

Article

An Integrated Approach to Identify Functional Areas for Bicycle Use with Spatial–Temporal Information: A Case Study of Seoul, Republic of Korea

Jiwon Lee  and Jiyoung Kim 

Social Eco Tech Institute, Konkuk University, Seoul 05029, Republic of Korea; vert802@konkuk.ac.kr

* Correspondence: elliekim@konkuk.ac.kr

Abstract

Identifying urban functional areas increasingly relies on data-driven approaches that utilize multimodal spatial information. There is a growing focus on purpose-oriented functional area identification with greater policy relevance. This paper proposes a data-driven methodology to identify functional areas from the perspective of bicycle users. To achieve this, line-based road network units were defined around bicycle stations, and spatial–temporal data such as Origin–Destination flows and Point of Interest information were semantically integrated to delineate functional areas. An experiment was conducted on 2628 public bicycle stations in Seoul, Republic of Korea, for May 2022, and a total of five functional areas were identified via a Co-Matrix Factorization-based fusion approach. Additionally, the proposed method was validated through visual evaluation and comparison with actual bicycle usage data. The results demonstrate that by simultaneously incorporating spatial–temporal information and latent connectivity, this approach identifies bicycle-friendly areas, even with low observed usage, highlighting its potential for policy applications.

Keywords: urban functional area; bicycle-sharing systems; semantic-based data fusion; co-matrix factorization; POI embedding; dynamic time warping



Academic Editor: Teodoro Semeraro

Received: 30 August 2025

Revised: 14 October 2025

Accepted: 15 October 2025

Published: 16 October 2025

Citation: Lee, J.; Kim, J. An Integrated Approach to Identify Functional Areas for Bicycle Use with Spatial–Temporal Information: A Case Study of Seoul, Republic of Korea. *Land* **2025**, *14*, 2069. <https://doi.org/10.3390/land14102069>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background and Objectives

The emergence of diverse geospatial big data and advancements in analysis technologies have accelerated research on urban functional area extraction. Approaches have shifted from direct field surveys [1] to data-based automated classification, emphasizing the importance of appropriate data and methodologies. These choices critically influence the accuracy of functional area identification and its applicability. Various studies have delineated urban functional areas using multiple data sources: remote sensing imagery that reveals visual spatial patterns [2,3], point of interest (POI) or social media data that capture socioeconomic characteristics [1,4–7], and mobility data reflecting intra-urban movement flows [1,8–10]. Previous studies have identified urban functional areas from a generalized perspective rather than for specific purposes. Studies using social media or mobile phone data often focus on representativeness in capturing broad urban dynamics instead of extracting tailored functions. However, such data can introduce regional and user-based biases that may distort the interpretation of functional areas [11].

However, for functional area extraction for a specific objective, the selected data do not necessarily need to have representativeness for an entire city. Functional areas can

be extracted by utilizing data or methodologies that can reflect the defined objective. For example, when extracting functional areas for the elderly, information on other age groups is not needed; instead, the analysis should be based on elderly mobility data and data on their preferred and frequently visited places. Cities are mostly spaces where diverse POIs are clustered, which makes it difficult to define an area with a single function. For instance, in the case of a station catchment area with a subway station and a department store nearby, it is difficult to unify the area's function into either transportation or shopping. However, when a precise target is defined, the area's meaning becomes clearer. For public transportation users, the area will be perceived as closer to a transportation function, whereas for private car users, it will be perceived as closer to a shopping function. Therefore, clarifying the purpose of functional area extraction and selecting data and methodologies suitable for it are crucial. This can be considered not just a simple classification but a process that enhances practical policy applicability and interpretability.

Accordingly, this paper proposes a functional area extraction methodology specialized for bicycle users. As an eco-friendly mode of transportation, bicycles have been highlighted as an alternative that can solve environmental and traffic problems. Particularly, they are very important as a means of connecting the first and last miles in terms of public transportation. This study aimed to identify functional areas by reflecting spatial and temporal characteristics based on actual bicycle usage data. Furthermore, by holistically utilizing multidimensional data, it seeks to clarify the unique spatial and temporal dynamics of bicycle use that have been overlooked in previous studies.

1.2. Related Work

With the rapid increase in data volume due to technological advancements, research on functional area extraction has also become active in recent years. Table 1 compares and summarizes such studies in terms of the extraction purpose, spatial unit, data, and methodology. As mentioned earlier, related works have primarily aimed at classifying the overall functions of cities, utilizing various datasets that reflect the representativeness of urban areas. Even when using data from a specific mode of transportation, analyses were conducted under the assumption that the dataset could sufficiently explain a region's mobility flows. For example, Chang et al. [12] and Zhao et al. [13] classified functional areas using bicycle usage data, judging that bicycle data, which are densely distributed across fine spatial units throughout the city, are useful for analyzing the city's fundamental spatial-temporal roles [13]. However, in Lee et al. [14], functional areas were derived for the specific research objective of trip-purpose inference. In that study, the functional area results were used only as input variables for the overall model, which makes it somewhat simplified. Therefore, to propose a functional area extraction methodology specialized for bicycle users, herein we review previous works in terms of spatial unit, data, and methodology, aiming to derive an analytical approach that can most appropriately reflect bicycle usage behavior.

Table 1. Analysis of related works on functional area identification.

Ref.	Purpose	Spatial Unit	Data	Methodology
Zhai et al. [1]	General	Administrative district	POI	Place2vec, K-means
Chang et al. [12]	General	Grid	Bicycle GPS, POI	Topic-modeling
Qian et al. [15]	General	Road and river block	RS image, Taxi GPS	MLC-ResNets, YOLO v3, K-means

Table 1. *Cont.*

Ref.	Purpose	Spatial Unit	Data	Methodology
Zhao et al. [13]	General	Parking spot coverage	Bicycle OD, POI	DNN
Hu et al. [16]	Fine-grained UFZ	Road segment	Taxi GPS, POI	Word2vec, GCNN
Lee et al. [14]	Trip-purpose inference	Parking spot coverage (Buffer zone)	POI	POI-type embedding, K-means
Niu & Silva [17]	General	Road block	POI	HSC, Doc2vec
Deng & He [18]	Fine-grained UFZ	Building-level polygon (Voronoi diagram)	Building, POI	LDA, SVM
Jing et al. [19]	Fine-grained UFZ	Hierarchical Grid	POI	LDA, Kernel Density
Jing et al. [20]	General	Road block	Taxi GPS, POI	CCMF, Spectral clustering
Qin et al. [21]	General	Road block	POI	Word2vec, RF
Yang et al. [22]	Fine-grained UFZ	Building-level polygon	Building, POI	Stacking Ensemble
Liu et al. [23]	General	Road block	Taxi GPS, POI	CA-RFM
Luo et al. [24]	General	Grid	POI	Kernel density, K-star
Wang & Feng [25]	General	Road block	RS image, Building, POI, Social media	VGG16, BERT, Random Forest
Zhang et al. [10]	General	Road block	Smart card OD, Building, POI	HGNN
Zhang et al. [26]	General	Land-use polygons (OSM)	RS image, POI, Smart card OD	TriNet

First, the spatial unit definition is a prerequisite for function classification of continuous and complex urban spaces. Accordingly, among studies that extract general urban functional areas, those focusing on extracting fine-grained Urban Functional Zones (UFZs), such as at the building level, are increasing. The spatial units mainly used in previous studies include administrative boundaries [1], block units based on road boundaries [10,15,17,20,21,23,25], grids [12,19,24], Voronoi polygons [18], and road networks [16], each with different strengths and weaknesses according to the research objective and data characteristics. For example, administrative boundaries have the limitation of not sufficiently reflecting actual mobility behaviors or spatial interactions; however, they offer practical advantages in that most policy documents and statistical data are collected and provided based on administrative units. Meanwhile, regular and standardized spatial units such as grids or Voronoi polygons are advantageous for processing large-scale spatial data, with high computational efficiency and scalability. However, these units also have the limitation of not sufficiently reflecting actual activity spaces or the spatial semantics of data. In studies utilizing bicycle data, spatial unit definitions have generally relied on grids [12] or station-based buffers and coverage areas [13,14], instead of on standardized units such as administrative districts. By contrast, the current study seeks to define spatial units that reflect the actual road network structure where movement occurs. While most existing urban-related research employed polygon-shaped spatial units, actual human movement often takes place along linear paths such as streets and roads. Consequently, the importance of line-based spatial units has recently increased. For instance, Wu and Li [27] proposed a “Hotstreet” analysis that extended conventional hotspot analysis to the street level for crime detection in London. This approach more precisely reflected incident locations and produced superior results in terms of analytical accuracy and visualization.

Although some studies on functional area extraction have applied road-network-based spatial units using trajectory data of transportation modes, most still remain at the block-unit level based on road boundaries. However, Hu et al. [16] derived functional areas at the road-segment level using taxi GPS data. Therefore, by applying line-based spatial units that directly reflect the bicycle road network, this study aims to complement conventional polygon-centered approaches and establish an analytical basis that more precisely captures bicycle usage behaviors.

Second, appropriate data selection for functional area extraction is a core element in clearly defining the functional area meaning. In early studies, many cases derived functional areas mainly from single data, particularly static land-use information such as POIs. However, for more accurate functional area extraction, the need of not only static data, such as POIs, remote sensing images, and building information, but also dynamic data, such as GPS, smart cards, and social media, has been increasingly emphasized. Accordingly, research utilizing both types of data concurrently has been increasing. Nevertheless, cases in which static and dynamic data have been actively fused and applied in analysis are still lacking. In most studies, a single dataset is used to extract features, and other data are often used only as supplementary material for validation. For example, Zhao et al. [13] identified land use by utilizing bicycle origin-destination (OD) data; however, in the feature extraction stage, only bicycle movement data were used, and POI data were used only for producing label information for training. Similarly, Hu et al. [16] extracted features at the road-segment level using taxi GPS data and predicted the function of each segment through a graph convolutional neural network (GCNN) model; however, as in other cases, POI data were used only as labeling material. In Chang et al. [12], bicycle pick-up and return data were aggregated by time period and topic modeling was applied to derive latent functional distributions and functional area identification, but POI data were used only when naming the clusters. By contrast, the current study has significance in that POI data and bicycle movement patterns were considered together in the cluster-naming process, meaning that regional functions were identified not merely by relying on land-use information but by reflecting actual mobility behavior. Recently, some studies have attempted to actively integrate static and dynamic data, but cases focusing on bicycle data are still lacking. Such data fusion is important not only in terms of academic significance but also in terms of policy and practical applicability. For example, Liu et al. [28] compared and analyzed static land-use and dynamic bicycle movement data to confirm inconsistencies between planning and actual use, and on this basis suggested directions for urban planning and transportation infrastructure improvement. Therefore, the current study seeks to derive functional areas that reflect spatial-temporal characteristics combining static and dynamic data by holistically utilizing POI and bicycle movement data to identify functional areas specialized for bicycle use.

