^{M_G}$ being the corresponding edge weight matrix, where $N_G = |\mathscr{V}|$ and $M_G = |\mathscr{E}|$ represent the number of vertexes and edges in G, respectively. The construction of the POI graph from POIs is illustrated in Fig. 3.

The construction of the graph structure mainly considers the spatial proximity between POIs to express the spatial context of each POI. Due to its merits (e.g., maximizing the minimum angle), the Delaunay Triangulation (DT) algorithm can assist in establishing rich spatial distribution information for POI graph (Kong et al., 2024; Xu et al., 2022; Yan et al., 2017). We unitize DT to construct the POI spatial adjacency to form $\mathscr{E}$. DT connects POIs into a triangular network based on their spatial distribution. The spatial adjacency of POIs in the geographic space is thus transformed into the neighborhood of each vertex in the graph. The more details could be found in Xu et al. (2022) with respect to DT-based POI spatial context graph construction. $\boldsymbol{W}_G$ describes spatially explicit constraint on message passing between nodes on POI graph, with edge weights used to treat information derived from disparate POIs within spatial contexts differently. Following the principle of distance decay effect, which emphasizes that the intensity of spatial interaction weakens as distance increases, edge weights can be naturally defined using a distance-based function that assigns higher weights to closely located POIs and lower weights to those farther apart. Given an edge connecting two POIs i and j in the graph G, its edge weight w_{ij} is defined as follows:

$$w_{ij} = \log\left((C+L^{\rho})\Big/\left(C+l_{ij}^{\rho}\right)\right)\cdot\mu \tag{1}$$

where L is diagonal length of the minimum bounding rectangle of all the POIs, ρ is an inverse distance factor,μ is a factor to differentiate cross-region edges (assigned a small value μ_{cross} to reflect the weaker spatial interaction across regions) and intra-region edges (assigned a larger value μ_{intra} to enforce spatial interaction within the same region), C is a constant to avoid infinity, μ_{cross}, μ_{intra}, C and ρ are set as 0.4, 1, 1 and 1.5 respectively, the above hyperparameters are set in reference to Huang et al. (2023). l_{ij} is the haversine distance between i and j. The final weight w_{ij} is transformed by a linear scaling to [0, 1].

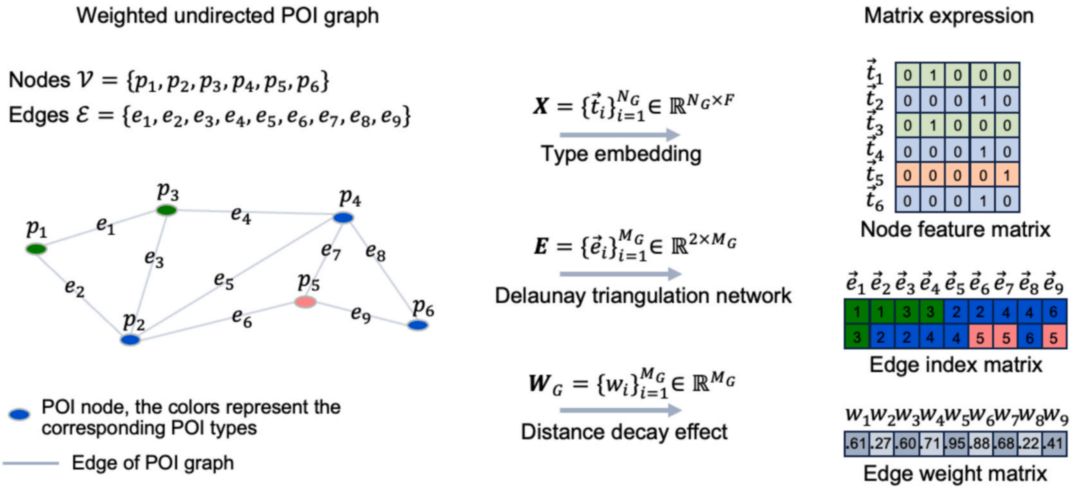
Along this line, we construct a POI graph covering the whole study area to serve as an input to the graph neural network (GNN), employing GNN to learn the spatial context information of the POIs. Existing GNN models generally treat all neighbor nodes (e.g., GCN) equally, but are unable to distinguish semantics between two POIs with the same spatial context. However, some types in the spatial context of POIs are more important than others in most cases, except for POIs with closer distances considered by distance decay weighting. The self-attention mechanism of the GAT provides a plausible solution to model heterogeneous semantic interactions (i.e., message passing processes) between POIs. Therefore, our encoder in the DCA is a one-layer GAT model $\phi_p : \mathbb{R}^F \rightarrow \mathbb{R}^{F'}$, and leverage self-attention mechanism to take the heterogeneous semantic influence between POIs into account, so that POI embeddings are delineated by sensing its type information and spatial context information.

Fig. 4 illustrates the encoding process of GAT. Given input graph node features, i.e., POI type embeddings $\boldsymbol{X} = \left\{\vec{t}_i\right\}_{i=1}^{N_G}$, $\vec{t}_i \in \mathbb{R}^F$ that describe type representations of POIs, the output POI embeddings that describe each individual POI representation, which convey its own type information and type information of the context and pass through GAT ϕ_p are denoted as $\boldsymbol{P} = \left\{\vec{p}_i\right\}_{i=1}^{n}$, $\vec{p}_i \in \mathbb{R}^{F'}$. The self-attention α of GAT ϕ_p is defined by a shared attention mechanism $\varphi : \mathbb{R}^{F'} \times \mathbb{R}^{F'} \rightarrow \mathbb{R}$. We introduce the distance decay effect to constrain the attention mechanism φ during model training, so that the model can better learn nuanced differences in spatial proximity, and prevent attention from over-relying on information of partial POIs. The distance decay effect constrained self-attention weights α_{ij} between given POI nodes i and j is expressed as the attention mechanism φ between the two nodes normalized by the softmax function.

$$\alpha_{ij} = \sigma_\alpha\left(w_{ij}\cdot\varphi\left(\boldsymbol{W}_\alpha \vec{t}_i, \boldsymbol{W}_\alpha \vec{t}_j\right)\right) \tag{2}$$

where w_{ij} is the weight of distance decay effect defined in Equation (1), $\boldsymbol{W}_\alpha \in \mathbb{R}^{F'\times F}$ is a learnable shared linear transformation that applies to all nodes, F represent dimension of the initial POI embeddings and F' represent dimension of the output POI embeddings of GAT ϕ_p, σ_α is a softmax to normalize attention weights, $\text{softmax}(\cdot) = \exp(\cdot)/\sum\exp(\cdot)$. Attention mechanism is a one-layer feedforward neural network, i.e., $\varphi\left(\boldsymbol{W}_\alpha \vec{t}_i, \boldsymbol{W}_\alpha \vec{t}_j\right) = \sigma_\varphi\left(\vec{a}^T\left[\boldsymbol{W}_\alpha \vec{t}_i \middle\| \boldsymbol{W}_\alpha \vec{t}_j\right]\right)$, where $\vec{a} \in \mathbb{R}^{2F'}$ is a learnable weight vector for φ, σ_φ is a LeakyReLU nonlinearity, $\text{LeakyReLU}(\cdot) = \max(0.2\cdot, \cdot)$, $\bullet^T$ represents transposition, and $\|$ is feature-wise concatenation operation.

We extend the distance decay effect constrained attention mechanism to a multi-head attention mechanism, averaging the spatial context


Weighted undirected POI graph
Nodes $\mathcal{V} = \{p_1, p_2, p_3, p_4, p_5, p_6\}$
Edges $\mathcal{E} = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9\}$
p_1
p_2
p_3
p_4
p_5
p_6
e_1
e_2
e_3
e_4
e_5
e_6
e_7
e_8
e_9
POI node, the colors represent the corresponding POI types
Edge of POI graph
$\boldsymbol{X} = \{\vec{t}_i\}_{i=1}^{N_G} \in \mathbb{R}^{N_G\times F}$
Type embedding
$\boldsymbol{E} = \{\vec{e}_i\}_{i=1}^{M_G} \in \mathbb{R}^{2\times M_G}$
Delaunay triangulation network
$\boldsymbol{W}_G = \{w_i\}_{i=1}^{M_G} \in \mathbb{R}^{M_G}$
Distance decay effect
Matrix expression
Node feature matrix
Edge index matrix
Edge weight matrix

$\vec{t}_1$	0	1	0	0	0
$\vec{t}_2$	0	0	0	1	0
$\vec{t}_3$	0	1	0	0	0
$\vec{t}_4$	0	0	0	1	0
$\vec{t}_5$	0	0	0	0	1
$\vec{t}_6$	0	0	0	1	0

$\vec{e}_1$	$\vec{e}_2$	$\vec{e}_3$	$\vec{e}_4$	$\vec{e}_5$	$\vec{e}_6$	$\vec{e}_7$	$\vec{e}_8$	$\vec{e}_9$
1	1	3	3	2	2	4	4	6
3	2	2	4	4	5	5	6	5

w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9
.61	.27	.60	.71	.95	.88	.68	.22	.41

Fig. 3. The construction of the POI graph.

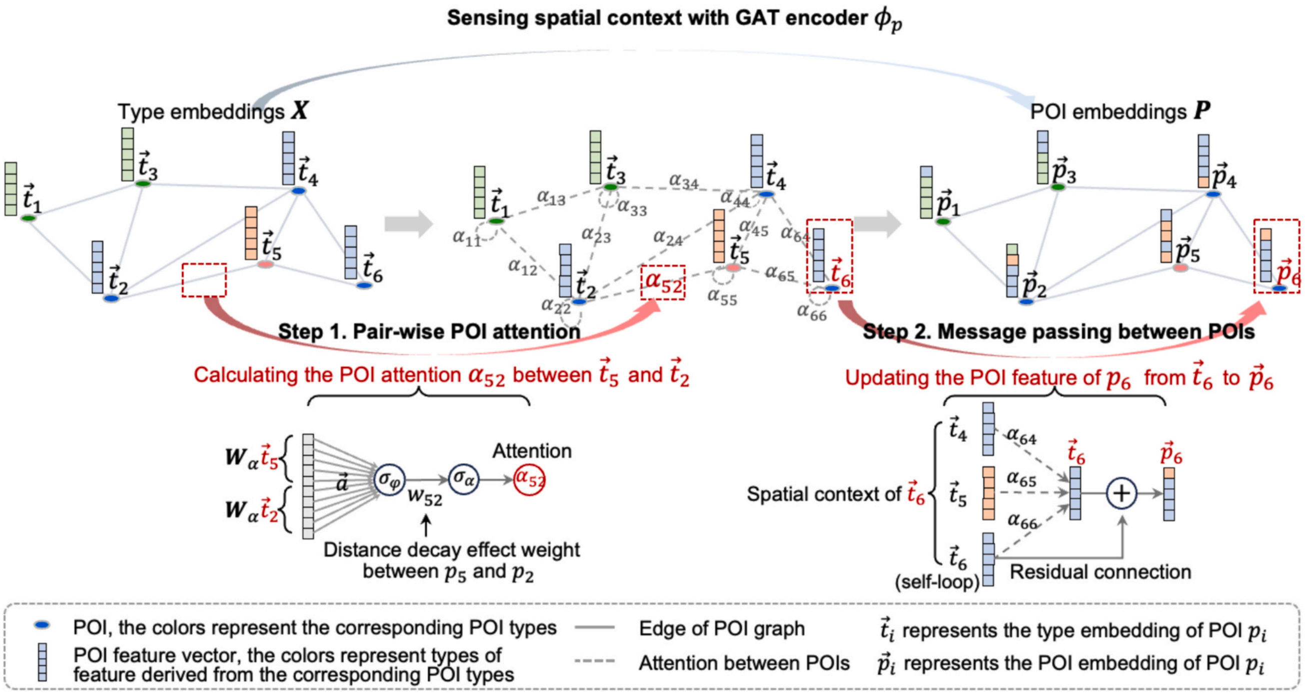

Sensing spatial context with GAT encoder ϕ_p
Type embeddings $\boldsymbol{X}$
$\vec{t}_1$ $\vec{t}_2$ $\vec{t}_3$ $\vec{t}_4$ $\vec{t}_5$ $\vec{t}_6$
α_{11} α_{12} α_{13} α_{22} α_{23} α_{33} α_{34} α_{24} α_{44} α_{45} α_{64} α_{52} α_{55} α_{65} α_{66}
POI embeddings $\boldsymbol{P}$
$\vec{p}_1$ $\vec{p}_2$ $\vec{p}_3$ $\vec{p}_4$ $\vec{p}_5$ $\vec{p}_6$
Step 1. Pair-wise POI attention
Calculating the POI attention α_{52} between $\vec{t}_5$ and $\vec{t}_2$
$\boldsymbol{W}_\alpha \vec{t}_5$
$\boldsymbol{W}_\alpha \vec{t}_2$
$\vec{a}$
σ_φ
w_{52}
σ_α
Attention
α_{52}
Distance decay effect weight between p_5 and p_2
Step 2. Message passing between POIs
Updating the POI feature of p_6 from $\vec{t}_6$ to $\vec{p}_6$
Spatial context of $\vec{t}_6$
$\vec{t}_4$ $\vec{t}_5$ $\vec{t}_6$
(self-loop)
α_{64} α_{65} α_{66}
$\vec{t}_6$ $\vec{p}_6$
Residual connection
POI, the colors represent the corresponding POI types
POI feature vector, the colors represent types of feature derived from the corresponding POI types
Edge of POI graph
Attention between POIs
$\vec{t}_i$ represents the type embedding of POI p_i
$\vec{p}_i$ represents the POI embedding of POI p_i