Third, to accurately extract functional areas, an appropriate analytical method is required that can effectively reflect the selected data. In previous studies, single datasets such as POIs were mainly used, applying algorithms such as Word2vec and LDA to extract features, followed by clustering or classification algorithms. However, when fusion of heterogeneous data such as static and dynamic data is required, the fusion method may vary depending on the data characteristics and research objectives. In general, although scholars have different perspectives on heterogeneous data fusion, it can largely be classified into data-, feature-, decision-, and hybrid-level fusion [11]. Data-level fusion integrates different data sources by converting them into a uniform format during the preprocessing stage, enabling single-model training. Its main advantage is that it allows joint learning, but it has the drawback that information loss may occur in the conversion process. Feature-level fusion extracts features independently from each data source and then combines them;

this approach has recently been widely applied as input in deep learning. It preserves the inherent representational power of the data, but additional processes such as normalization and dimensionality reduction are required. Wang and Feng [25] fused various data for functional area extraction, extracting scene features from remote sensing images (128 dimensions), building object features (14 dimensions), socioeconomic features based on POIs (14 dimensions), and human activity features based on Weibo (50 dimensions), combining them into a 206-dimensional feature data for functional area to classification. Decision-level fusion combines the results after performing independent predictions for each dataset, maintaining classifier independence and preventing model collapse in the event of fusion failure. Yang et al. [22] classified urban functional areas using building and POI data; the building data were modeled with a dynamic GCNN (DGCNN) to extract morphological characteristics by constructing a graph structure of relationships among buildings within each block, while the POI data were modeled using Word2vec-based embeddings. The final functional area classification was then performed using a stacking ensemble classifier. As this method generates prediction values by feeding the features extracted from each data source independently into random forest classifiers and then combines these predictions in a logistic regression-based meta-classifier to make the final decision, it can be regarded as decision-level fusion. Hybrid-level fusion combines the above approaches, enabling the incorporation of diverse data characteristics, but it has the limitation of increased model complexity and computational demand. Zhang et al. [26] proposed TriNet, a tripartite neural network structure with three independent branches, namely, ImgNet, POINet, and TrajNet, for urban functional area identification. In their framework, different types of data were processed and integrated by assigning them to the three independent branches. They used data-level fusion by rasterizing vector data in ImgNet and processing them together with satellite imagery; additionally applying feature-level fusion by integrating the vectors extracted from each neural network. In this study, a hybrid data fusion approach was applied to derive functional areas specialized for bicycle use, while employing semantic-level fusion in the process of combining static and dynamic data. Semantic-level fusion is not merely merging but a high-dimensional integration technique that incorporates the meanings and roles inherent in each dataset into interpretation, enabling multidimensional functional area analysis. Jing et al. [20] proposed Co-Matrix Factorization, a semantic-based fusion method using taxi GPS data and POI information, which has the advantage of complementing missing or imbalanced data with contextual shared information. That is, by simultaneously decomposing different matrices while sharing common latent factors, it naturally performs data fusion and makes it possible to extract functional areas that reflect multidimensional semantics. Therefore, this approach is suitable for the characteristics of bicycle data, where OD matrices between stations often have many missing values. Through combined static and dynamic information, it can extract even latent semantic-level functional information, thereby contributing to policy-making. In summary, this study sets bicycle stations as the basic spatial analysis unit, extracts features from bicycle movement data containing temporal information and from land-use data containing spatial information at the feature level, and then integrates them at the semantic level to derive functional areas. This approach integrally reflects bicycle usage behaviors and spatial contexts, allowing for more precise interpretation of functional areas and further providing policy implications such as the strategic placement of bicycle infrastructure and the promotion of bicycle use.

The main contributions of this study can be summarized as follows. First, a purpose-oriented model for functional area identification related to bicycle use is proposed, unlike previous research in which urban structures were classified generally. Second, it establishes line-based spatial units that reflect actual bicycle networks and integrates spatial-temporal

datasets aligned with the analytical purpose. This enhances consistency between mobility behaviors and spatial structures, offering a new perspective on defining spatial units for mobility-based functional area analysis. Third, it employs semantic-level fusion, capturing the complex nature of urban functions associated with bicycle use, which leads to deeper interpretations and more relevant policy insights.

2. Materials and Methods

2.1. Study Area

This study examines Seoul, Republic of Korea, to identify bicycle-oriented functional areas based on spatial–temporal information. As the capital, Seoul houses approximately a quarter of the national population and ranks sixth in the Global Power City Index; it is recognized as a leading global city [<https://mori-m-foundation.or.jp>]. As of May 2022, the Ddareungi public bicycle-sharing system operated 2628 stations, significantly outpacing the city’s 302 subway stations (Figure 1). Although Seoul’s Ddareungi bicycle stations are relatively evenly distributed across the city to ensure spatial equity among regions, in reality, they tend to be concentrated in central business districts (CBDs), high-density residential areas, and major transportation hubs. In recent years, the city has actively expanded bicycle roads and improved related infrastructure as part of efforts to establish an eco-friendly transportation system, and the number of stations has continued to increase. As of 2021, Seoul’s bicycle-road network accounted for approximately 20 percent of the total road length, forming a connected system linking the downtown area with the Han River. Expansion projects are currently underway to enable continuous riding along the major east–west and north–south corridors.

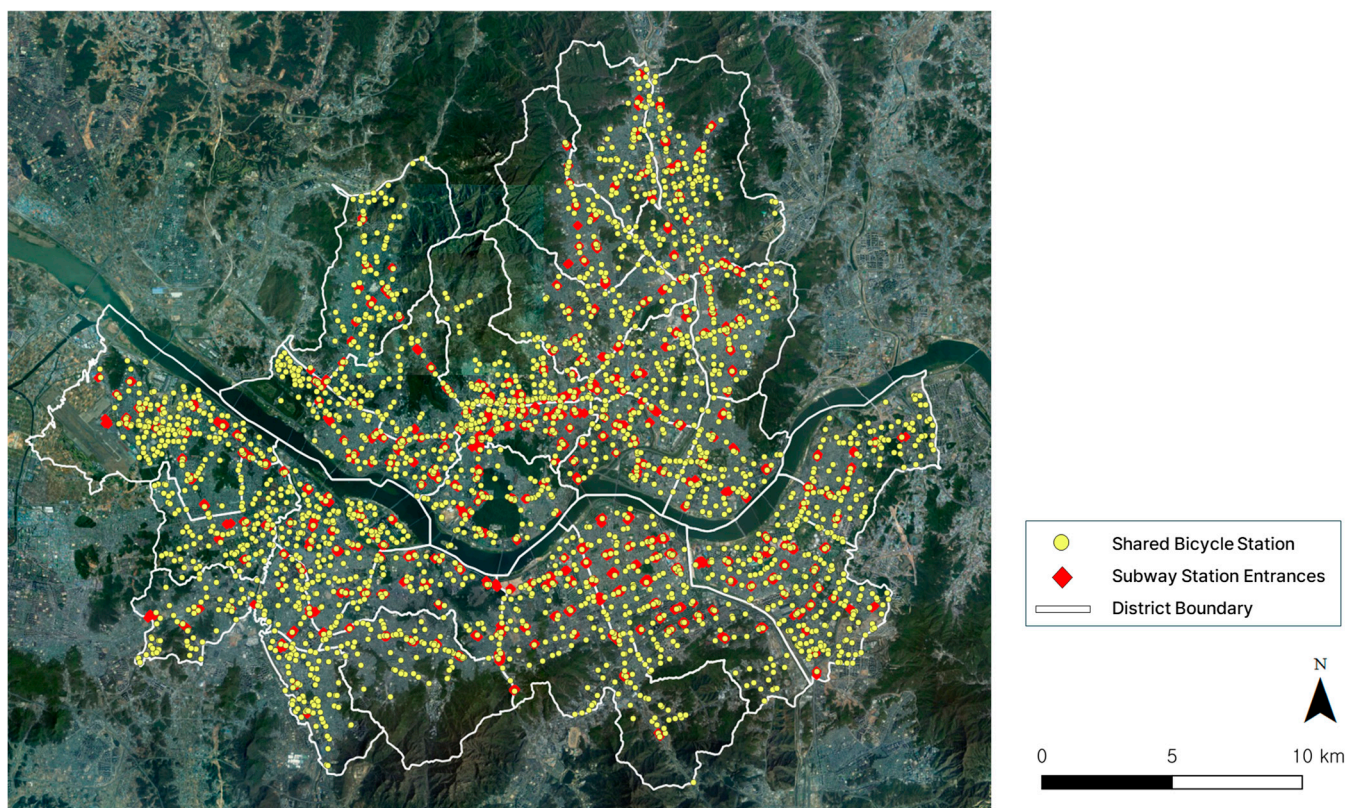


Figure 1. Study area in Seoul, Republic of Korea.

Bicycle use in Seoul is primarily for leisure rather than daily commuting, resulting in a relatively low-mode share compared to major global cities. Demand is concentrated along

the Han River and specific high-demand areas, with OD patterns showing trips clustered in these locations. Many users rent bicycles for leisure at specific stations and return them to the same location, reflecting a preference for circular trip patterns.

2.2. Data Source

2.2.1. Bicycle Network Data

This study established spatial units specialized for bicycle use by collecting and processing bicycle network data from OpenStreetMap (OSM) (<https://www.openstreetmap.org>). OSM, an open-source spatial database built through global user participation, provides road network information as nodes and edges, facilitating analysis in environments such as Python 3.13.9. Only roads classified as network type = bike were extracted from the OSM dataset for this study, enabling the analysis of line-based road networks accessible from bicycle stations.

2.2.2. Bicycle OD Data

This study used public bicycle station data and trip records from Seoul Open Data Plaza [<https://data.seoul.go.kr>] to analyze bicycle use. The station dataset includes attributes such as station ID, name, installation date, operational method, address, and geographic coordinates for geospatial processing. The trip dataset offers detailed information such as bicycle ID, rental time, rental station ID and name, dock count, return time, return station ID and name, return dock count, trip duration, and travel distance, facilitating OD mobility pattern analysis. Campbell et al. [29] revealed that weather conditions such as temperature and precipitation have a significant impact on bicycle usage, with usage particularly decreasing in winter. Accordingly, an examination of Seoul's bicycle rental records for 2022 confirmed a seasonal bias, as shown in Figure 2. However, as the purpose of this study was to propose a functional area extraction methodology specialized for bicycle use, the data for May, the month with the highest usage, was selected for analysis. Based on this, data from May 2022 consisting of 2628 stations and 4,933,660 rental records were used for analysis.