Fig. 4. Illustration of spatial context-constrained message passing with GAT encoder.

information aggregated by all attention heads to update the information of each node to stabilize the self-attention training process. Thus, for the POI node update of K independent attention mechanisms, the output feature of node i is expressed as $\vec{t}'_i = \sigma_p\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in\mathscr{C}_i}\alpha_{ij}^k \boldsymbol{W}_\alpha^k \vec{t}_j\right)$, where $\mathscr{C}_i$ is a spatial context set (i.e., 1-hop neighbors) of node i in the POI graph (including i), attention weight α_{ij}^k comes from k-th attention mechanism φ^k, and σ_p is a ReLU nonlinearity, $\mathrm{ReLU}(\cdot) = \max(0, \cdot)$. Finally, we add a residual connection, which results in the encoding process of ϕ_p for output embedding $\vec{p}_i$:

$$\phi_p\left(\vec{t}_i\right) = \vec{t}'_i + \vec{t}_i \tag{3}$$

Each POI embedding is updated by integrating its type information and spatial context information based on multi-head self-attention mechanism of GAT ϕ_p. Multi-head self-attention mechanism reflects the heterogeneous semantic interaction of POIs from multiple different perspectives (multi-independent attention heads), and is constrained by distance decay effect. At this point, the GAT ϕ_p can take the distance decay effect and the heterogeneous semantic interaction of POIs into account to sense POI spatial context.

2.2.2. *Generating POI embeddings via DCA forward propagation*

The previous section described how GAT in DCA encodes POI spatial context information to shape POI embeddings. However, the POI taxonomy usually consists of three type levels, namely the first-level, second-level, and third-level. The POI embeddings are shaped by spatial context information on a single type level, which ignores type context, i.e., the hierarchical correlations between types inherent in the POI taxonomy. Type context can provide a view of the intrinsic type hierarchy of POIs to complement the spatial context, and is crucial for finer semantic shaping of POI embedding. It is natural to assume that embeddings between the first ~ second-level or second ~ third-level with hierarchical affiliations should be interdependent. In order to establish the semantic interdependence in response to type context, we design a type hierarchical aggregation neural network architecture. In the forward propagation process of this network, the type information of POI embeddings is aggregated in a bottom-up way, and the structural relationship of type information between POI embeddings on the three type levels is established. Fig. 5 illustrates the hierarchical aggregation of type information from POI embeddings, for instance, the type embedding of third-level type “Enterprise” is aggregated by the POI embeddings of second-level type “Company” and “Factory”, and the type embedding of second-level type “Company” also conveys the type information of POI embeddings of third-level type “Medical Company” and “Mining Company”.

As a start of the DCA forward propagation, we randomly initialize type embedding $\boldsymbol{X}^{\mathrm{f}} = \left\{\vec{t}^{\,\mathrm{f}}_i\right\}_{i=1}^{N_G}$ on first-level on type-wise. Then, we can generate POI embeddings $\boldsymbol{P}^{\mathrm{f}}$ on first-level on node-wise through GAT encoder, the POI embedding on first-level of POI i can be calculated as $\vec{p}^{\,\mathrm{f}}_i = \phi_p\left(\vec{t}^{\,\mathrm{f}}_i\right)$. The next step is to aggregate POI embeddings on first-level, depending on the hierarchical relationships between first ~ second-level, and generate type embeddings $\boldsymbol{X}^{\mathrm{s}} = \left\{\vec{t}^{\,\mathrm{s}}_i\right\}_{i=1}^{N_G}$ on second-level on type-wise with an aggregator. For the aggregation function, we employ element-wise average pooling that conforms to permutation invariance, aligning with the absence of any inherent order relationship within the type hierarchy. Let l and $l' \in \{\mathrm{f}, \mathrm{s}, \mathrm{t}\}$ represent the type levels of the previous and next hierarchical levels, respectively. This aggregation function can be formalized as follows.

$$\mathrm{AGG}\left(\boldsymbol{P}^{l}\right) = \sigma_a\left(\frac{1}{\left|\mathscr{N}_j^{l'}\right|}\sum_{j\in\mathscr{N}_j^{l'}}\vec{p}^{\,l}_i\right) \tag{4}$$

where σ_a is a sigmoid nonlinearity. $\mathscr{N}_j^{l'}$ represents the set of POIs belonging to type j on l'-level.

Given the POI embeddings belonging to the second-level type j, we aggregate POI embeddings on first-level that share the same second-level type to calculate their common type embedding $\vec{t}^{\,\mathrm{s}}_j$ on second level as $\boldsymbol{X}^{\mathrm{s}} = \mathrm{AGG}\left(\boldsymbol{P}^{\mathrm{f}}\right)$.

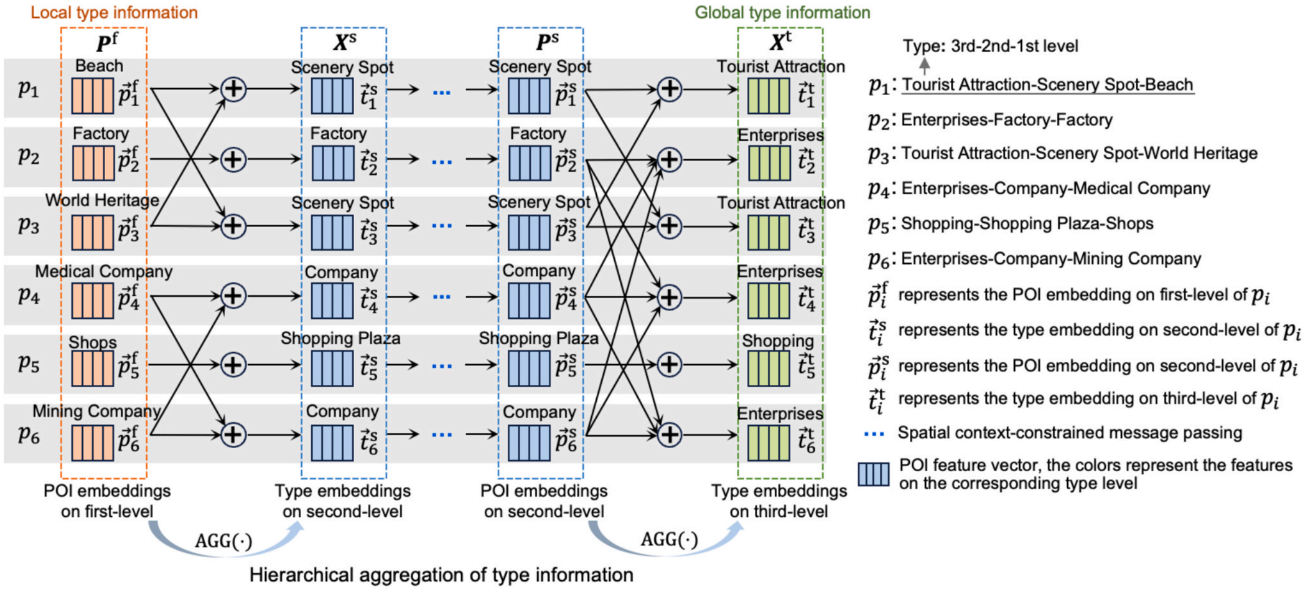

Local type information
Global type information
$\boldsymbol{P}^{\mathrm{f}}$
$\boldsymbol{X}^{\mathrm{s}}$
$\boldsymbol{P}^{\mathrm{s}}$
$\boldsymbol{X}^{\mathrm{t}}$
Beach
Factory
World Heritage
Medical Company
Shops
Mining Company
Scenery Spot
Factory
Scenery Spot
Company
Shopping Plaza
Company
Tourist Attraction
Enterprises
Tourist Attraction
Enterprises
Shopping
Enterprises
p_1 p_2 p_3 p_4 p_5 p_6
POI embeddings on first-level
Type embeddings on second-level
POI embeddings on second-level
Type embeddings on third-level
AGG(·)
AGG(·)
Hierarchical aggregation of type information
Type: 3rd-2nd-1st level
p_1: Tourist Attraction-Scenery Spot-Beach
p_2: Enterprises-Factory-Factory
p_3: Tourist Attraction-Scenery Spot-World Heritage
p_4: Enterprises-Company-Medical Company
p_5: Shopping-Shopping Plaza-Shops
p_6: Enterprises-Company-Mining Company
$\vec{p}_i^{\mathrm{f}}$ represents the POI embedding on first-level of p_i
$\vec{t}_i^{\mathrm{s}}$ represents the type embedding on second-level of p_i
$\vec{p}_i^{\mathrm{s}}$ represents the POI embedding on second-level of p_i
$\vec{t}_i^{\mathrm{t}}$ represents the type embedding on third-level of p_i
··· Spatial context-constrained message passing
POI feature vector, the colors represent the features on the corresponding type level

Fig. 5. Illustration of type context-constrained information aggregation with aggregator.

Thus, the type embeddings $\boldsymbol{X}^{\mathrm{s}}$ on second-level convey the information from all POI embeddings belonging to first-level types it subsumes. In other words, $\boldsymbol{X}^{\mathrm{s}}$ convey the information of all type embeddings of its first-level types and the type embeddings of the POIs in its spatial context. Subsequently, the type embeddings $\boldsymbol{X}^{\mathrm{s}}$ are fed to GAT ϕ_p to generate the POI embedding $\boldsymbol{P}^{\mathrm{s}}$ on second-level, i.e., $\overrightarrow{p}_i^{\mathrm{s}} = \phi_p\left(\overrightarrow{t}_i^{\mathrm{s}}\right)$.

In the subsequent step, we apply the aggregation function to derive the type embeddings $\boldsymbol{X}^{\mathrm{t}}$ on third-level from the POI embeddings on second-level in the same way. Then, the calculation of type embeddings on third-level is conducted as $\boldsymbol{X}^{\mathrm{t}} = \mathrm{AGG}(\boldsymbol{P}^{\mathrm{m}})$. Finally, with the help of GAT, the spatial context information of POI is injected into type embeddings on third-level to generate POI embeddings $\boldsymbol{P}^{\mathrm{t}}$ on third-level, i.e., $\overrightarrow{p}_i^{\mathrm{t}} = \phi_p\left(\overrightarrow{t}_i^{\mathrm{t}}\right)$.

In the forward propagation process of DCA, due to the fixed nature of spatial relationships among POIs, we employ a shared-weight GAT encoder, along with $\mathscr{E}$ and $\boldsymbol{W}_G$, to calculate POI embeddings on three type levels. Thus far, we have gradually obtained POI embeddings on three type levels along DCA forward propagation. POI embeddings on first-level serve as the bottom layer of information propagation, providing local type information, while POI embeddings on third-level serve as the top layer for information propagation, conveying global type information. Throughout the forward propagation in DCA, GAT encoder constrains information flow between POIs, establishing semantic interdependence among POI embeddings on node-wise (i.e., spatial context). Simultaneously, the aggregator constrains unidirectional information flow between type hierarchies of POIs, establishing semantic interdependence among POI embeddings on type hierarchy-wise (i.e., type context).

2.2.3. *Sensing type context with type infomax*

We obtain POI embedding on three type-levels through the forward propagation of DCA. However, the bottom-up aggregation strategy employed by DCA solely senses unidirectional hierarchical relationships among the POI embeddings on three type-levels, resulting in the unidirectional type context awareness of POI embeddings rather than bidirectional. In other words, the POI embeddings of higher type level convey the type information of the lower type level, but not vice versa. In this context, there is a need for an explicit constraint to sense bidirectional hierarchical relationships (i.e., type context), encouraging the entire network to focus on type context information of POI types during the process of encoding spatial context information. To this end, type context of POIs is regarded as self-supervised signals to define a meaningful optimization objective to guide DCA training.

In principle, there should be stronger interdependence among POI embeddings within the type context. For instance, POI embeddings on second-level belong to the same third-level types should have a closer semantic connection. In view of this, we employ MI to measure the interdependence between type hierarchies, and follow the insights from deep graph infomax (DGI) (Veličković et al., 2019) and hierarchical graph infomax (HGI) (Huang et al., 2023) which both capture the structural relationship by maximizing the MI between the local information representations and global information representations. Intuitively, we choose POI embeddings on second-level as the final output POI representation embeddings, thus POI embeddings on third-level and first-level are naturally regarded as local type information representations and global type information representations. Subsequently, the optimization objective of DCA can be designed as a second-level-centric type infomax to capture the type hierarchical relationships of POIs (i.e., type context), which maximizing the MI between POI embeddings on first $\sim$ second-level and second $\sim$ third-level. And, a contrastive learning mechanism is adopted to prevent feature collapse issues. The mechanism behind second-level-centric type infomax for sensing type context is to constrain the information flow between POI embeddings on first $\sim$ second $\sim$ third level, by maximizing the MI between POI embeddings that belong to some type context, and minimizing the MI between POI embeddings from different type contexts.

As the POI embeddings on three type levels are obtained by DCA forward propagation, negative sampling is further implemented to fetch positive and negative samples for calculating type infomax loss. Specifically, given a POI p_k with embedding on second-level $\overrightarrow{p}_k^{\mathrm{s}}$, any POI embeddings on third-level that subsumes the second-level type of p_k are positive samples, while the others are regarded as negative samples to $\overrightarrow{p}_k^{\mathrm{s}}$. Similarly, we build the positive and negative samples on first-level to $\overrightarrow{p}_k^{\mathrm{s}}$. The entire negative sampling is completed by traversing all POI embeddings on second-level. The optimization of DCA model can be expressed mathematically as minimizing the following objectives:

$$\mathscr{L}_{\mathrm{fs}} = -\left(\frac{1}{n}\sum_{k=1}^{N_P}\sum_{i=1}^{n_k}\log\left(\mathscr{D}_{\mathrm{fs}}\left(\overrightarrow{p}^{\mathrm{f}}_{i}, \overrightarrow{p}^{\mathrm{s}}_{k}\right)+\epsilon\right)+\frac{1}{\widetilde{n}}\sum_{k=1}^{N_P}\sum_{i=1}^{\tilde{n}_k}\log\left(1 - \mathscr{D}_{\mathrm{fs}}\left(\overrightarrow{\widetilde{p}}^{\mathrm{f}}_{i}, \overrightarrow{p}^{\mathrm{s}}_{k}\right)+\epsilon\right)\right) \tag{5}$$

$$\mathscr{L}_{\mathrm{st}} = -\left(\frac{1}{n}\sum_{k=1}^{N_P}\sum_{i=1}^{n_k}\log\left(\mathscr{D}_{\mathrm{st}}\left(\overrightarrow{p}^{t}_{i}, \overrightarrow{p}^{s}_{k}\right)+\epsilon\right)+\frac{1}{\widetilde{n}}\sum_{k=1}^{N_P}\sum_{i=1}^{\tilde{n}_k}\log\left(1 - \mathscr{D}_{\mathrm{st}}\left(\overrightarrow{\widetilde{p}}^{\mathrm{t}}_{i}, \overrightarrow{p}^{\mathrm{s}}_{k}\right)+\epsilon\right)\right) \tag{6}$$

where $\mathscr{L}_{\mathrm{fs}}$ and $\mathscr{L}_{\mathrm{st}}$ respectively represent the loss functions on positive and negative samples between first $\sim$ second-level and second $\sim$ third-level, n_k and $\widetilde{n}_k$ are the number of positive and negative samples corresponding to p_k, N_P is the total number of POIs, $n = \sum_{k=1}^{N} n_k$ and $\widetilde{n} = \sum_{k=1}^{N} \widetilde{n}_k$ are the total number of positive and negative samples, $\overrightarrow{p}^{\mathrm{f}}_{i}$ and $\overrightarrow{\widetilde{p}}^{\mathrm{f}}_{i}$ are positive and negative samples on the first-level to p_k respectively, $\overrightarrow{p}^{\mathrm{t}}_{i}$ and $\overrightarrow{\widetilde{p}}^{\mathrm{t}}_{i}$ are positive and negative samples on the third-level to p_k respectively, ϵ is a small positive constant. $\mathscr{D}_{\mathrm{fs}} : \mathbb{R}^{F\times F}\rightarrow\mathbb{R}$ and $\mathscr{D}_{\mathrm{st}} : \mathbb{R}^{F\times F}\rightarrow\mathbb{R}$ are discriminator, which are employed as a proxy for maximizing the type hierarchy MI between first $\sim$ second-level and second $\sim$ third-level respectively. $\mathscr{D}_{\mathrm{fs}}$ and $\mathscr{D}_{\mathrm{st}}$ can be calculated by a bilinear scoring function referred to a discriminator scoring used in Veličković et al. (2019), i.e., $\mathscr{D}(\boldsymbol{u}, \boldsymbol{v}) = \sigma_{\mathscr{D}}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{W}_{\mathscr{D}}\boldsymbol{v})$, where $\sigma_{\mathscr{D}}$ is a sigmoid nonlinearity and $\boldsymbol{W}_{\mathscr{D}}$ is a learnable scoring matrix.

By combining $\mathscr{L}_{\mathrm{fs}}$ and $\mathscr{L}_{\mathrm{st}}$, we obtain the optimization objective of DCA for sensing type context. As $\mathscr{L}_{\mathrm{fs}}$ and $\mathscr{L}_{\mathrm{st}}$ respectively guide the modeling of interdependencies first $\sim$ second-level and second $\sim$ third-level, A trade-off emerges in POI embeddings (second-level) between emphasizing local (first-level) and global (third-level) type information during representation learning. We set λ as the weight for the joint losses to balance this trade-off. Considering that $\mathscr{L}_{\mathrm{fs}}$ and $\mathscr{L}_{\mathrm{st}}$ share the GAT encoder in DCA, we employ the GradNorm technique (Chen et al., 2018), an effective gradient normalization technique in multi-task learning. GradNorm technique is utilized to search the optimal balance point dynamically, so that parameters of the shared GAT encoder can converge to a robust state which is useful across all losses. The λ is dynamically adjusted during each iteration in the model training process based on the ℓ_2 norm of the gradients of the shared GAT encoder, rather than using a fixed λ for the entire training stage, remedying the gradient domination issue of GAT ϕ_p. The final optimization objective (loss function) of DCA is as follows:

$$\mathscr{L} = \lambda_{\mathrm{fs}}\mathscr{L}_{\mathrm{fs}} + \lambda_{\mathrm{st}}\mathscr{L}_{\mathrm{st}} \tag{7}$$

where λ_{fs} and $\lambda_{\mathrm{st}}(> 0)$ are learnable weights for $\mathscr{L}_{\mathrm{fs}}$, $\mathscr{L}_{\mathrm{st}}$, which are dynamically adjusted by GradNorm technique during the model training process.

Since $\mathscr{L}_{\mathrm{st}}$ encourages the model to capture global type features that represent general commonalities among POI embeddings on second-level and $\mathscr{L}_{\mathrm{fs}}$ encourages the model to focus on local type features, the joint loss $\mathscr{L}$ with the second-level-centric type infomax objective can guide both local (first-level) and global (third-level) type information to flow adaptively into POI embeddings on second-level at a similar rate constrained by GradNorm. Under the guidance of $\mathscr{L}$, GAT encoder ϕ_p is constrained to pay attention to type context while encoding spatial context information. Specifically, the gradients from backpropagation encourage increasing attention to positive samples while decreasing attention to negative samples in self-attention training, and further influence the information flows of POI embeddings between global and local type information. As illustrated in Fig. 6, DCA pushes away POI embeddings for different contexts (i.e., belong to different third-level types or subsume different first-level types), as well as pulls closer POI embeddings with same type context. With regard to POI embeddings with same third-level types but different first-level types or same first-level types but different third-level types, DCA pushes them away and pulls them closer synchronously, and falls them to the adaptive balance point with the help of GradNorm. Hence, POI embeddings on second-level are sculpted from both the global (third-level) and local (first-level) perspectives result in sensing type context. At last, POI embeddings on second-level shaped by both type context and spatial context jointly, and output as the final dual context aware POI representation embeddings.

2.3. Fine-tuning for geographic mapping

Since the DCA is trained in a self-supervised manner, the learned POI representations are generic and task-agnostic. Therefore, we select several representative geographic mapping tasks, including urban function, region popularity, and housing price mapping, which play a pivotal role in urban planning and management (Gu et al., 2025; Hou et al., 2024; Manvi et al., 2024; Qin et al., 2022b; Zhang et al., 2024a), to evaluate the effectiveness of the learned representations.

The first step is to acquire region embeddings by aggregating POI embeddings within the region. While more sophisticated aggregation architectures such as multi-head attention-based POI aggregation (Huang et al., 2023) are likely to be more appropriate for aggregation functions, we limit our focus to the aspect of POI representation learning, and employ a simple element-wise average pooling to generate region embeddings for ease of comparison with baselines. For the

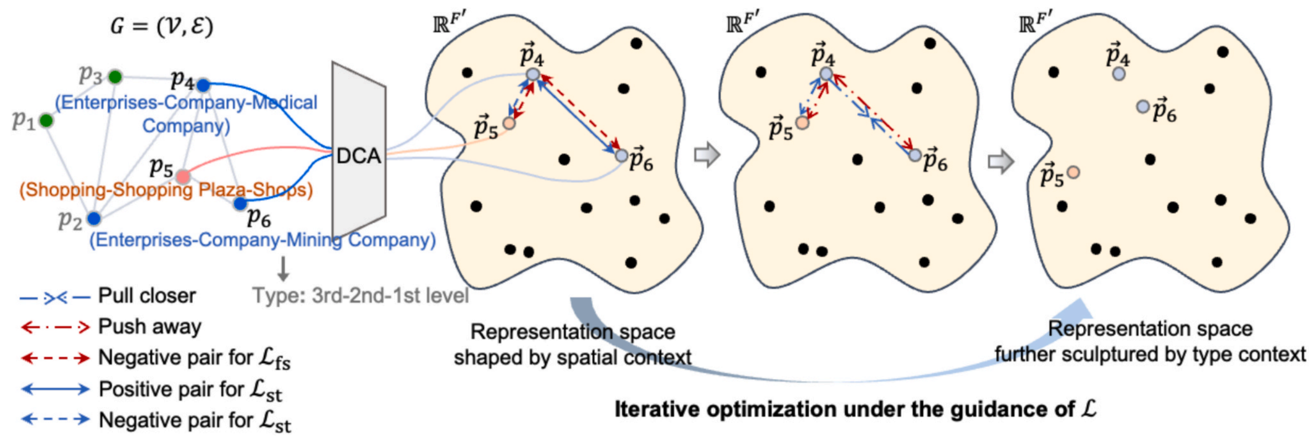

$G = (\mathcal{V}, \mathcal{E})$
p_1
p_2
p_3
p_4
(Enterprises-Company-Medical Company)
p_5
(Shopping-Shopping Plaza-Shops)
p_6
(Enterprises-Company-Mining Company)
Type: 3rd-2nd-1st level
DCA
$\mathbb{R}^{F'}$
$\vec{p}_4$
$\vec{p}_5$
$\vec{p}_6$
Pull closer
Push away
Negative pair for $\mathcal{L}_{fs}$
Positive pair for $\mathcal{L}_{st}$
Negative pair for $\mathcal{L}_{st}$
Representation space shaped by spatial context
Representation space further sculptured by type context
Iterative optimization under the guidance of $\mathcal{L}$

Fig. 6. Illustration of sensing type context with type infomax.

embedding of region r_j, it can be calculated by the following:

$$\overrightarrow{r}_j = \frac{1}{|\mathcal{N}_{r_j}|}\sum_{i\in\mathcal{N}_{r_j}} \overrightarrow{p}_i \tag{8}$$

where $\mathcal{N}_{r_j}$ is the set of POIs fall in r_j.