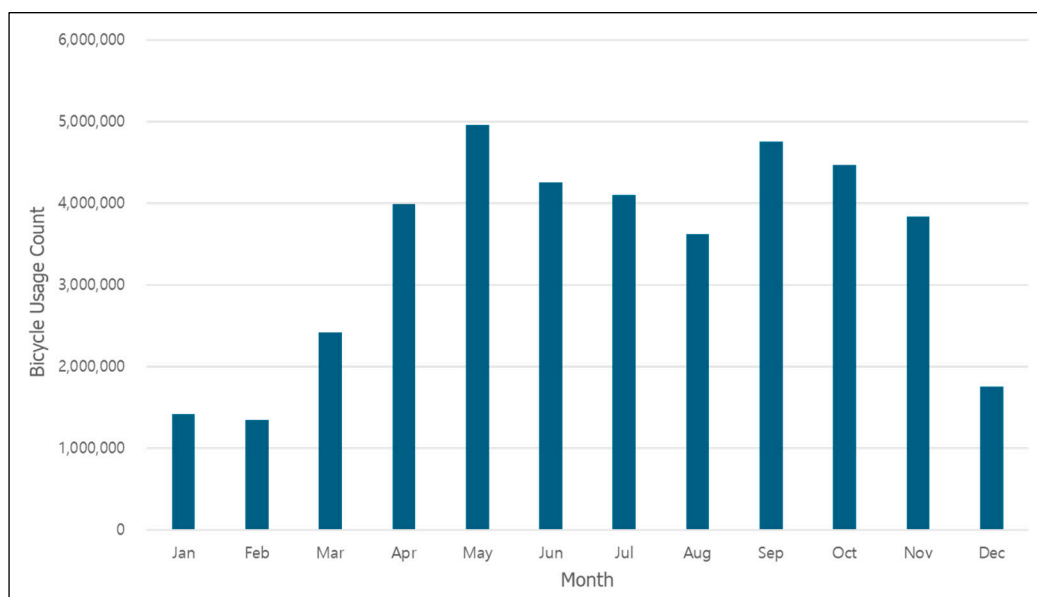


Figure 2. Monthly bicycle usage in Seoul in 2022.

2.2.3. POI Data

This study extracted static information related to bicycle use by integrating multiple land-use-related POI datasets. To identify functional areas for bicycle users, POI categories were selected based on trip purpose items from the Household Travel Survey. Instead of relying on a single dataset, various public datasets were collected and utilized. Building data came from the Basic Building Address Map provided by the Ministry of the Interior and Safety [<https://business.juso.go.kr>], while commercial facility information was obtained from Local Administration Licensing Data via LocalData [<https://www.localdata.go.kr>]. For categories difficult to capture through building data alone, subway station data were sourced from the Korea Transport Database [<https://www.ktdb.go.kr>], and bus stop and park data were obtained from the Seoul Open Data Plaza.

Due to differences in geometry and attribute structures, a data-level fusion approach was used to integrate the datasets. Building data in polygons was converted to point features by extracting centroids, while address-based datasets, such as commercial facilities, were geocoded into spatial objects. Through this process, approximately 780,000 POI records were compiled, providing static information for bicycle usage pattern analysis (Table 2).

Table 2. POI types considering trip purposes.

Category	POI Type	Count	Related Trip Purpose
Residential	Detached house	315,214	Return home, To visit relatives
	Apartment	30,311	
	Others	103,969	
Industrial	Public enterprise	1132	Go to work, Back to work
	Private company	92,349	
	Factory	7960	
Public Service	School	9114	Go to school
	Academy	5835	To attend academy classes
	Job-related service	6390	For job-related (work) reasons
	Hospital	2704	To get medical treatment at the hospital
	Pharmacy	5130	
Commercial	Shopping	29,586	To buy something (shopping, food packaging, etc.)
	Leisure	12,002	For recreation/sports
Green Space	Park	1884	/tourism/leisure
Dining	Restaurant	102,514	To eat
	Café	22,826	
	Bar	17,600	
Transportation	Bus stop	10,905	To pick up or drop off someone
	Subway station	1735	
	Parking lot	656	
Total count		780,016	

2.3. Methodology

This paper proposes a bicycle-specific functional area identification methodology by integrating spatial and temporal information. Line-based spatial units were established to reflect bicycle use characteristics, and feature information was extracted from spatial-temporal datasets. The extracted features were semantically fused to derive latent functional information. Clustering was then performed on this integrated information, and the semantic characteristics of each cluster were analyzed to classify functional areas for bicycle use. The research framework is illustrated in Figure 3.

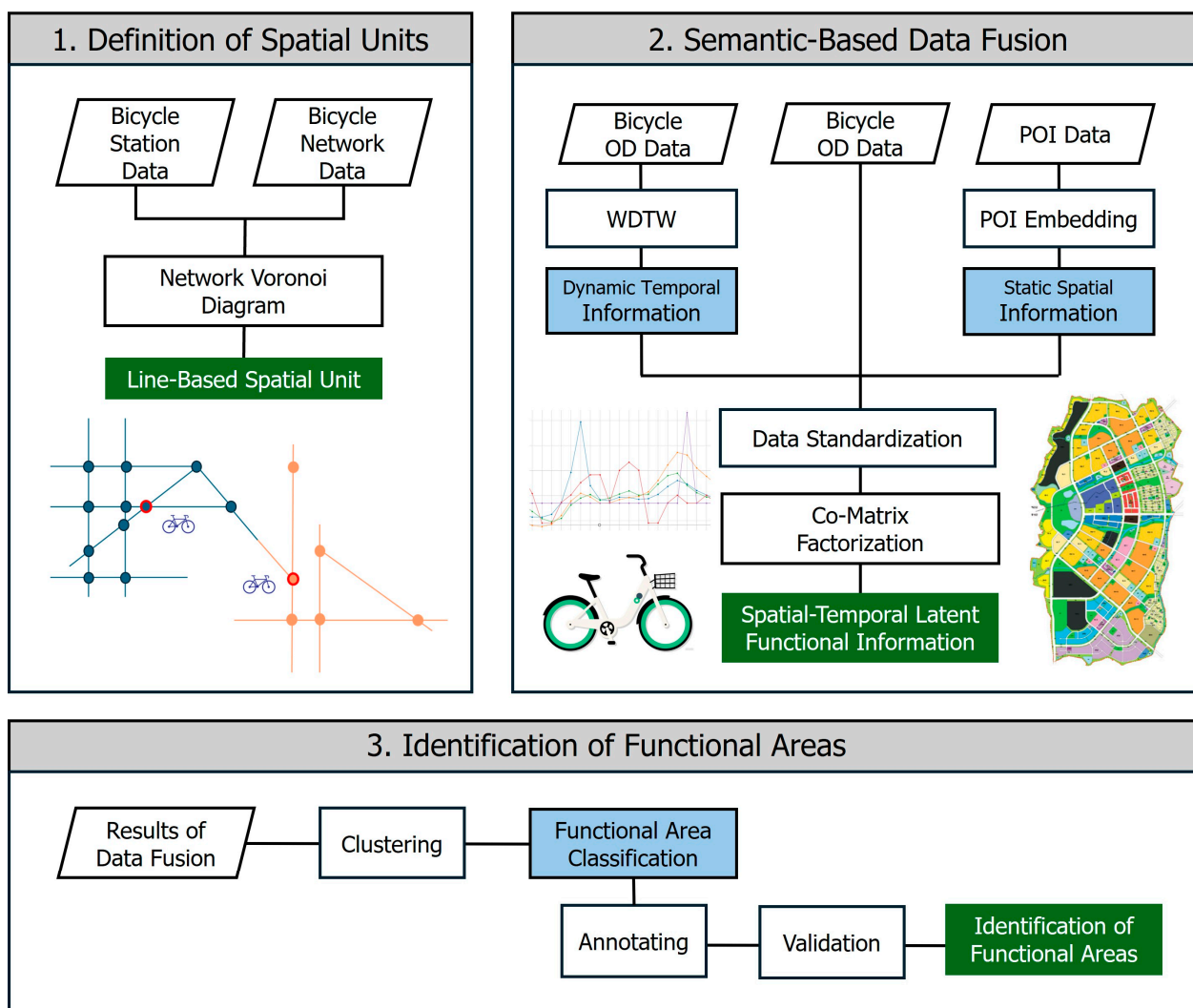


Figure 3. Research framework.

2.3.1. Method for Defining Spatial Units Using the Bicycle Network

This section outlines the procedure for dividing and extracting road networks accessible from bicycle stations to define spatial units for bicycle use. Line-based spatial units were defined using the Network Voronoi Diagram (NVD). Unlike a conventional Voronoi diagram, which uses Euclidean distance to partition space, the NVD utilizes network distance, allowing for more realistic space partitioning. This method is particularly beneficial in complex urban areas, as it reflects actual travel paths [30] and is well-suited for bicycle-oriented functional area identification.

Following the definition by Okabe et al. [31], the NVD was applied to divide the road network with respect to bicycle stations. Let there be n generators, $P = \{P_1, P_2, \dots, P_n\}$ on network $G = (V, E)$. The Voronoi region of each generator P_i , denoted as Vor_i , is defined as

$$Vor_i = \{p \in G \mid d_s(p, p_i) \leq d_s(p, p_j), \forall j \neq i\} \quad (1)$$

where $d_s(p, p_j)$ represents the shortest path distance along the network, and Vor_i corresponds to the set of points whose shortest network distance is minimized to generator P_i .

The procedure for applying this concept to bicycle stations and bicycle road networks, which consist of nodes and edges, is as follows. First, for each bicycle station, the nearest node on the network is selected as its central node, ensuring that each station is matched to only one central node without duplication. Next, Dijkstra's algorithm calculates the

shortest path distance from each central node, and nodes are classified into the region of their closest central node. Then, edges are classified according to their nodes: if both nodes of an edge belong to the same cell, the edge inherits that cell's value. However, if the nodes of an edge belong to different cells, the edge is split, and each segment is assigned to the corresponding cell. Through this process, the bicycle road network is partitioned and indexed by station, which was ultimately adopted as the line-based spatial unit in this study.

2.3.2. Semantic-Based Data Fusion Method Considering Spatial–Temporal Information

This section presents a semantic-based data fusion method to integrate spatial–temporal datasets related to bicycle use, enhancing functional area identification accuracy. The fusion process utilizes Co-Matrix Factorization to extract and combine dynamic and static information from a semantic perspective.

- Semantic-Based Data Fusion Method Using Co-Matrix Factorization

This subsection discusses the integration of heterogeneous datasets (dynamic and static information) at a semantic level using Co-Matrix Factorization. This method goes beyond data merging to capture the intrinsic meanings and roles of each dataset, allowing for the extraction of latent functional information. While various semantic-based data fusion approaches exist, this study adopted the methodology proposed by Jing et al. [20].

Co-Matrix Factorization simultaneously factorizes multiple data matrices using a shared latent representation. This method addresses missing values, enhances semantic information, and leverages complementary effects across datasets. Widely used in recommender systems, it decomposes a rating matrix into user and item latent matrices for predicting ratings of unrated items. Though limited in urban research, Kang et al. [32] applied Non-negative Matrix Factorization (NMF) to uncover taxi drivers' travel patterns and define supply areas. Subsequently, Kang et al. [33] analyzed place-visiting patterns among social groups, extracting latent correlations. While NMF operates on a single matrix, Co-Matrix Factorization jointly factorizes multiple matrices, sharing latent matrices and facilitating the learning of shared information and contextual similarities, making it ideal for data fusion and mutual reinforcement.

This study applied Co-Matrix Factorization to identify bicycle-use functional areas using three datasets: (i) an OD matrix (connectivity between spatial units), (ii) an OD temporal frequency matrix (dynamic temporal information), and (iii) a POI matrix (spatial semantic information). Because bicycle trips are typically short, OD flows are concentrated between nearby stations, leading to many missing values in the OD matrix. Following Jing et al. [32], knowledge transfer was utilized to transfer learned similarities between datasets, addressing missing values and data sparsity, and generating a refined OD matrix. The overall process is expressed as follows:

$$L(H, M, N) = \frac{1}{2} \|W(Q - (MN^T))\|_F^2 + \frac{\alpha}{2} \|P - (MH^T)\|_F^2 + \frac{\beta}{2} \|R - (NN^T)\|_F^2 + \frac{\gamma}{2} (\|H\|_F^2 + \|M\|_F^2 + \|N\|_F^2) \quad (2)$$

where matrix Q denotes the bicycle OD matrix, P represents the POI matrix, and R corresponds to the temporal similarity matrix of bicycle OD flows. All three matrices share the same spatial units defined by bicycle stations, and matrix factorization is performed on this common basis. Specifically, P is approximated as MH^T , R as NN^T , and Q as MN^T , where M represents the latent spatial representations of stations, N captures the latent temporal characteristics, and H represents the latent semantic features of POIs. The model learns a shared latent structure across the three datasets through this structure, enabling complementary integration of semantic information derived from heterogeneous sources.

In addition, W denotes a weight matrix that reflects the reliability of observed values in Q , assigning smaller weights to missing entries or low-confidence items. The first, second, and third terms of the objective function minimize discrepancies between observed and approximated matrices to incorporate the corresponding information. By contrast, the final term performs $L2$ regularization on M , N , and H to control model complexity. Here, α adjusts the importance of the loss term related to P , β controls the importance of the loss term related to R , and γ weights the regularization term to prevent overfitting and constrain the magnitudes of M , N , and H . Finally, matrices M , N , and H are optimized through Stochastic Gradient Descent (SGD). This process yields the refined matrix Q' , in which spatial and temporal information are jointly integrated.

In this study, we followed the framework of Jing et al. [20] to extract latent functional information by complementing bicycle OD data. However, when constructing the temporal similarity matrix R and the POI matrix P , we refined and enhanced the construction procedures so that the semantic information associated with bicycle use could be more explicitly reflected. Details of these improvements are provided in the following section.

- Method for Extracting Dynamic Temporal Information

This section outlines the procedure for extracting dynamic information by generating an OD matrix from bicycle trip data to analyze temporal travel patterns. The OD matrix represents trips between bicycle stations, indicating spatial connectivity strength, and is commonly used for functional area identification in previous studies [9,24]. This study aims to complement dynamic and static information centered on the OD matrix, focusing on the temporal usage pattern of bicycles at each station.

Kang et al. [32] used the Dynamic Time Warping (DTW) method to measure similarity in time-series data by warping the time axis, unlike the rigid Euclidean distance that compares values at the same time steps. DTW quantifies temporal similarity between two stations through inflow pattern graphs. However, standard DTW has limitations, such as excessive distortion of the time axis for similar-shaped series and equal treatment of all time intervals, regardless of their significance. To address these issues, prior studies have proposed weighting schemes [5,34].

Lee et al. [34] analyzed spatial-temporal demand patterns of Citi Bike in New York City, USA, highlighting that conventional DTW may misidentify morning (commuting to work) and evening (commuting home) peaks as identical patterns. They addressed this by applying a weighting scheme that penalizes larger temporal misalignments, reducing similarity scores for significant time differences. This enhanced the distinction between morning and evening commuting demands.

Building on these insights, the present study applied Weighted Dynamic Time Warping (WDTW) to analyze the temporal patterns of bicycle trips. To incorporate weights into DTW, the point-wise distance $d(x_i, y_j)$ is multiplied by a weight depending on the phase difference $|i - j|$, as follows:

$$Vd_w(x_i, y_j) = w(|i - j|) \cdot d(x_i, y_j) \quad (3)$$

The weight function adopts a modified logistic function, as follows:

$$w(t) = \frac{w_{max}}{1 + \exp(-\delta(t - \epsilon))} \quad (4)$$

In (4), the weighting parameters consist of three components: w_{max} , δ , and ϵ . Here, w_{max} denotes the upper bound of the weight function, δ represents the slope that controls the strength of the penalty, and ϵ corresponds to the midpoint of the time series, at which the penalty begins to increase rapidly. Each parameter can be manually specified by the researcher.

After applying the weighting scheme, the minimum cumulative cost matrix is obtained according to (5),

$$D_w(i, j) = d_w(x_i, y_j) + \min(D_w(i-1, j-1), D_w(i-1, j), D_w(i, j-1)) \quad (5)$$

and the final WDTW distance between time series X and Y is calculated as follows:

$$DTW_w(X, Y) = \min_{p \in P} \sqrt{\sum_{k=1}^K d_w(p_k)} \quad (6)$$

In (5), the initialization is defined as $D_w(0, 0) = 0$, $D_w(i, 0) = \infty$, $D_w(0, j) = \infty$. Path constraints include the following: Endpoint, where the path begins at the upper-left corner and ends at the lower-right corner; Continuity, where only one step is allowed at a time; and Monotonicity, which prohibits backward movement along the time axis.

In (6), P denotes the set of all possible warping paths, p_k represents the position (i, j) at the k -th observation along a warping path, and K is the length of the warping path. This procedure quantifies temporal similarity between stations, and the results are organized into a matrix R , which is subsequently used as an input for Co-Matrix Factorization.

- Method for Extracting Static Spatial Information

This section outlines the procedure for constructing and preprocessing spatial datasets, including POI data, to capture land-use characteristics. Jing et al. [20] calculated POI densities to extract spatial semantics, aggregating the number of POIs in six categories into six-dimensional vectors for each unit. While this method reflects the distribution of specific POI types, it does not fully represent the functional properties of areas with mixed POI types.

To address this limitation, the present study adopted the POI embedding method by Zhai et al. [1], based on Word2Vec, to extract static features of each station. POI embedding learns spatial relationships among POIs and converts categories into high-dimensional vectors. Initially used to capture semantic relationships between POI categories or identify urban functional areas [1,17,35], it has recently become an input feature for urban analytics models. For example, Yang et al. [36] integrated POI embedding with an LDA topic model to create functional area vectors that merge POI vectors and topic distributions, subsequently classifying UFZs using an SVM classifier. Similarly, Yang et al. [37] used POI vectors as initial node attributes in a GCNN to classify urban functional areas.

This study applied the POI embedding method to vectorize the semantic characteristics of areas surrounding bicycle stations. Instead of directly adopting Zhai et al. [1]'s framework, the selection and aggregation of POI categories were refined to represent functional areas specialized in bicycle use. A training model generated embedding vectors for 20 POI categories, listed in Table 1. For each central POI, the k nearest POIs were defined as contextual POIs to create the training dataset. Distance-based data augmentation was implemented to ensure that POI categories closer together produced more similar semantic vectors, utilizing the distance coefficient proposed by Yan et al. [38] (7).

$$\beta = \left[\frac{1 + \frac{\sum_{k=1}^{|L|} p_{ik}}{|L|}}{1 + d^\alpha(l_i, l_j)} \right] \quad (7)$$

where $|L|$ denotes the total number of POIs, $d(l_i, l_j)$ represents the distance between POIs l_i and l_j , and α is the inverse distance factor, which was set to 1 in this study.

The training dataset was used to develop a Skip-gram model in Word2Vec that predicts surrounding POIs from a central POI. The output was converted to probabilities using a SoftMax function, generating embedding vectors for each POI type. Semantic information was aggregated for each defined line-based spatial unit. While Zhai et al. [1] calculated the mean vector by summing all POI embeddings in each spatial unit and dividing by their count, this study introduced a weighting scheme to better represent bicycle mobility characteristics.

The weighting procedure involved identifying the three nearest bicycle road segments for each POI and calculating their distances. These distances were divided into five ranges, each assigned differential weights. Following Zhao et al. [39], POIs classified as Park and Transportation were positively weighted to enhance bicycle use. Detailed weighting criteria are listed in Table 3. This process derived semantic vectors representing the spatial characteristics of each station, organized into a matrix P , which served as input data for Co-Matrix Factorization.

Table 3. Weighting criteria for aggregating static information by spatial unit.

Criterion	Class	Weight Value
POI Category	Residential, Industrial, Public Service, Commercial, Dining,	1
	Green Space (Park), Transportation	2
Distance from POI to Bicycle Road Network	0–10 m	1
	10–50 m	0.50
	50–100 m	0.33
	100–500 m	0.25
	≥500 m	0.20

2.3.3. Clustering and Interpretation Approach for Functional Area Identification

This section explains the functional area identification procedure using the updated matrix Q' , which was derived from bicycle OD data (Q), POI data (P), and temporal similarity data (R) through Co-Matrix Factorization. The matrix Q' is constructed based on the shared latent factors M (between Q and P) and N (between Q and R), thereby compensating for the sparsity and missing values in the original Q while simultaneously learning the common latent structures across the three datasets. Each element of Q' thus represents the degree of functional association between two regions in semantic and temporal contexts, even when direct observations are absent.