Region representations derived from the aggregation of generic POI representations can be competent for region scale downstream tasks, with an additional classifier or regressor accepting the task corresponding labels of regions for supervised learning. We follow the adapter-based delta-tuning paradigm (Ding et al., 2023) and employ a multilayer perceptron (MLP) $\phi_z : \mathbb{R}^{F'} \rightarrow \mathbb{R}^{F_d}$ as a task head tailored to specific tasks, where F' is the dimensions of region embeddings and F_d is the dimensions of output. We freeze the parameters of a pre-trained DCA and finetune the task head to adapt to specific geographic mapping tasks.

3. Experiment and results

3.1. Experimental setting

3.1.1. Study area and data

We chose the fifth ring road area of Beijing, China, a mature urban area, to verify the effectiveness of DCA. Five datasets are involved in this study. The spatial distribution of POIs and urban regions are shown in Fig. 7.

- POI data are harvested from the API of Amap (a.k.a. Gaode Map, https://lbs.amap.com) in 2018. POI is vector point data, comprises several fields, including longitude, latitude, first-level type, second-level type, and third-level type. There is a POI type taxonomy tree (https://lbs.amap.com/api/webservice/download) to connect first ∼ second ∼ third-level, which consist of 12 third-level types, 100 second-level types and 872 first-level types. 208,929 POIs are obtained after data cleaning, including removing duplicate POIs, eliminating POIs with missing fields. A POI graph with 208,929 vertexes and 935,841 edges is constructed using DT, serves as the input of DCA.
- Essential urban land use categories (EULUC) data (Gong et al., 2020) is a dataset of urban land functional use in China in 2018, produced based on high-resolution images, mobile-phone locating-request data, POIs and nighttime light images. Labels on level I of EULUC are used as the ground truth to subsequent function mapping: residential, commercial, industrial, transportation, public management and service. A total of 2,553 samples serves as the input of MLP.
- Housing price data are collected from Lianjia (https://lianjia.com/). A total of 150,516 price data are average pooling to each region to serve as the ground truth for housing price mapping.
- Check-in data are collected from Weibo (https://weibo.com/). Referencing (Li et al., 2024; Zhang et al., 2023b), a total of 1,003,960 geotagged check-in data are aggregated within each region serve as the ground truth for region popularity mapping.

3.1.2. Evaluation metrics

In the context of evaluating the effectiveness of DCA, well-established evaluation metrics are adopted to gauge the performance of DCA on different geographic mapping tasks. Specifically, referring to (Xu et al., 2022; Zhang et al., 2023b), we treat urban function mapping as a multi-class classification problem and select the evaluation metrics below: overall accuracy (OA), kappa coefficient (Kappa), and macro-averaged F1 score (MacF1). We treat housing price mapping and region popularity mapping as regression problems and select the following evaluation metrics: mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2). Those metrics are defined as follows (the symbols ↑ and ↓ denote that higher and lower values are better performance, respectively).

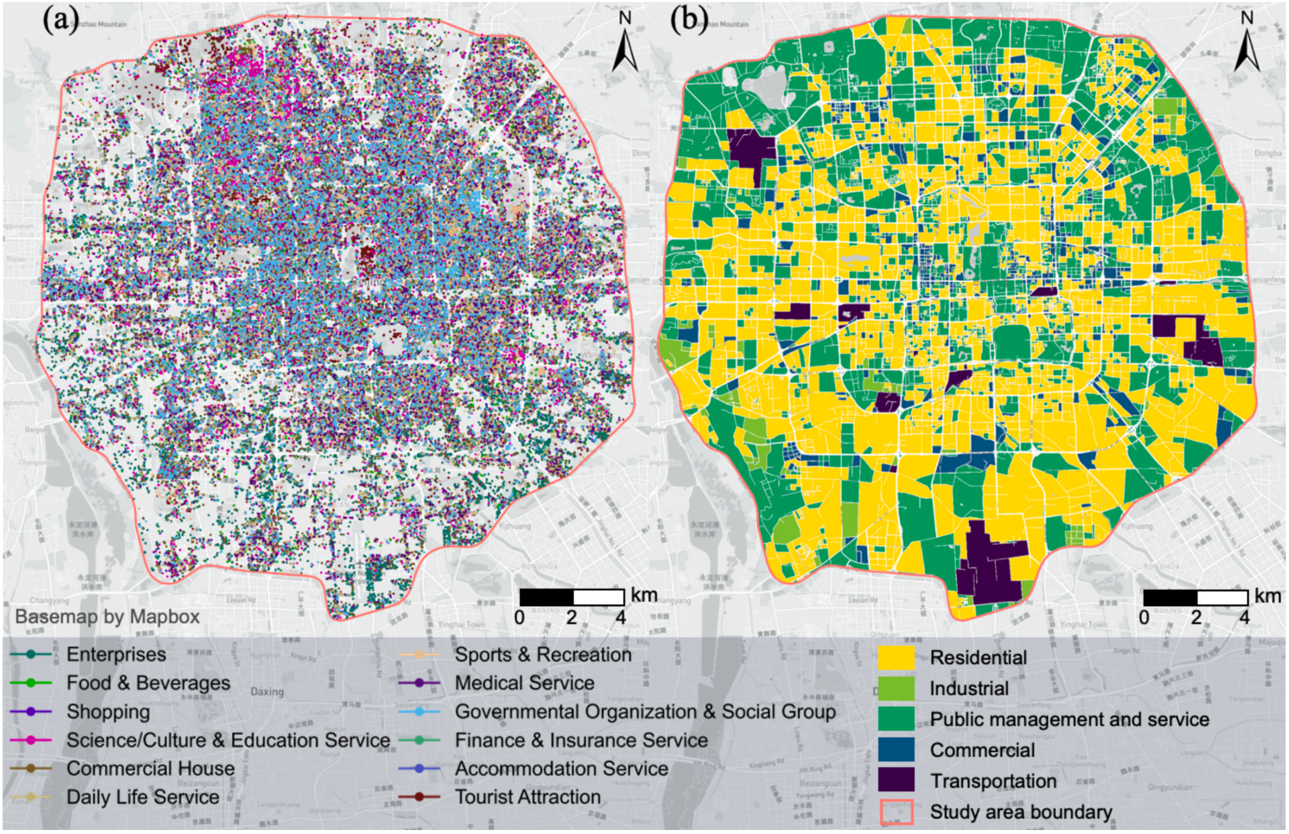


Fig. 7. The spatial distribution of (a) POIs in the study area colored by their third-level types, (b) Urban regions with EULUC functional labels.

- OA ↑: $OA = \frac{TP+TN}{TP+TN+FP+FN}$ which measures the proportion of correct predictions relative to the ground truth.
- Kappa ↑: $Kappa = \frac{OA-P_e}{1-P_e}$ which measures the consistency between the overall classification results and the ground truth
- MacF1 ↑: $MacF1 = \frac{1}{k}\sum_{i=1}^{k} F_1^i$ which measures the average performance of the model across all classes.
- MAE ↓: $MAE = \frac{1}{n}\sum_{i=1}^{n} \left|y_i - \widehat{y}_i\right|$ which measures the average absolute difference between predictions and ground truth.
- RMSE ↓: $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(y_i - \widehat{y}_i\right)^2}$ which measures the standard deviation of prediction errors.
- R^2 ↑: $R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2}$ which evaluates goodness-of-fit of the model.

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively, $P_e = \frac{(TP+FN)(TP+FP)+(FP+TN)(FN+TN)}{(TP+TN+FP+FN)^2}$, $F_1^i = \frac{2\cdot Precision \bullet Rrecall}{Precision+Recall}$, $Precision = \frac{TP}{TP+FP}$, $Recall = \frac{TP}{TP+FN}$, k is the number of function categories. n is the size of testing set, y_i and $\widehat{y}_i$ are prediction and ground truth of sample i, and $\overline{y}$ is the average of the ground truth.

3.1.3. Baseline models

Several well-accepted (stable) models ranging from classical to state-of-the-art are used as the baselines to compare with the proposed DCA. For the implementation of the baselines, we follow the convention of previous studies and only involve the second-level of POIs. Note that the baselines only involve the generation of region embeddings, and MLP task head with the same architecture are used for the geographic mapping task to be consistent with DCA.

- GCN (Xu et al., 2022): This is a supervised model for the end-to-end geographic mapping task which learns spatial context information of POIs. It is trained by an unsupervised task of node classification in this work, and used as the representation learning model, but not for urban function mapping. This ensures consistency with DCA to facilitate performance comparison. For the setup, we employ a one-layer GCN, and the dimension of the output layer set to 64. It is trained on the DT-based POI graph same as DCA.
- Word2Vec (CBOW architecture) (Yao et al., 2017): This model first samples the textual sequence of POIs using a greedy algorithm, and learns bidirectional spatial context information of POIs. The output layer dimension is set to 64.
- Latent Dirichlet Allocation (LDA) (Liu et al., 2017): This method is a topic model which infers the probability distribution among zones-functional topics (i.e., clusters of second-level types)-POIs. The topic distribution for each zone generated by LDA is regarded as a zone representation for downstream tasks. The output dimension of the zone embeddings is set to 64.
- Term Frequency-Inverse Document Frequency (TFIDF) (Liu et al., 2020): This model encodes urban functional zones based on the frequency distribution of POI types. Since TFIDF cannot actively set a dimension, the dimension of zone embeddings matches the number of second-level types.

3.2. Generating POI embeddings

3.2.1. Implementation details

We instantiate a DCA model with the tuned hyperparameter combination {$d = 64$, h=4} (cf. Section 3.3.3 for the hyperparameter tuning), and deploy it on a single NVIDIA Quadro RTX 8000 GPU for training. We adopt an Adam optimizer with an initial learning rate of 1 × 10^{-3}, and use a gradient clipping technique (constrain the ℓ_2 norm of the gradients no more than 0.9) to accelerate model training. We use a linear learning rate warmup technique in the first 20 epochs (training iterations) to stabilize the model training. We train the model for a maximum of 200 epochs w/o early stopping strategy, and the model with the lowest loss in all training epochs is retrieved as the trained model, and the corresponding output embeddings on second-level are retrieved as the final POI embeddings.

3.2.2. Visualization analysis of POI embedding space

Mapping POI embeddings to a 2-D space facilitates the visualization of embedding space, and analyzing its semantic lay out in the embedding space. We adopt the t-distributed stochastic neighbor embedding (t-SNE), an effective non-linear dimension reduction algorithm which is generally better than other algorithms (e.g., principal component analysis) for high dimensionality (Liu et al., 2020; van der Maaten and Hinton, 2008). As illustrated in Fig. 8, the embeddings of each individual POI are mapped to a 2-D space, where colors are rendered according to their third-level types. Fig. 8 (a) depicts the spatial context aware POI embeddings derived from GCN. The embedding layout from GCN is completely determined by the second-level type information of POIs itself and its spatial context information, which indicates that the more similar the spatial context of the POIs (including itself) is, the closer the distance is in the embedding space. The actual embedding layout from GCN exhibits a noticeable clustering pattern, wherein POI embeddings of the same type generally tend to converge to each other. Moreover, the distribution of POI embeddings belonging to the same type does not strictly cluster solely based on their type information, which reflects the semantic meaning of the corresponding spatial context of each unique POI. Fig. 8 (b) depicts the dual context aware POI embeddings derived from DCA, whose embedding layout is affected by both the spatial and type contexts, which indicates that the similar the spatial contexts and type contexts between POIs are, the closer they are in the embedding space. The embedding layout of Fig. 8 (a) and (b) reflects the Third Law of Geography, i.e., the more similar the environment of POIs is, the more similar the features are.