Jing et al. [20] identified urban functional areas using spectral clustering on the Q' matrix. In this study, the P matrix, derived from high-dimensional semantic embeddings created by Word2Vec, supports using K-means clustering, which is better suited for Euclidean vector spaces. To cluster high-dimensional, complex data more effectively, t-SNE was applied to reduce the dimensionality of the Q' matrix while preserving its essential structure, followed by K-means clustering on the reduced representation.

In the interpretation stage, two indicators were calculated to clarify the functional roles of each cluster: Frequency Density (FD) and Category Proportion (CP), respectively defined in (8) and (9).

$$FD_i = \frac{n_i}{N_i} \quad (8)$$

$$CP_i^k = \frac{n_i^k}{\sum_j n_j^k} \quad (9)$$

where i denotes the POI category. Let n_i be the number of POIs of category i within a cluster, and N_i the total number of POIs of category i across all clusters. Then, FD_i represents the frequency density of category i , calculated as the relative concentration of that category compared to its overall distribution. Similarly, CP_i^k refers to the category proportion of POI type i within a given cluster k .

Furthermore, as the Q' matrix generated in this study incorporates not only POI information but also actual bicycle OD flows and temporal information, the functional interpretation of each cluster was performed by jointly considering indicators such as Average OD Trips per Station (Q matrix) and Average Enhanced OD Trips per Station (Q' matrix). Specifically, the sum of Q' values for each bicycle station indicates its connectivity to other regions in the latent space, providing insight into cluster characteristics.

3. Results

3.1. Bicycle Network-Based Spatial Unit Definition

A spatial partitioning process was conducted for 2628 bicycle stations in Seoul, utilizing the bicycle road network from OSM and NVD. The nearest node to each station was identified as the central node using a KD-Tree-based nearest neighbor search algorithm, ensuring one-to-one matching without duplication. Dijkstra's algorithm classified neighboring nodes based on shortest path distances, resulting in 117,110 nodes extracted and stored in a dictionary structure.

Edges corresponding to classified nodes were then partitioned. When an edge connected nodes from different classes, the Shapely library (version 1.8.5) split the edge, assigning a unique classification to each segment. This process classified a total of 343,946 edges, with results stored in GeoJSON format. These classified edges served as line-based network spatial units in this study.

Figure 4 illustrates the results. Figure 4a shows network nodes extracted from bicycle stations, each assigned an index corresponding to its central node's station ID. Figure 4b presents edges derived from Figure 4a, where each edge inherits the station index of its constituent nodes, enabling final classification of road networks by station. This network-based spatial unit facilitates aggregation of surrounding spatial information, such as POIs, around roads accessible from bicycle stations.

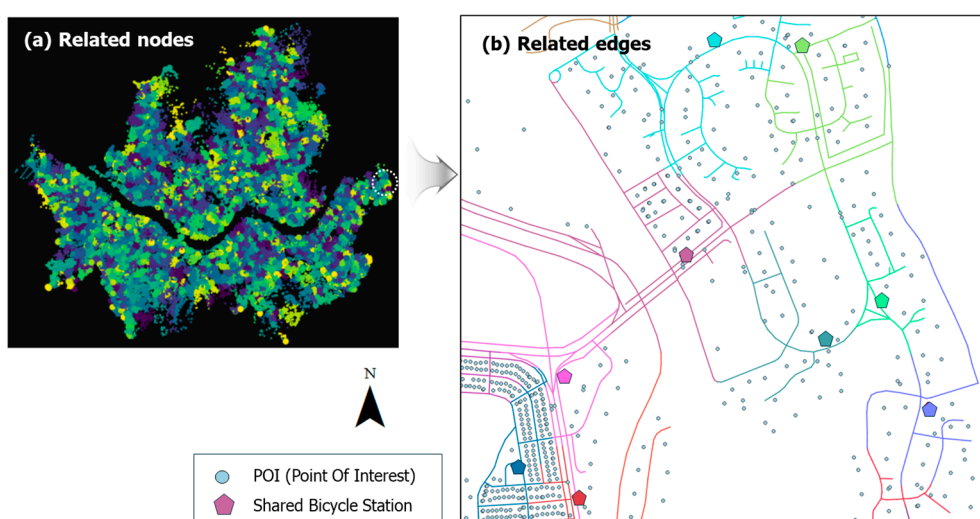


Figure 4. Spatial unit delineation results using a bicycle-road-network-based NVD.

3.2. Semantic Data Fusion with Spatial–Temporal Information

3.2.1. Extraction of Dynamic Temporal Information

Dynamic bicycle movement data (Q, R) were incorporated into Co-Matrix Factorization using an OD matrix and a temporal similarity matrix. Matrices were constructed from Seoul’s public bicycle trip records across 2628 stations, with stations lacking rental or return activity assigned a zero value.

The OD matrix (Q) in Table 4 aggregates trip counts between stations. Trips were primarily concentrated between neighboring areas, with rare long-distance trips resulting in many zero values. Seoul’s public bicycle system displayed notable patterns at specific stations with high rental and return frequencies. Stations for recreational purposes often had overlapping origins and destinations. For instance, Ttukseom Resort (Exit 1) and Yeouinaru Station (Exit 1) near Han River Park served as major origins and destinations for bicycle trips.

Table 4. OD trip aggregation results by station (Q matrix).

		Bicycle Station ID																				
Bicycle Station ID		102	103	104	105	106	107	108	109	111	112	...	5078	5301	5305	5306	5751	5752	5753	5851	5852	5853
	102	397	101	99	34	75	284	68	39	19	18	...	0	0	0	0	0	0	0	0	0	0
	103	77	369	51	52	153	75	56	51	24	29	...	0	0	0	0	0	0	0	1	0	0
	104	125	53	125	16	38	71	63	32	15	29	...	0	0	0	0	0	0	0	2	0	0
	105	33	42	10	104	52	18	116	69	21	50	...	0	0	0	0	0	0	0	0	0	0
	106	129	198	33	59	505	30	85	222	44	68	...	0	0	0	0	0	0	0	1	0	0
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	5752	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
5753	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	
5851	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	195	10	0	
5852	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	8	10	0	
5853	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	

2628 rows $\times$ 2628 columns.

Next, the temporal similarity matrix (R) between stations was derived using the WDTW algorithm to quantify similarities in hourly bicycle usage patterns. This matrix served as input for Co-Matrix Factorization alongside the OD matrix. The WDTW computation involved first aggregating bicycle returns for each station by time interval. A standardization process was conducted using LOWESS (Locally Weighted Scatterplot Smoothing) and Z-score normalization. In the LOWESS step, the frac parameter was set to 0.2, meaning each predicted value incorporated 20% of neighboring data points, minimizing noise in the time-series data. For the WDTW computation, parameter settings were adopted from Kang et al. [32]. Specifically, the maximum weight $w_{\max}$ was set to 1, the slope δ to 0.1, and the midpoint ϵ to 1. The resulting temporal similarity matrix (R) is presented in Table 5.

Table 5. Temporal similarity matrix between stations (R matrix).

		Bicycle Station ID										
Bicycle Station ID		102	103	104	105	106	107	108	109	111	112	...
	102	0	1.415115	6.027184	0.550858	0.747530	9.262674	1.389063	0.809668	0.867471	0.442452	...
	103	1.415115	0	4.849054	0.865617	0.582082	6.033004	3.235862	0.963805	0.758531	1.599929	...
	104	6.027184	4.849054	0	7.341838	8.133845	1.489657	2.484759	7.716670	8.134696	6.002384	...
	105	0.550858	0.865617	7.341838	0	0.288882	7.001657	1.915456	0.455287	0.465555	0.470225	...
	106	0.747530	0.582082	8.133845	0.288882	0	7.181926	2.525698	0.284783	0.179199	1.007133	...
	...	...	...	...	...	...	...	...	...	...	...	...
	5752	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
	5753	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
	5851	1.449870	0.659773	8.660493	0.722720	0.323853	7.359159	3.190453	0.471608	0.343099	2.041195	...
	5852	2.272780	1.498987	8.745269	1.404750	1.011186	5.701480	3.729711	0.828786	0.714414	2.519627	...
	5853	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...

2628 rows $\times$ 2628 columns.

Unlike the OD matrix, stations without trip records are not assigned a value of zero but retained as null. In WDTW, smaller values indicate higher similarity; replacing missing values with zero would wrongly suggest that stations with no data have the highest

similarity, causing misinterpretation. As can be seen in Table 5, diagonal values for the same station are zero, indicating perfect temporal similarity.

3.2.2. Extraction of Static Spatial Information

To incorporate static spatial information (P) for Co-Matrix Factorization, semantic data from POIs was aggregated within line-based spatial units surrounding each bicycle station. POI types were vectorized, constructing tuples for each POI with its 10 nearest neighbors. A distance-based augmentation coefficient assigned higher frequencies to closer POI pairs. This process generated 8,416,826 original tuples and an additional 2,721,077 augmented tuples, resulting in a total of 11,137,903 tuples for training.

The Skip-gram Word2Vec model was trained on tuples with an embedding dimension of 70 and 10,000 iterations. The output vectors represent hidden layer embeddings of each POI type, encoding their semantic meanings. These embeddings were assigned to individual POIs and aggregated by bicycle station, factoring in POI types and distance weights as outlined in Table 3. Consequently, each bicycle station is represented as a 70-dimensional semantic vector, serving as static spatial information (Table 6).

Table 6. Results of POI-based semantic information extraction by station (P matrix).

Bicycle Station ID	Embedding Results (70 dimensions)										
	0	1	2	3	4	5	6	7	8	9	...
102	0.416487	−0.158637	−0.226452	0.254378	0.600380	−0.204288	0.291262	−0.271597	−0.084234	0.358788	...
103	0.396833	−0.165042	−0.304261	0.267791	0.563749	−0.209214	0.281475	−0.292068	−0.024150	0.329308	...
104	0.461867	0.034045	−0.287962	0.431353	0.640563	−0.420478	0.436185	−0.112449	−0.110945	0.303682	...
105	0.633366	0.166209	−0.248414	0.559907	0.654447	−0.385225	0.476219	−0.088170	−0.269816	0.326406	...
106	0.540244	0.024397	−0.379808	0.441713	0.577836	−0.296120	0.318313	−0.184423	−0.064344	0.230602	...
...	...	...	...	...	...	...	...	...	...	...	...
5752	−0.094440	−0.896108	0.297623	−0.151369	0.488541	0.009165	0.388839	−0.504931	−0.091359	0.908376	...
5753	0.392842	−0.420355	−1.375567	−0.135741	−0.513409	1.289962	−0.123697	−0.160912	−0.757437	−0.454109	...
5851	−0.030416	−0.632836	0.040875	−0.037444	0.443812	0.006919	0.294876	−0.426770	0.048235	0.612541	...
5852	0.218871	−0.281096	−0.230428	0.168613	0.411862	−0.051084	0.290885	−0.330305	−0.012623	0.366870	...
5853	1.154471	0.524940	−0.619453	0.747115	0.707610	−0.258740	0.497085	−0.042117	−0.558134	0.035194	...

2628 rows $\times$ 2628 columns.

3.2.3. Integrated Semantic-Based Data Fusion

Finally, the Q , P , and R matrices described in previous sections were integrated using Co-Matrix Factorization, resulting in the updated matrix Q' , which incorporates spatial–temporal information. Data normalization was performed prior to fusion. The OD matrix (Q) underwent Min–Max scaling to align with trip-count properties. The POI matrix (P), comprising semantic embeddings, was normalized using $L2$ -normalization for stability in the inner-product-based loss function. For the temporal similarity matrix (R), stations with no trip records were assigned null values, replaced by the maximum matrix value to indicate low similarity. Min–Max normalization was then applied and inverted, designating 1 as maximum similarity and 0 as minimum similarity.

The preprocessed matrices were fused using Co-Matrix Factorization as described in (2). Latent vectors M , N , and H were initialized randomly and updated iteratively via SGD until convergence. Following Jing et al. [20], parameters were set to $\alpha = 0.1$, $\beta = 0.5$, and $\gamma = 0.01$. The learning rate and number of epochs were fixed at 0.001 and 100, respectively, ensuring stable convergence of the loss function. Figure 5 shows the loss over epochs, confirming stable model convergence.

Table 7 shows the updated Q' matrix. Unlike the original OD matrix (Q), which reflects direct trip frequencies, Q' integrates static spatial semantics (P) and temporal mobility similarities (R). As a result, missing or sparse connections in the original Q were complemented. For example, low or absent trip frequencies between two stations did not eliminate their high similarity in P and R , suggesting latent interactions captured in the Q' matrix.

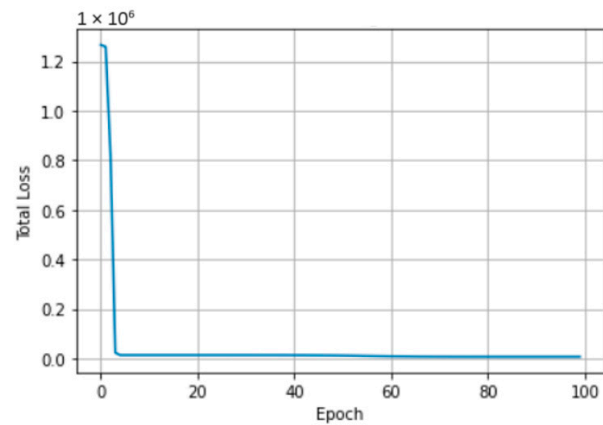


Figure 5. Visualization of the loss curve.

Table 7. Enhanced OD trips considering spatial–temporal information (Q' matrix).

Bicycle Station ID												
Bicycle Station ID		102	103	104	105	106	107	108	109	111	112	...
	102	0.002537	0.002569	0.002378	0.002566	0.002561	0.002332	0.002522	0.002527	0.002550	0.002526	...
	103	0.001776	0.001791	0.001528	0.001795	0.001805	0.001483	0.001718	0.001804	0.001809	0.001765	...
	104	0.001055	0.001064	0.000916	0.001066	0.001071	0.000890	0.001023	0.001069	0.001073	0.001048	...
	105	0.000824	0.000832	0.000735	0.000833	0.000835	0.000716	0.000806	0.000830	0.000834	0.000819	...
	106	0.001449	0.001462	0.001260	0.001464	0.001471	0.001225	0.001406	0.001468	0.001473	0.001440	...
	...	...	...	...	...	...	...	...	...	...	...	...
	5752	−0.000067	−0.000085	−0.000399	−0.000071	−0.000038	−0.000431	−0.000184	0.000020	−0.000008	−0.000076	...
	5753	0.000836	0.000858	0.000999	0.000848	0.000826	0.001005	0.000906	0.000778	0.000803	0.000839	...
	5851	0.001120	0.001133	0.001020	0.001133	0.001133	0.000997	0.001103	0.001123	0.001131	0.001115	...
5852	0.000423	0.000437	0.000563	0.000429	0.000412	0.000572	0.000479	0.000379	0.000396	0.000426	...	
5853	−0.000517	−0.000533	−0.000660	−0.000525	−0.000506	−0.000668	−0.000575	−0.000470	−0.000489	−0.000520	...	

2628 rows $\times$ 2628 columns.

3.3. Clustering Result and Functional Area Identification

Using the updated Q' matrix with spatial–temporal information, functional areas specialized for bicycle use were identified via clustering. The Q' matrix was reduced to low-dimensional vectors via t-SNE, followed by K-means clustering. The optimal number of clusters (k) was determined using silhouette scores, which peaked at 0.3702 for $k = 5$. Thus, five bicycle-use functional areas were identified, as shown in Figure 6. Clusters were characterized via dynamic and static spatial information analysis, with detailed results in Table 8.

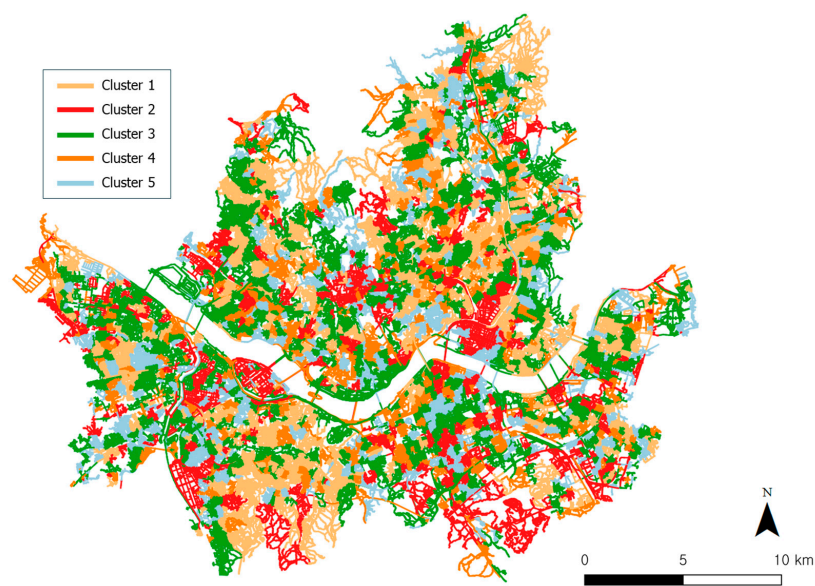


Figure 6. Bicycle functional area identification results with spatial–temporal information.

Table 8. Computed indicators for functional area identification.

			Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Dynamic Temporal Information	Average OD Trips per Station (Q matrix)		1761.531	1774.680	2213.198	1808.058	1419.763
	Average Enhanced OD Trips per Station (Q' matrix)		3.234	2.867	3.673	3.037	2.730
Static Spatial Information	Frequency Density	Residential	0.306	0.079	0.334	0.127	0.154
		Industrial	0.204	0.156	0.328	0.144	0.168
		Public Service	0.220	0.147	0.351	0.146	0.136
		Commercial	0.220	0.153	0.341	0.142	0.144
		Green Space	0.227	0.123	0.355	0.139	0.157
		Dining	0.181	0.172	0.340	0.164	0.143
		Transportation	0.210	0.148	0.339	0.149	0.154
	Category Proportion	Residential	0.195	0.081	0.140	0.126	0.146
		Industrial	0.130	0.160	0.137	0.142	0.159
		Public Service	0.140	0.150	0.147	0.145	0.128
		Commercial	0.140	0.156	0.143	0.141	0.136
		Green Space	0.145	0.126	0.149	0.137	0.148
		Dining	0.116	0.176	0.142	0.162	0.136
		Transportation	0.134	0.152	0.142	0.147	0.146

Dynamic information analysis revealed the Average OD Trips per Station (Q Matrix) and Average Enhanced OD Trips per Station (Q' Matrix) for each cluster. The Q Matrix reflects actual trip volumes, while the Q' Matrix indicates potential connectivity strength through spatial-temporal data. Cluster 3 showed the highest values in both measures, confirming it as the most bicycle-friendly area. By contrast, Cluster 5 had the lowest values, indicating the weakest bicycle usage. Thus, these clusters clearly differentiate between areas with strong and weak bicycle activity. However, the values of Q and Q' were not always directly proportional. For instance, Clusters 2 and 4 had higher Average OD Trips than Cluster 1, yet Cluster 1 exhibited a higher Average Enhanced OD Trips value. This indicates that Cluster 1 may be more bicycle-friendly than trip counts suggest.

The cluster classification using this information is as follows. Based on the data presented in Table 8, the functional areas were named according to the POI information, actual demand (Q Matrix), and latent demand (Q' Matrix) results.

- Cluster 1: Residential-Oriented, Medium-Flow with Strong Latent Zone.

The residential area proportion is relatively higher than in other clusters, and its share within the cluster is large; therefore, it was classified as a residential-oriented area. By contrast, industrial and dining proportions are relatively low, making it identifiable as a traditional residential area. The current demand for bicycle use is at a medium level, almost similar to Cluster 2, but the latent demand was found to be somewhat higher.

- Cluster 2: Dining-Industrial Mixed, Medium-Flow with Weak Latent Zone.

The dining, industrial, and commercial proportions are high, while the residential proportion is very low, exhibiting the characteristics of a typical central business district, which is the exact opposite of Cluster 1. The demand for bicycle use is at a medium level; however, in terms of latent demand, it is the second lowest after Cluster 5. Major business districts such as the Seoul City Hall area and Yeouido are included in this cluster.

- Cluster 3: Multi-functional Core, High-Flow with Strong Latent Zone.

This is a multi-functional area with the highest proportion across all POI categories, and, especially in terms of green space, it is remarkably higher than other clusters. It shows

the highest values in both actual bicycle traffic and latent demand, making it a specialized area for bicycle use. Most of the areas adjacent to the Han River, which passes through the center of Seoul, belong to this cluster, which corresponds to the fact that bicycle use in Seoul is largely centered on parks and waterfront spaces for leisure and recreational purposes.

- Cluster 4: Dining-Oriented, Medium-Flow with Moderate Latent Zone.