It is obvious that spatial context aware POI embeddings exhibit a relatively more scattered and dispersed distribution. Dual context aware POI embeddings illustrate the similar layout to spatial context aware POI embeddings, but distinct isolation boundaries among different embedding clusters. This aligns with the expected results, as DCA learns POI embeddings by jointly embedding both spatial context and type context, making them more informative and discriminative. Conversely, forcibly pulling closer the embeddings belonging to the same third-level type by injecting type context information may potentially destroy the spatial context semantic meaning of the original POI embeddings.

We randomly sample different rectangular regions from clusters on two third-level types in the POI embedding space derived from DCA. As illustrated in Fig. 9, we extract POIs from the four regions of the POI embedding space, forming corresponding sets of POIs denoted as $\mathscr{S}_a$, $\mathscr{S}_b$, $\mathscr{S}_c$ and $\mathscr{S}_d$. These POI sets serve as a probe to analyze the influence of both type information and spatial context information in the shaping process of POI embeddings. The proportion of second-level types in first-order spatial context (i.e., 1-hop neighbors, as DCA utilizes a one-layer GAT encoder to sense spatial context) of all POIs within each POI set are calculated. The bar chart in Fig. 9 depicts the top 10 second-level types for each POI set in terms of the type statistics of the spatial context. Interestingly, we find that $\mathscr{S}_a$, $\mathscr{S}_b$, $\mathscr{S}_c$ and $\mathscr{S}_d$ are all sets purely of a single type, i.e., the semantics of POI embeddings is consistent with the actual POI type semantics, indicating a powerful encoding capability of DCA for the intrinsic type information of POIs. The spatial context type statistics indicates to a certain extent which context types primarily influence the embeddings of the sampled POI sets, given that the message passing between POIs relies on the spatial context of POIs. The results of spatial context type statistics in Fig. 9 demonstrate a general consistency between the layout of POI embeddings and their spatial context, that $\mathscr{S}_a$, $\mathscr{S}_b$, $\mathscr{S}_c$ and $\mathscr{S}_d$ exhibit similar dominant third-level types both in the ambient of their embedding space and their spatial

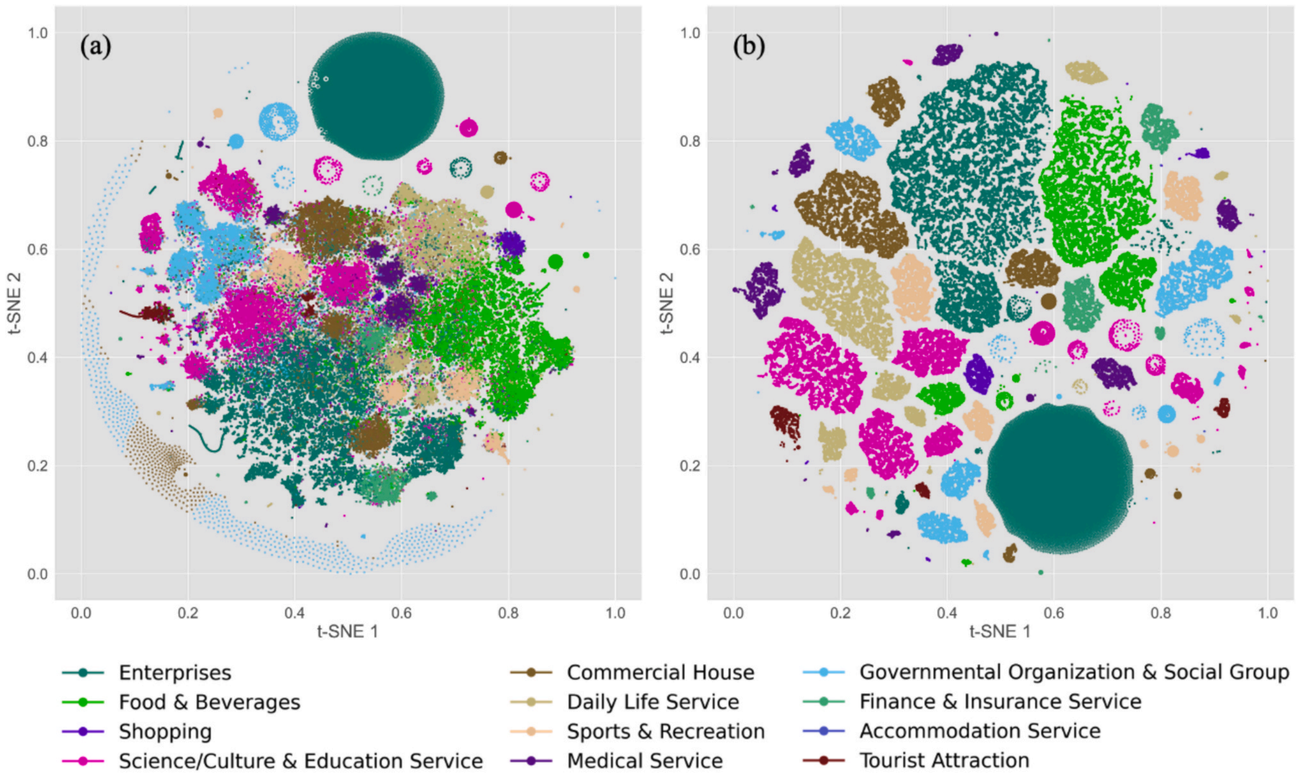


Fig. 8. POI 2-D embedding space by t-SNE is derived from (a) a learned GCN model, (b) a learned DCA model. The POIs are colored by their third-level types. There is no specific meaning to the two axes (i.e., t-SNE 1 and t-SNE 2) of the POI embedding space.

context. In addition, it can be found that the semantic layout and spatial context of some POI embeddings are not entirely consistent. For instance, in the case of $\mathscr{S}_a$ representing the second-level type “Research Institution”, and its spatial context includes a high proportion of second-level type “Residential Area”, but their distributions in the embedding space are not close. A reasonable explanation is that the information in POI embeddings is synchronously determined by spatial context and actual POI attention. Even if there are more edge connections between the two types of POI nodes, DCA can constrain the message passing between them by reducing attention, leading to differences between semantic layout and spatial context.

3.3. Performance of geographic mapping tasks

3.3.1. Implementation details

To finetune the trained DCA for different geographic mapping tasks, the task head is deployed as a shallow MLP network ϕ_z with a hidden layer of 128 neurons (with 1D-batchnorm and ReLU nonlinearity), and a final affine transformation layer. MLP ϕ_z is optimized by minimizing cross-entropy loss for urban function mapping and mean squared error loss for housing price mapping and region popularity mapping, with stochastic gradient descent (SGD) used as the optimizer. Before training, we exclude regions with sparse POIs (less than 10 POIs) from the study, and the remaining data are randomly split into training, validation, and test sets in a ratio of 6:2:2 for model training and evaluation. In the training stage, we set the learning rate to 5×10^{-3}, and the model was trained in minibatch mode with the batch size of 32 for 200 epochs. All geographic mapping experiments (including training, validation, testing processes and dataset random shuffling) are repeated 10 times with unfixed random seeds for reliability.

3.3.2. Comparison with baselines

We report the performance of the DCA and baselines on the test set on three geographic mapping tasks, as shown in Table 1. We observe that LDA and TFIDF perform poorly on these geographic mapping tasks because both solely consider the frequency distribution features of POIs. Contrary to our expectations, TFIDF with naïvely modeling frequency distribution feature did not exhibit the worst performance on urban function mapping. This could be attributed to the fact that the POI features extracted based on frequency in the production of EULUC data (i.e., total number and proportion of each type of POIs within each region) which makes TFIDF easier to capture the correlation between POIs and urban functions. Different from LDA and TFIDF, which are non-representation learning statistical language models, other models take POI spatial context into account and achieve considerable performance. This suggests that POI representations that introduce spatial context information will be more efficient. The performance of Word2Vec is inferior to DCA and GCN, because it lacks the description information of the uniqueness of POIs, and the sequential structure input of Word2Vec is challenged to describe the actual complex spatial context of POIs, resulting in Word2Vec learning POI type embeddings at a rough level.

As expected, DCA achieves the optimal performance, outperforming GCN, which all employ the GNN architecture. On one hand, DCA inherits the powerful encoding capability of GAT, and on the other hand, DCA injects crucial type context information to POI embeddings. In addition, the effectiveness of POI embeddings derived from the DCA can also be supported by combining the visualization results of DCA and GCN (cf. Fig. 8). Additionally, we provide a spatial visualization of geographic mapping performance across all regions in Fig. 10. We did not include zoomed-in views of selected areas, as this might unintentionally introduce selective emphasis. The global perspective intuitively reveals that the prediction errors of all models are spatially uniform, without exhibiting systematic spatial bias.

3.3.3. Parameter sensitivity analysis

We tune the hyperparameters of DCA on a preset hyperparameter space on urban function mapping to investigate the sensitivity of our proposed DCA to its hyperparameters. The two important hyper-

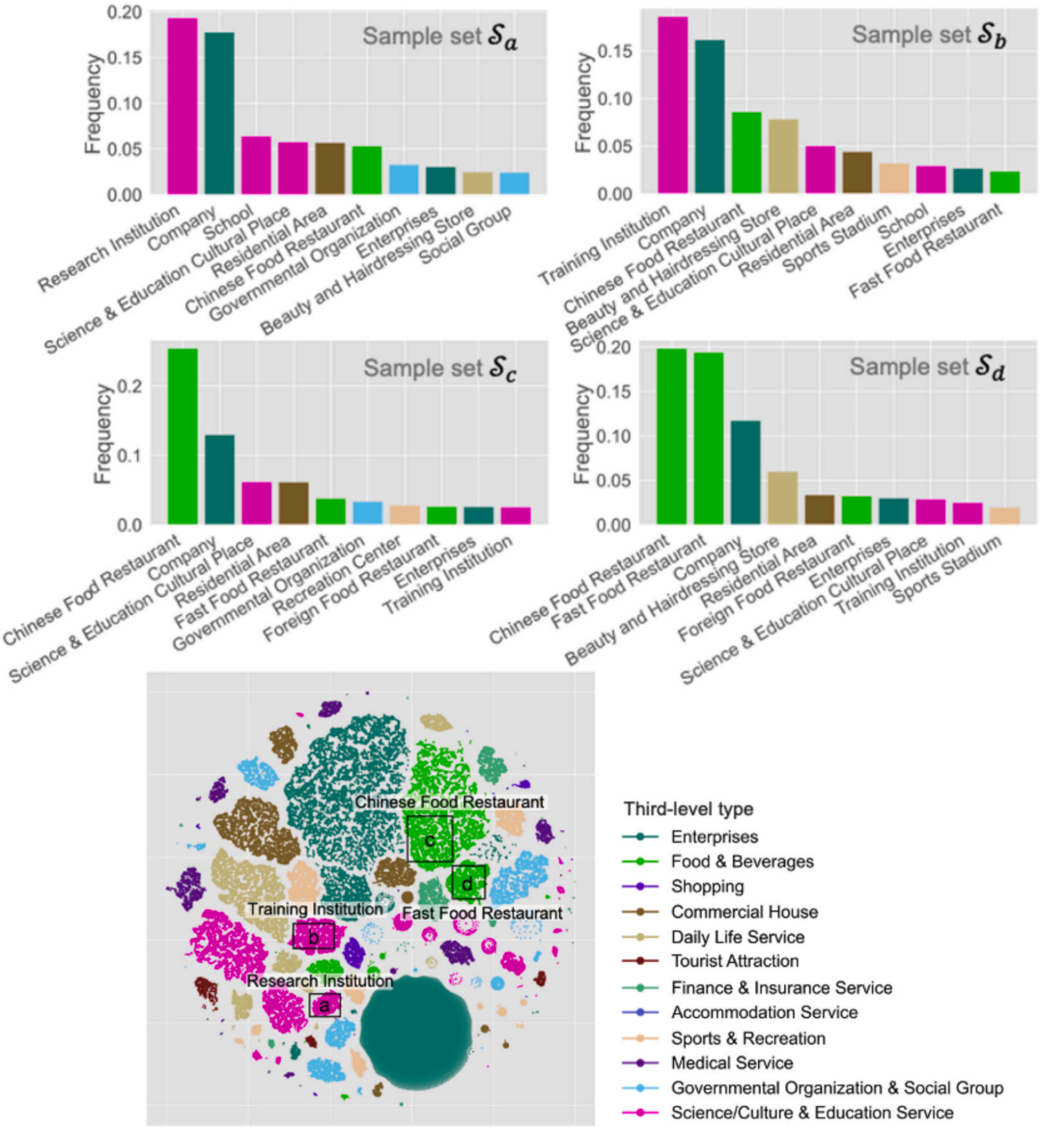


Fig. 9. Local statistics on second-level of the POI embedding space. The horizontal coordinate of the bar chart corresponds to second-level types, and the vertical coordinate indicates the proportion of POI types within its set. The color of the bar chart is rendered according to third-level types.