The dining proportion is the highest, and the industrial and commercial proportions are also high, though not as much as in Cluster 2, so it can be interpreted as a quasi-commercial area. Actual bicycle traffic is higher than in Cluster 1, but the latent demand appears somewhat lower. However, there is no significant difference in the absolute values, and both are classified as medium-level usage areas.

- Cluster 5: Industrial–Green Mixed, Low-Flow with Weak Latent Zone.

Although all POI categories show a balanced distribution, there are no distinctive strengths in this area. Within the cluster, the industrial and green space proportions are relatively high, but not as differentiated as in Clusters 2 or 3. However, both current and latent demand for bicycles are the lowest; therefore, this cluster can be interpreted as a low-usage area for bicycles. Interestingly, some central business districts such as Yeouido and Gangnam are included, which results in a relatively high industrial proportion. This outcome can be interpreted as stemming from the fact that the Q' matrix reflects not only POI-based spatial information but also spatial–temporal factors.

3.4. Validation of the Identified Functional Areas

This study identified functional areas based on enhanced bicycle OD data incorporating spatial–temporal information. Due to the lack of ground truth data, a quantitative accuracy assessment was not possible. Validation was carried out through (i) visual inspection of representative regions within each cluster to evaluate the plausibility of identified bicycle-use functional areas, and (ii) comparison with actual OD trip data, aggregated by station and time interval, to confirm accurate reflection of dynamic information. Figure 7 presents representative bicycle stations selected from each cluster to validate the study results. The selection of representative stations was based on the quantitative characteristics of each cluster summarized in Table 8. When multiple stations exhibited similar indicator values, stations including well-known landmarks of Seoul—such as Gwanghwamun, the Han River, or the Trade Center—were given priority. This selection approach allows for an intuitive understanding of the spatial meaning of each cluster and enhances the objectivity of visual validation by presenting the actual geographic coordinates (latitude and longitude) together.

Satellite imagery from Vworld [<https://www.vworld.kr>] was used for visual validation (Table 9), while dynamic information was validated through time-series OD data for selected stations (Figure 8). Validation results indicate the following. Near station 2140 in Cluster 1, the area is a residential-centered bicycle-use zone with dense single-family housing and Dorim Stream. Although the total OD value was lower (3205 trips) than that of Clusters 2 (3305 trips) and 4 (3269 trips), its latent connectivity strength was higher (5.478 vs. 3.769 in Cluster 2 and 4.609 in Cluster 4), suggesting high bicycle-friendliness despite lower trip volumes. Cluster 2, including the Gwanghwamun area (station 303), is a central business district with increased usage during morning commuting. Cluster 3, near Ttukseom Resort and Han River Park, recorded the highest number of trips (22,268), showing distinct usage patterns. Cluster 4, around Mapo-gu Office and the Agricultural and Fishery Market, has high morning usage but lower evening usage than Cluster 1. Finally, Cluster 5, near Samseong Station (station 2232), includes major business districts such as the Trade Center but shows minimal bicycle activity, classifying it as a low-use area.

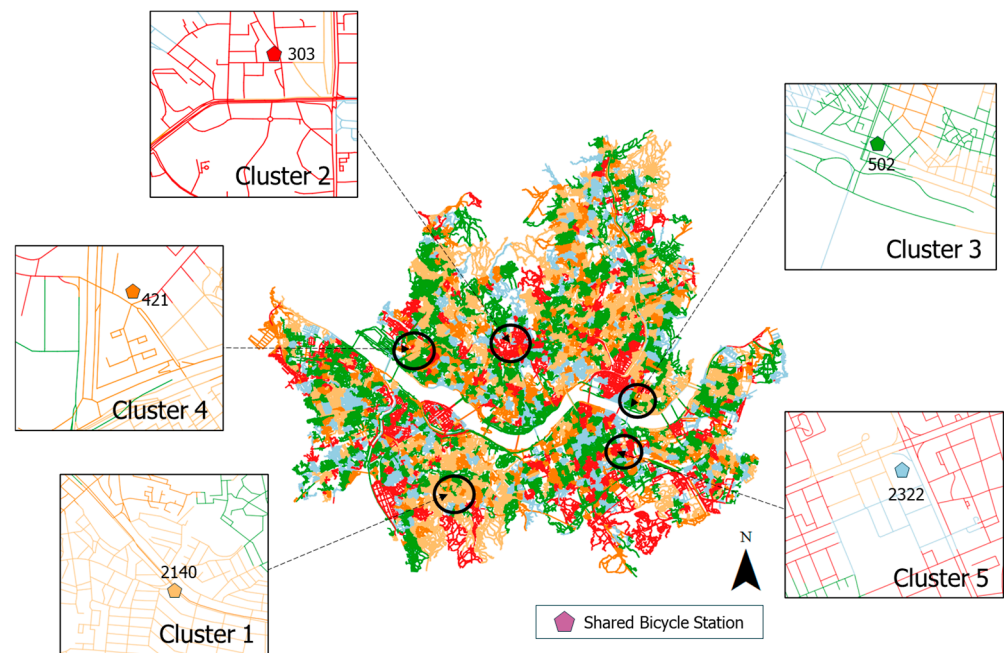
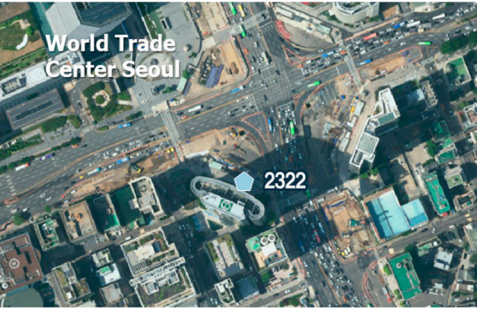


Figure 7. Representative bicycle stations by cluster.

Table 9. Interpretive validation of functional area identification using satellite imagery.

Cluster	Bicycle Station ID, Name, Location	Satellite Imagery
Cluster 1	ID: 2140 Sillim 1-gyo Intersection (Lat: 37.47842789 Lon: 126.9318619)	
Cluster 2	ID: 303 In front of Gwanghwamun Station, Exit 1 (Lat: 37.57176971 Lon: 126.9746628)	
Cluster 3	ID: 502 In front of Ttukseom Resort Station, Exit 1 (Lat: 37.53186035 Lon: 127.0671921)	

Table 9. Cont.

Cluster	Bicycle Station ID, Name, Location	Satellite Imagery
Cluster 4	ID: 421 In front of Mapo-gu District Office (Lat: 37.56574631 Lon: 126.9018631)	
Cluster 5	ID: 2322 Samseong Station, Exit 3 (Lat: 37.50809097 Lon: 127.0631027)	

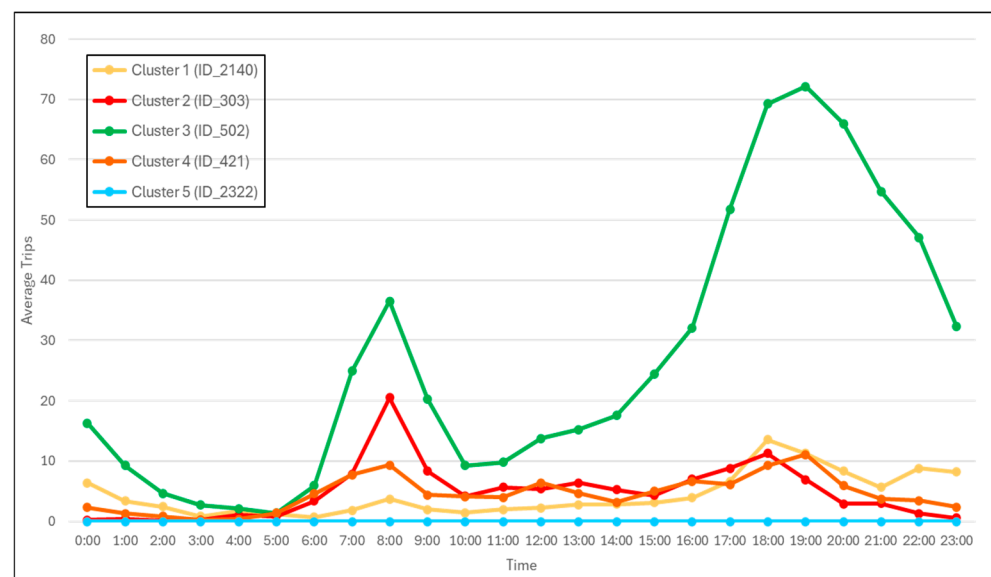


Figure 8. Average hourly trip volume for representative stations by cluster.

The validation results indicate that the proposed method effectively identifies bicycle-use functional areas by integrating spatial-temporal information. This approach enables more accurate extraction of hard-to-delineate functional areas than only static data such as POI or land-use data. However, it is difficult to sufficiently verify the methodology through visual evaluation alone. Therefore, in this study, under the assumption that the updated Q' matrix reflects latent functional intensity, the results were validated using future data (May 2025). For improved accuracy, only the top 10 stations in terms of bicycle traffic volume were examined, comparing actual traffic (Q) with latent functional intensity (Q'). The left column of Table 10 presents the Q (sum of actual traffic) and Q' values (sum of latent functional intensity) calculated based on bicycle trip records from May 2022. The

right column presents the sum of the Q values for the top 10 stations in terms of bicycle traffic in May 2025. The analysis showed that In front of Ttukseom Resort Station, Exit 1 and In front of Yeouinaru Station, Exit 1, which were the 1st and 2nd stations in May 2022, respectively, ranked somewhat lower in May 2025. It is noteworthy that the Q' values partly corresponded to the actual future data. For example, in the case of Magongnaru Station Exit 2 and Behind Magongnaru Station, Exit 5, although their actual traffic volume was not high in 2022, the calculated Q' values were relatively high, and in the 2025 data their actual traffic also rose to the top ranks. Similarly, Magongnaru Station, Exit 3 was outside the top 10 in 2022, but because its Q' value was high, it ranked within the top 5 in 2025. However, it should be noted that the May 2025 data used for comparison included approximately 100 additional stations compared to 2022, so it cannot be regarded as a complete ground truth. Additionally, the Q' values did not precisely predict the actual future data, so methodological limitations still exist. Nevertheless, it was confirmed that the Q' matrix, which reflects spatial–temporal information, can partly explain future usage patterns, and this is considered to have sufficient significance for use as policy reference material.

Table 10. Top 10 bicycle stations by OD results: Comparison between May 2022 and May 2025.