Table 1
Performance of geographic mapping tasks.

Model	Urban function			Housing price			Region popularity		
	OA↑	Kappa↑	MacF1↑	MAE↓	RMSE↓	R^2↑	MAE↓	RMSE↓	R^2↑
DCA	**0.7268**	**0.5112**	**0.6016**	**13439.34**	**17234.82**	**0.441**	**210.65**	**331.58**	**0.411**
GCN	0.6842	0.4869	0.5804	14421.65	18412.44	0.312	252.08	375.30	0.322
Word2Vec	0.6578	0.4477	0.5567	16017.65	19245.62	0.228	267.41	397.86	0.209
LDA	0.5950	0.3983	0.5191	17044.05	20964.35	0.175	306.90	528.78	0.147
TFIDF	0.6031	0.4124	0.5316	17862.87	22185.63	0.123	451.42	624.24	0.118

parameters of DCA are the dimension of the POI embeddings (d) and the number of attention heads (h) of the GAT encoder. We use a grid search method to tune $d \in \{6, 32, 64\}$ and $h \in \{1, 2, 4\}$ (higher values not reported due to GPU memory limitations). All 9 combinations in the Cartesian product of the two parameter sets are tested. To illustrate the sensitivity to each hyperparameter, we report the performance trends by varying one while keeping the other fixed at its optimal value. The results are presented in Fig. 11.

We observe that setting d to 16 results in poor performance, which can be attributed to the dimension of embeddings is too low to express sufficient POI information. It can be found that the marginal returns diminish during the increase of d, and the best performance is achieved when $d = 64$. Regarding h, the best performance is attained with $h = 4$, primarily due to the increased number of attention heads leads to higher expressiveness of the model. In view of this, it is suggesting that a merit of DCA is insensitive to different hyperparameter settings.

3.3.4. *Ablation study*

We conduct ablation experiments over a number of key components of DCA in order to investigate their impacts on model performance. We

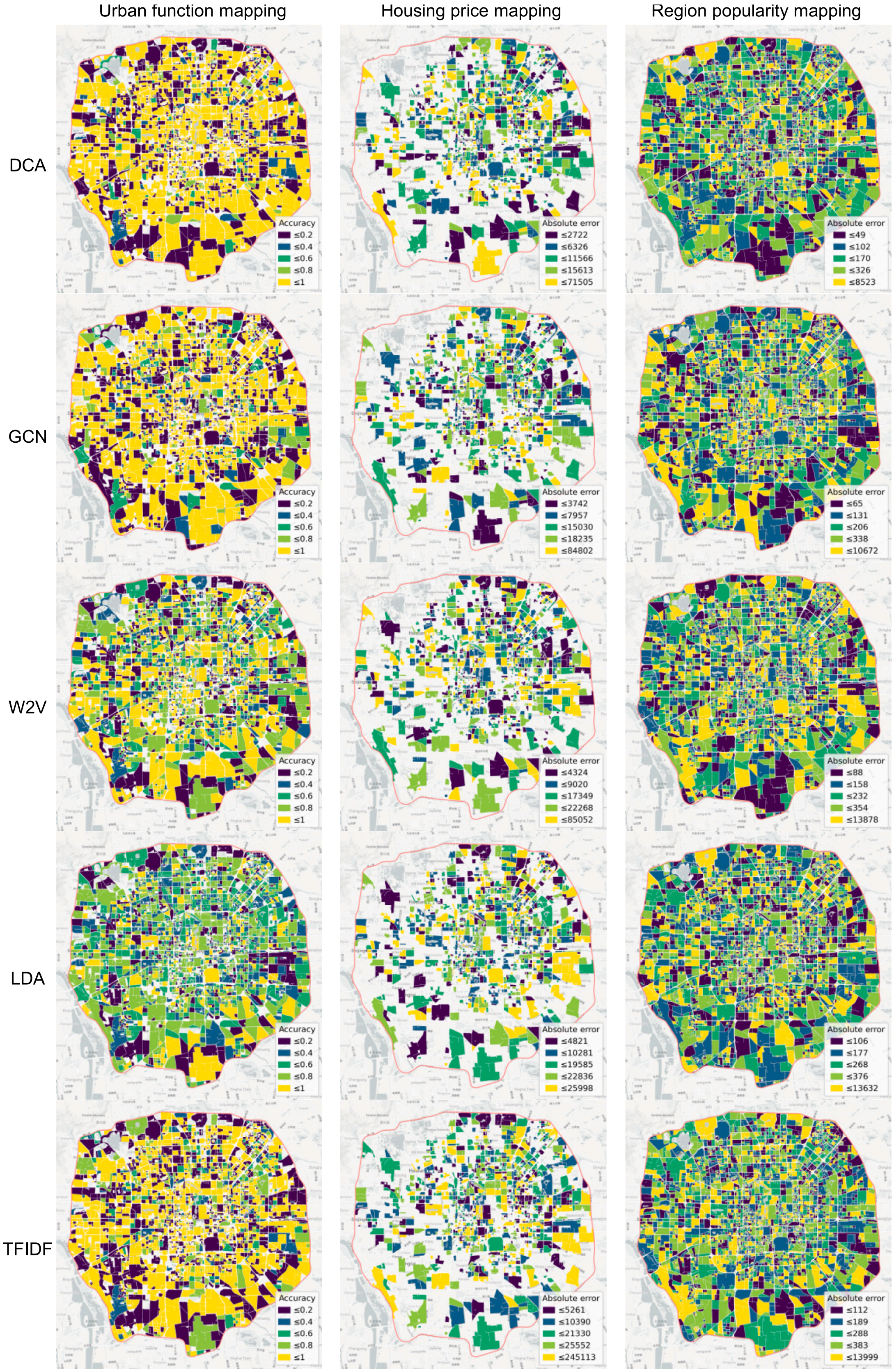


Fig. 10. The spatial distribution of geographic mapping performance.

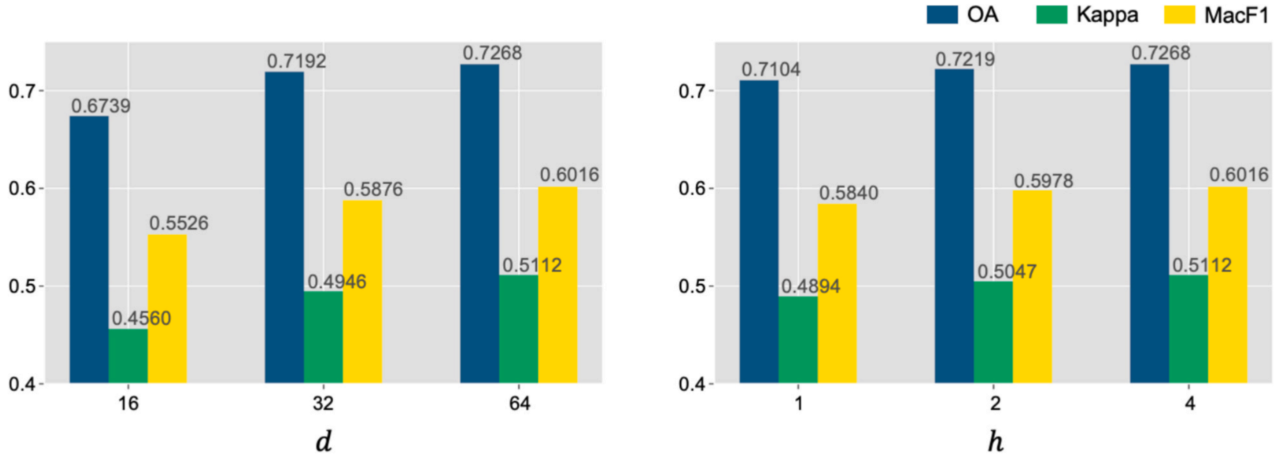


Fig. 11. Parameter sensitivity analysis results for the POI representation dimension *d*, and the number of attention heads *h*.

employ the following DCA variants.

(1) w/o $\mathscr{L}_{st}$: it drops the second ∼ third-level type infomax component, i.e., drops $\mathscr{L}_{st}$ from $\mathscr{L}$ (Equation (7) by setting the fixed values $\lambda_{st} = 0$, $\lambda_{fs} = 1$ in $\mathscr{L}$ for the whole model training stage (in this case, the GradNorm technique is dropped together).
(2) w/o $\mathscr{L}_{fs}$: it drops the first ∼ second-level type infomax component, i.e., drops $\mathscr{L}_{fs}$ from $\mathscr{L}$ (Equation (7) by setting the fixed values $\lambda_{fs} = 0$, $\lambda_{st} = 1$ in $\mathscr{L}$ for the whole model training stage (in this case, the GradNorm technique is dropped together).
(3) w/o DDW: it drops distance decay weighting, i.e., drops w_{ij} from Equation (2).
(4) DCA-GCN: it replaces GAT encoder with one-layer GCN to access the impact of the GAT encoder.
(5) DCA-add: it fuses POI embeddings on tree type level using element-wise addition operation to result the final POI embeddings, rather than POI embeddings on second-level.
(6) DCA-concat: it fuses POI embeddings on tree type level using feature-wise concatenation operation to result the final POI embeddings, rather than POI embeddings on second-level.

From Table 2, it is evident that the performance of w/o $\mathscr{L}_{st}$ surpasses that of w/o $\mathscr{L}_{fs}$ on different geographic mapping tasks. Due to the bottom-up aggregation strategy, second-level type information is derived from first-level types. A reasonable inference is that the local information supervised by $\mathscr{L}_{fs}$ contributes more significantly to the learning of POI embeddings, while the global information supervised by $\mathscr{L}_{st}$ plays a relatively weaker auxiliary role. The performance difference between DCA and w/o DDW serves to decouple the impact of the distance decay effect on model performance, providing empirical evidence that incorporating this effect can regularize the graph attention mechanism to perform geographically meaningful message passing between POIs, thereby facilitating the learning of valuable geographic information. The observed performance decline in DCA-GCN compared to DCA highlights the importance of modeling heterogeneity in semantic interactions, as graph attention-based message passing in DCA allows for modeling of varying semantic significance of POIs within their spatial context, whereas graph convolution-based message passing in DCA-GCN assumes uniform importance across POIs within the spatial context. However, DCA-GCN still outperforms the baseline GCN on these geographic mapping tasks (cf. Table 1), indicating that the performance of DCA is not solely dependent on the spatial context, and type context information remains crucial. DCA-add and DCA-concat is slightly worse than DCA, which can be reasonably attribute to the type context-aware loss function of DCA enables POI embeddings on second-level to convey information of first- and third-level, so that the feature fusion makes the information of POI embedding too redundant. All ablations show inferior performance compared to DCA, which indicates each component plays a pivotal role to the superiority of our DCA model.

4. Discussion

4.1. POI semantic interaction analysis

Since the strength of message passing between POIs in DCA depends on the multi-head self-attention, we can reveal the semantic interaction process of POIs through the attention mechanism. The attention weights serve to quantify the strength of semantic interaction between POIs. To be specific, the attention weights are extracted by averaging all GAT attention heads, and retrieved to be used as POI semantic interactions.

We visualize the attention between all POIs in Fig. 12. Fig. 12 (a) renders the semantic interaction map of POIs (spatial visualization of the attention between POIs, which is distributed along the edges of POI graph), and it can be observed that the semantic interaction among POIs is evenly distributed in space. Note that due to the statistical attention of self-loops (i.e., attention to oneself), the origin–destination points overlap in space and are not displayed. In Fig. 12 (b), we find that the attention strength among different POIs follows a long-tail distribution,

Table 2
Results of the ablation experiments on geographic mapping tasks.

Model	Urban function			Housing price			Region popularity		
	OA↑	Kappa↑	MacF1↑	MAE↓	RMSE↓	R^2↑	MAE↓	RMSE↓	R^2↑
DCA	**0.7268**	**0.5112**	**0.6016**	**13439.34**	**17234.82**	**0.441**	**210.65**	**331.58**	**0.411**
w/o $\mathscr{L}_{st}$	0.7137	0.4947	0.5892	13774.44	17813.43	0.405	218.73	347.52	0.401
w/o $\mathscr{L}_{fs}$	0.7102	0.4898	0.5830	13852.84	17987.82	0.397	227.48	366.14	0.389
w/o DDW	0.7114	0.5026	0.5893	13536.34	17478.26	0.429	223.54	354.32	0.390
DCA-GCN	0.7149	0.4997	0.5895	13606.71	17583.60	0.425	217.61	344.68	0.404
DCA-add	0.7044	0.4805	0.5803	13633.79	17604.21	0.418	226.97	359.07	0.385
DCA-concat	0.7103	0.4932	0.5902	13551.62	17438.04	0.436	220.03	346.52	0.396

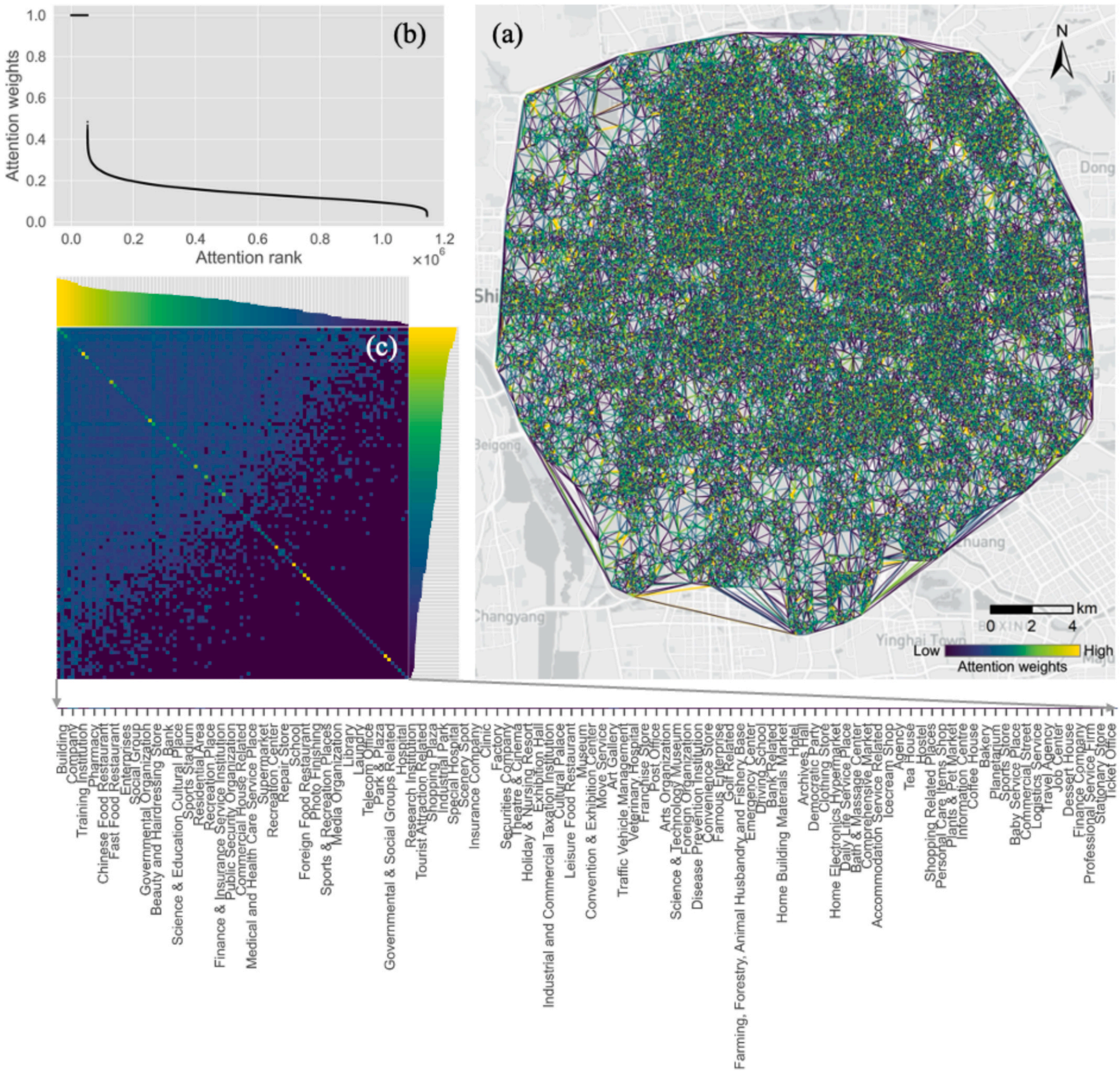


Fig. 12. The POI attention distribution. (a) POI semantic interaction map: the spatial distribution of attention between individual POIs along the DT network, (b) the statistical distribution of POI attention, (c) the statistical distribution of attention on second-level, the heat map represents pair-wise attention between POI types (the tick labels of the two axes are the same), and the bar chart represents the total attention obtained by each type.

indicating that a small number of edges obtained a large amount of attention. In other words, only a small portion of POIs exhibit relatively significant role in message passing, while most POIs have limited semantic interaction.

We observe differences in attention weights across types on second-level, as shown in Fig. 12 (c). The heatmap represents the mean attention between POI types, while the bar chart shows the sum attention between each type and all types. The heatmap illustrates that some types are more likely to semantically interact with certain types. The diagonal of the heatmap indicates that all POI types are more focused on their own types, which means that POI tend to retain their own type information during semantic interaction. The bar chart explains that certain POI types aggregate more attention in the message passing process along the POI graph, indicating a tendency to semantically interact with their spatial context and disseminate their own information and receive contextual information. This suggests that POI attention is related to each unique POI and its type, rather than being relevant to spatial distribution.

4.2. Limitations and potential

There are some limitations to the application and designing of DCA model. Although the DCA model adopts the homogeneous graph, a well-accepted approach in the literature, to establish the spatial context, this approach may impose certain limitations on the spatial context awareness of POI representations. Since this study focuses on sensing the dual context of POIs jointly, DCA adopts a one-layer GAT that solely considers first-order spatial context. While deeper GNNs can capture higher-order spatial dependencies, they also introduce more trainable parameters, making it challenging to decouple the performance changes brought by expanded spatial context from those resulting from increased model complexity. In light of this, a more in-depth exploration of high-order spatial contexts for POIs is a worthwhile direction for future work. As a natural choice, aggregation function (Equation (4) applies the average pooling to the DCA forward propagation, whereas any sophisticated aggregation function conforms to permutation invariance can serve as an alternative to average pooling aggregation function, e.g., self-attention mechanism (Vaswani et al., 2017), leading to flexible model variants. Furthermore, specific limitations with respect to type infomax deserve attention. POI data from different sources often adopt inconsistent type taxonomies, leading to variations in type granularity and hierarchical structure. This could limit the effectiveness of type infomax in learning hierarchical type information. DCA employ random negative sampling in the current implementation of type infomax. While it is simple and effective, it may result in easy negatives that provide limited learning signals. A promising direction for improvement lies in introducing hard negative sampling, e.g., selecting semantically similar yet type hierarchically inconsistent POIs, as the fact of that using hard negatives can benefit contrastive representation learning (Robinson et al., 2021). However, this also raises challenges in defining the boundary between hard negatives and positives, especially in POI type taxonomy where semantic similarity and hierarchical type structures may overlap or be ambiguous, warranting further investigation to improve the model. It is promising to explore more advanced DCA variants by overcoming the above limitations in future work.

On the other hand, DCA has significant potential for different application scenarios beyond the scope of this work. Firstly, DCA learns POI representation embeddings does not involve the modifiable area unit problem. The learned POI embeddings can be mapped to urban regions of different scales to adapt to corresponding tasks, e.g., blocks, traffic analysis zones, census units, and building footprints. Moreover, since DCA is a self-supervised representation learning model, it is not bound by the supervision signal of any particular task that leads to generic POI representations. Therefore, we believe that DCA can be well-suited for a broader range of downstream tasks, such as POI recommendation (Cui et al., 2022), location matching (Mousset et al., 2020), population density mapping (Huang et al., 2023), crime prediction (Zhang et al., 2023c), and traffic speed forecasting (Zhang et al., 2023d). Considering that DCA learns representation of each unique POI, it is necessary to consider whether there is some explicit correlation between the task and POIs before further applying DCA to various downstream tasks. In addition, if the type taxonomy of some POI data does not define three levels, but two levels (e.g., POIs come from Baidu Map with only the first level and second level), the network architecture and loss function of DCA can be simply adjusted to adapt to the two levels. Another possible solution is to try POI name as the bottom type level and apply the DCA model. Additionally, as DCA adopts inductive learning GAT encoder. Compared with GCN, it can be easily generalized and applied to new POI data (POI graphs of different cities or local regions), as long as the POIs share the same type taxonomy. In light of the above, the proposed DCA model holds substantial potential for widespread applications. Given the ubiquity and accessibility of POI data, DCA exhibits promising prospects which could be extended or transferred to a broader spectrum of POI-related scenarios.

5. Conclusions

In this study, we proposed a novel POI representation learning approach DCA, to learn POI representations in a self-supervised manner. To the best of our knowledge, DCA pioneers in learning POI representations by jointly embedding the external spatial structure and the internal type hierarchy of POIs. We evaluated the DCA on three typical geographic mapping tasks, and the results outperforms all baseline models. We also visually analyzed the layout of POI embeddings, and conducted parameter sensitivity analysis and ablation analysis for DCA. The experimental results all demonstrate the robustness and superiority of DCA. This study provides a new insight to mine deep information from static POIs, and enhance our understanding of urban system.

In the future, we consider exploring the capabilities of DCA across various geographic mapping with different city scales. We believe that the POI graph plays a crucial role in facilitating message passing between POIs. Introducing blocking effect which from tangible or intangible geographical objects (e.g., rivers, roads, and administrative boundary) to constrain POI graph construction will enhance spatial context awareness of POIs. Furthermore, multi-modal data fusion will be considered so that the POI embedding is not limited to endogenetic spatial and type context awareness, such as incorporating temporal context awareness, leading to more comprehensive POI embeddings for describing urban information.

CRediT authorship contribution statement

Quan Qin: Writing – original draft, Methodology, Conceptualization. **Tinghua Ai:** Writing – review & editing, Supervision, Funding acquisition. **Shishuo Xu:** Writing – review & editing. **Yan Zhang:** Writing – review & editing. **Weiming Huang:** Writing – review & editing. **Mingyi Du:** Writing – review & editing. **Songnian Li:** Writing – review & editing.

Declaration of competing interest

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

Acknowledgement

The authors thank the editors and anonymous reviewers for their insightful comments and constructive suggestions. This work was supported by the National Natural Science Foundation of China [No.42394065].

Data availability

The authors do not have permission to share data.

References

Andrade, R., Alves, A., Bento, C., 2020. POI mining for land use classification: a case study. ISPRS Int. J. Geo Inf. 9 (9), 1–23. https://doi.org/10.3390/ijgi9090493.

Bai, L., Huang, W., Zhang, X., Du, S., Cong, G., Wang, H., Liu, B., 2023. Geographic mapping with unsupervised multi-modal representation learning from VHR images and POIs. ISPRS J. Photogramm. Remote Sens. 201, 193–208. https://doi.org/10.1016/j.isprsjprs.2023.05.006.

Cao, J., Wang, X., Chen, G., Tu, W., Shen, X., Zhao, T., Chen, J., Li, Q., 2025. Disentangling the hourly dynamics of mixed urban function : a multimodal fusion perspective using dynamic graphs. Inf. Fusion 117, 102832. https://doi.org/10.1016/j.inffus.2024.102832.

Chen, Z., Badrinarayanan, V., Lee, C.Y., Rabinovich, A., 2018. GradNorm: gradient normalization for adaptive loss balancing in deep multitask networks. Int. Conf. on Mach. Learn. 2, 1240–1251.

Chen, Y., Huang, W., Zhao, K., Jiang, Y., Cong, G., 2025. Self-supervised representation learning for geospatial objects: A survey. Information Fusion 103265.

Cui, Y., Sun, H., Zhao, Y., Yin, H., Zheng, K., 2022. Sequential-knowledge-aware next POI recommendation: a meta-learning approach. ACM Trans. Inf. Syst. 40 (2). https://doi.org/10.1145/3460198.

Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., Hu, S., Chen, Y., Chan, C.M., Chen, W., Yi, J., Zhao, W., Wang, X., Liu, Z., Zheng, H.T., Chen, J., Liu, Y., Tang, J., Li, J., Sun, M., 2023. Parameter-efficient fine-tuning of large-scale pre-trained language models. Nat. Mach. Intell. 5 (3), 220–235. https://doi.org/10.1038/s42256-023-00626-4.

Gong, P., Chen, B., Li, X., Liu, H., Wang, J., Bai, Y., Chen, J., Chen, X., Fang, L., Feng, S., Feng, Y., Gong, Y., Gu, H., Huang, H., Huang, X., Jiao, H., Kang, Y., Lei, G., Li, A., Xu, B., 2020. Mapping essential urban land use categories in China (EULUC-China): preliminary results for 2018. Sci. Bulletin 65 (3), 182–187. https://doi.org/10.1016/j.scib.2019.12.007.

Gu, T., Zhao, H., Yue, L., Guo, J., Cui, Q., Tang, J., Gong, Z., Zhao, P., 2025. Attribution analysis of urban social resilience differences under rainstorm disaster impact: insights from interpretable spatial machine learning framework. Sustain. Cities Soc. 118, 106029. https://doi.org/10.1016/j.scs.2024.106029.

Hou, C., Zhang, F., Kang, Y., Gao, S., Li, Y., Duarte, F., Li, S., 2024. Transferred bias uncovers the balance between the development of physical and socioeconomic environments of cities. Ann. Am. Assoc. Geogr. 115 (1), 148–166. https://doi.org/10.1080/24694452.2024.2412173.

Hu, S., He, Z., Wu, L., Yin, L., Xu, Y., Cui, H., 2020. A framework for extracting urban functional regions based on multiprototype word embeddings using points-of-interest data. Comput. Environ. Urban Syst. 80, 101442. https://doi.org/10.1016/j.compenvurbsys.2019.101442.

Huang, W., Cui, L., Chen, M., Zhang, D., Yao, Y., 2022. Estimating urban functional distributions with semantics preserved POI embedding. Int. J. Geogr. Inf. Sci. 1–26. https://doi.org/10.1080/13658816.2022.2040510.

Huang, W., Zhang, D., Mai, G., Guo, X., Cui, L., 2023. Learning urban region representations with POIs and hierarchical graph infomax. ISPRS J. Photogramm. Remote Sens. 196, 134–145. https://doi.org/10.1016/j.isprsjprs.2022.11.021.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., Bhaduri, B., 2020. GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. Int. J. Geogr. Inf. Sci. 34 (4), 625–636. https://doi.org/10.1080/13658816.2019.1684500.

Kong, B., Ai, T., Zou, X., Yan, X., Yang, M., 2024. A graph-based neural network approach to integrate multi-source data for urban building function classification. Comput. Environ. Urban Syst. 110, 102094. https://doi.org/10.1016/j.compenvurbsys.2024.102094.

Li, Y., Huang, W., Cong, G., Wang, H., Wang, Z., 2023. Urban region representation learning with OpenStreetMap building footprints. In: The 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 1363–1373. https://doi.org/10.1145/3580305.3599538.

Li, Z., Huang, W., Zhao, K., Yang, M., Gong, Y., Chen, M., 2024. Urban region embedding via multi-view contrastive prediction. Proce. AAAI Conf. on Artificial Intelligence 38 (8), 8724–8732. https://doi.org/10.1609/aaai.v38i8.28718.

Liu, K., Qiu, P., Gao, S., Lu, F., Jiang, J., Yin, L., 2020. Investigating urban metro stations as cognitive places in cities using points of interest. Cities 97, 102561. https://doi.org/10.1016/j.cities.2019.102561.

Liu, P., Biljecki, F., 2022. A review of spatially-explicit GeoAI applications in Urban Geography. Int. J. Appl. Earth Obs. Geoinf. 112 (July), 102936. https://doi.org/10.1016/j.jag.2022.102936.

Liu, X., He, J., Yao, Y., Zhang, J., Liang, H., Wang, H., Hong, Y., 2017. Classifying urban land use by integrating remote sensing and social media data. Int. J. Geogr. Inf. Sci. 31 (8), 1675–1696. https://doi.org/10.1080/13658816.2017.1324976.

van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (11), 2579–2605.

Mai, G., Yao, X., Xie, Y., Rao, J., Li, H., Zhu, Q., Li, Z., Lao, N., 2024. SRL : towards a general-purpose framework for spatial representation learning. In: Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems, pp. 465–468. https://doi.org/10.1145/3678717.3691246.

Manvi, R., Khanna, S., Mai, G., Burke, M., Lobell, D., Ermon, S., 2024. GeoLLM: extracting geospatial knowledge from large language models. Int. Conference on Learning Representations.

Mousset, P., Pitarch, Y., Tamine, L., 2020. End-to-End Neural matching for semantic location prediction of tweets. ACM Trans. Inf. Syst. 39 (1). https://doi.org/10.1145/3415149.

Niu, H., Silva, E.A., 2021. Delineating urban functional use from points of interest data with neural network embedding: a case study in Greater London. Comput. Environ. Urban Syst. 88, 101651. https://doi.org/10.1016/j.compenvurbsys.2021.101651.

Psyllidis, A., Gao, S., Hu, Y., Kim, E.-K., McKenzie, G., Purves, R., Yuan, M., Andris, C., 2022. Points Of Interest (POI): a commentary on the state of the art, challenges, and prospects for the future. Comput. Urban Sci. 2 (1), 20. https://doi.org/10.1007/s43762-022-00047-w.

Qin, Q., Xu, S., Du, M., Li, S., 2022a. Identifying urban functional zones by capturing multi-spatial distribution patterns of points of interest. Int. J. Digital Earth 15 (1), 2468–2494. https://doi.org/10.1080/17538947.2022.2160841.

Qin, Q., Xu, S., Du, M., Li, S., 2022b. Urban functional zone identification by considering the heterogeneous distribution of points of interests. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 5 (4), 83–90. https://doi.org/10.5194/isprs-annals-V-4-2022-83-2022.

Robinson, J., Chuang, C.Y., Sra, S., Jegelka, S., 2021. Contrastive learning with hard negative samples. Int. Conf. on Learn. Representations 1–28.

Su, Y., Zhong, Y., Zhu, Q., Zhao, J., 2021. Urban scene understanding based on semantic and socioeconomic features: from high-resolution remote sensing imagery to multi-source geographic datasets. ISPRS J. Photogramm. Remote Sens. 179 (January), 50–65. https://doi.org/10.1016/j.isprsjprs.2021.07.003.

Veličković, P., Casanova, A., Liò, P., Cucurull, G., Romero, A., Bengio, Y., 2018. Graph attention networks. Int. Conf. on Learn. Representations. https://doi.org/10.1007/978-3-031-01587-8_7.

Veličković, P., Fedus, W., Hamilton, W.L., Bengio, Y., Liò, P., Devon Hjelm, R., 2019. Deep graph infomax. 7th International Conference on Learning Representations.

Wang, J., Biljecki, F., 2022. Unsupervised machine learning in urban studies : a systematic review of applications. Cities 129, 103925. https://doi.org/10.1016/j.cities.2022.103925.

Wang, X., Cheng, T., Law, S., Zeng, Z., Yin, L., Liu, J., 2024. Multimodal contrastive learning of urban space representations from POI data. Comput. Environ. Urban Syst. 120 (April), 102299. https://doi.org/10.1016/j.compenvurbsys.2025.102299.

Xiao, C., Zhou, J., Xiao, Y., Huang, J., Xiong, H., 2024. ReFound : crafting a foundation model for urban region understanding upon language and visual foundations. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 3527–3538. https://doi.org/10.1145/3637528.3671992.

Xu, Y., Zhou, B., Jin, S., Xie, X., Chen, Z., Hu, S., He, N., 2022. A framework for urban land use classification by integrating the spatial context of points of interest and graph convolutional neural network method. Comput. Environ. Urban Syst. 95, 101807. https://doi.org/10.1016/j.compenvurbsys.2022.101807.

Yan, B., Mai, G., Janowicz, K., Gao, S., 2017. From ITDL to Place2Vec–reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In: Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 1–10. https://doi.org/10.1145/3139958.3140054.

Yan, X., Ai, T., Yang, M., Yin, H., 2019. A graph convolutional neural network for classification of building patterns using spatial vector data. ISPRS J. Photogramm. Remote Sens. 150, 259–273. https://doi.org/10.1016/j.isprsjprs.2019.02.010.

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., Mai, K., 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. Int. J. Geogr. Inf. Sci. 31 (4), 825–848. https://doi.org/10.1080/13658816.2016.1244608.

Yao, Y., Zhu, Q., Guo, Z., Huang, W., Zhang, Y., Yan, X., 2023. Unsupervised land-use change detection using multi-temporal POI embedding. Int. J. Geogr. Inf. Sci. 1–24. https://doi.org/10.1080/13658816.2023.2257262.

Zhang, D., Xu, R., Huang, W., Zhao, K., Chen, M., 2023a. Towards an integrated view of semantic annotation for pois with spatial and textual information. Int. Joint Conf. on Artificial Intelligence 2441–2449. https://doi.org/10.24963/ijcai.2023/271.

Zhang, L., Long, C., Cong, G., 2023b. Region embedding with intra and inter-view contrastive learning. IEEE Trans. Knowl. Data Eng. 35 (9), 9031–9036. https://doi.org/10.1109/TKDE.2022.3220874.

Zhang, P., Yang, M., Wang, Y., Yang, T., Yu, H., Yan, X., 2024a. Integrating metro passenger flow data to improve the classification of urban functional regions using a heterogeneous graph neural network. Int. J. Digital Earth 17 (1), 2443468. https://doi.org/10.1080/17538947.2024.2443468.

Zhang, Q., Huang, C., Xia, L., Wang, Z., Yiu, S., Han, R., 2023c. Spatial-temporal graph learning with adversarial contrastive adaptation. Int. Conf. on Mach. Learn. 202, 41151–41163.

Zhang, Y., Chen, Z., Zheng, X., Chen, N., Wang, Y., 2021. Extracting the location of flooding events in urban systems and analyzing the semantic risk using social sensing data. J. Hydrol. 603, 127053. https://doi.org/10.1016/j.jhydrol.2021.127053.

Zhang, Y., Huang, W., Yao, Y., Gao, S., Cui, L., Yan, Z., 2024b. Urban region representation learning with human trajectories: a multi-view approach incorporating transition, spatial, and temporal perspectives. Giscience & Remote Sensing 61 (1). https://doi.org/10.1080/15481603.2024.2387392.

Zhang, Y., Zhao, T., Gao, S., Raubal, M., 2023d. Incorporating multimodal context information into traffic speed forecasting through graph deep learning. Int. J. Geogr. Inf. Sci. 1–27. https://doi.org/10.1080/13658816.2023.2234959.

Zhao, Y., Qi, J., Trisedya, B.D., Su, Y., Zhang, R., Ren, H., 2023. Learning region similarities via graph-based deep metric learning. IEEE Trans. Knowl. Data Eng. 1–14. https://doi.org/10.1109/TKDE.2023.3253802.