May 2022 (Results of This Study)			May 2025 (Reference Data)	
Bicycle Station Name	Q Total	Q' Total	Bicycle Station Name	Q Total
In front of Ttukseom Resort Station, Exit 1	22,268	10.193	Hangang Park Mangwon Entrance	15,740
In front of Yeouinaru Station, Exit 1	20,965	6.864	Magongnaru Station, Exit 2	15,575
Hangang Park Mangwon Entrance	20,851	12.656	In front of Ttukseom Resort Station, Exit 1	13,137
Magongnaru Station, Exit 2	17,839	21.467	Lotte World Tower	11,499
Bongnimgyo Traffic Island	14,777	13.376	Magongnaru Station, Exit 3	10,211
Lotte World Tower	14,558	10.608	Behind Magongnaru Station, Exit 5	10,179
Sindaebang Station, Exit 2	11,918	11.906	Near Balsan Station, Exits 1 and 9	8429
In front of Guro Digital Complex Station	11,679	9.691	Olympic Park Station, Exit 3	8361
Olympic Park Station, Exit 3	11,374	12.732	In front of Yeouinaru Station, Exit 1	8313
Behind Magongnaru Station, Exit 5	10,870	19.925	Yeongdeungpo-gu Office Station, Exit 1	8227

4. Discussion

The bicycle-specialized functional area extraction results considering spatial–temporal information can provide insights in various fields, unlike conventional urban functional area classifications. First, from the perspective of bicycle users, the results of this study allow intuitive identification of bicycle-friendly areas and help recognize major functional areas by purpose, such as leisure, commuting, and shopping. Particularly, as both static and dynamic information were reflected in the functional area classification, the results can be more intuitively understood even by general users. Second, from the perspective of policymakers, our findings can serve as a basis for setting priorities in the allocation of bicycle infrastructure. Particularly, the Q' -matrix-based results make it possible to identify areas with low current demand but high latent demand, thereby enabling proactive infrastructure investment and policy intervention. Furthermore, the proposed approach is not limited to bicycle policy, it could also be utilized as useful reference material in urban planning at a macro level, such as land-use planning or the functional reorganization of urban spaces. For example, during the updating the Q matrix with the Q' matrix supplemented by spatial–temporal information, the results may vary dynamically depending on the characteristics of the urban space around bicycle stations. Thus, when public transport facilities such as subways are expanded or parks are created near stations, those POIs act as favorable elements for bicycle use, increasing the Q' matrix values. Therefore, the findings of this study hold significant meaning not only as a direct tool for bicycle policy formulation but also as a decision-making support tool for urban planners. Third, these

results can also be applied to bicycle-related service operations. They can be directly used to improve operational efficiency in rebalancing bicycle supply and retrieval, and they are also useful for service expansion strategies such as customized pricing schemes or new service planning. This provides practical implications for both the public and private sectors, including shared bicycle operating companies.

Despite these major advantages, this study has a limitation in that it did not consider actual bicycle GPS data on movement paths, but instead set units by partitioning the road network based on station location information alone. The left side of Figure 9 shows the proposed method, in which the road network is partitioned (red lines) through a network Voronoi diagram based on station location information, and the related POIs are aggregated accordingly to compile static spatial information. However, actual bicycle use does not occur uniformly across all road networks. The right side of Figure 9 presents hypothetical GPS data (yellow dots) as an example, showing that in reality, usage is concentrated in specific sections such as bicycle roads along the Han River, while in residential areas such as apartment complexes it is likely to be relatively low. In this study, due to restrictions in GPS data acquisition, only station-based OD data were utilized; therefore, these movement path characteristics were not sufficiently reflected. Nevertheless, the analysis was conducted by considering not only static spatial information but also dynamic information in the form of bicycle traffic volume at each station. Consequently, the station in Figure 9 was classified into Cluster 3, which is a bicycle-specialized area. In future work, the functional area extraction accuracy and interpretability needs to be enhanced by incorporating actual GPS-based movement path data.

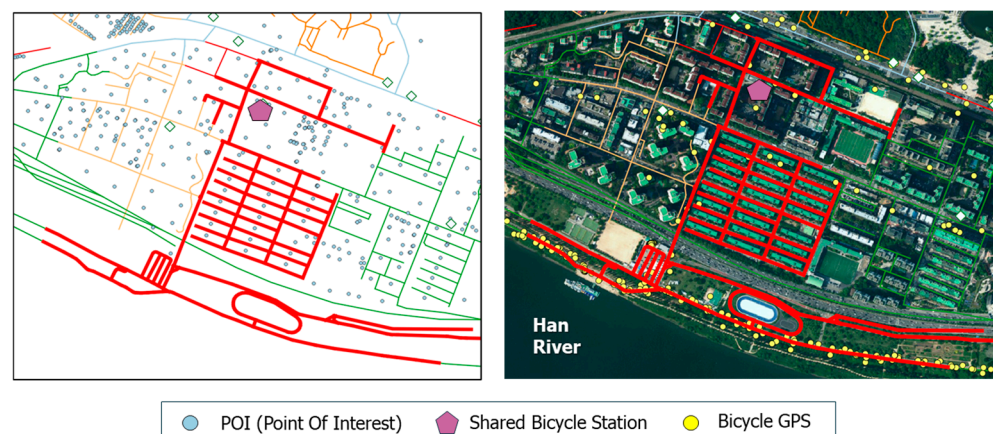


Figure 9. Comparison of considering actual bicycle paths.

5. Conclusions

Research on data-driven functional area identification has become specialized, necessitating tailored approaches. Unlike conventional studies that identified urban functional areas from a general perspective, this study proposes a functional area identification methodology from the perspective of bicycle users, enhancing practical applicability. The three core stages of functional area identification—spatial unit definition, data selection, and methodological framework—are designed specifically for bicycle use.

For spatial unit definition, NVD algorithms partitioned the road network into line-based units accessible from bicycle stations, enabling the aggregation of diverse spatial datasets. Dynamic (bicycle trip) and static (POI) information were jointly utilized to capture the spatial–temporal characteristics of bicycle use, avoiding reliance on a single dataset. To effectively integrate multimodal data, a semantic-based hybrid fusion method was introduced. Co-Matrix Factorization combined temporal (time-series travel patterns)

and spatial (POI-based semantics) information into the original OD matrix, producing an enhanced OD matrix (Q') from which latent functional areas were identified.

The proposed approach was applied to 2628 public bicycle stations in Seoul, republic of Korea, classifying the city into five functional areas: (i) Residential-Oriented, Medium-Flow with Strong Latent Zone, (ii) Dining-Industrial Mixed, Medium-Flow with Weak Latent zone, (iii) Multi-Functional Core, High-Flow with Strong Latent zone, (iv) Dining-Oriented, Medium-Flow with Moderate Latent zone, and (v) Industrial-Green Mixed, Low-Flow with Weak Latent zone. The proposed semantic-based fusion method reflects both spatial and temporal information. Although there is no ground truth data that can accurately identify this, our results were verified through visual evaluation using satellite imagery and comparison with reference data for diagnosing latent potential. These findings confirmed that the proposed method effectively delineated areas reflecting bicycle use characteristics.

These results demonstrate that functional areas, which are not easily identified using single-source or static datasets such as land use, can be effectively extracted through semantic fusion with dynamic information. By incorporating latent connectivity beyond observed trip volumes, the proposed method reveals bicycle-friendly areas that may have low current usage but high potential affinity. This offers substantial policy implications: even in areas with low present bicycle use, potential demand can be identified by considering environmental context and mobility pattern similarities, allowing planners to anticipate future demand growth. Such insights are crucial for decisions related to new station placement, bicycle infrastructure expansion, and urban mobility policies.

Despite these advantages, this study has the following limitations. First, the various factors that influence bicycle usage behavior were not sufficiently reflected. Unlike general means of transportation, bicycles are greatly affected by weather, seasonal and climatic conditions, topographical features, and demographic characteristics such as age and gender. However, as this study focused on developing a functional area identification methodology by integrating spatial-temporal information, these factors were not sufficiently considered. In fact, Cluster 5 was classified as a low-bicycle-usage area; however, the static information showed a high proportion of green space. This could be due to the fact that the green spaces included in this cluster are mostly located in highland areas, which do not have a positive impact on bicycle traffic. Second, the proposed methodology can only be applied to dock-based shared bicycle systems. In many cities, various forms of micromobility such as dock-less shared bicycles or electric scooters are currently being operated; however, our methodology relies on bicycle stations as spatial generators, making it difficult to apply directly to these transportation modes. Furthermore, as only OD data between stations, rather than GPS data, were used to define spatial units, actual bicycle riding paths were not sufficiently reflected. Third, the Co-Matrix Factorization approach proposed in this study is based on the semantic integration of information across datasets, and thus does not fully capture the topological structure among bicycle stations. In reality, bicycle stations tend to interact with spatially adjacent stations; however, such spatial interactions were not explicitly considered in the model. Moreover, the results may vary depending on the quality and completeness of the OD and POI datasets, implying the potential presence of data bias.

Therefore, in future work, the factors affecting bicycle usage behavior need to be more broadly analyzed and applied so that the results can reflect robust and representative characterization of actual bicycle use. Particularly, if GPS-based movement path data are incorporated, it could be possible to encompass various types of shared mobility and reflect actual riding characteristics more precisely. In addition, when designing data fusion models, methodological extensions that can incorporate the topological structure of bicycle stations and road networks should be considered. Finally, the proposed methodology could

be further enhanced in terms of its practical applicability and policy contribution if the validation methods are improved through quantitative verification such as user surveys and related urban planning data.

Author Contributions: Conceptualization, J.L. and J.K.; methodology, J.L. and J.K.; software, J.L.; validation, J.K.; formal analysis, J.L.; writing—original draft preparation, J.L.; writing—review and editing, J.K.; visualization, J.L.; supervision, J.K.; project administration, J.K.; funding acquisition, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2022-00143804).

Data Availability Statement: The data presented in this paper are available upon reasonable request from the corresponding author.

Acknowledgments: During the preparation of this manuscript the authors used ChatGPT (version GPT-5, OpenAI) to enhance its language clarity. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhai, W.; Bai, X.; Shi, Y.; Han, Y.; Peng, Z.-R.; Gu, C. Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Comput. Environ. Urban Syst.* **2019**, *74*, 1–12. [\[CrossRef\]](#)
2. Chen, T.-H.K.; Qiu, C.; Schmitt, M.; Zhu, X.X.; Sabel, C.E.; Prishchepov, A.V. Mapping horizontal and vertical urban densification in Denmark with Landsat time-series from 1985 to 2018: A semantic segmentation solution. *Remote Sens. Environ.* **2020**, *251*, 112096. [\[CrossRef\]](#)
3. Du, S.; Du, S.; Liu, B.; Zhang, X. Mapping large-scale and fine-grained urban functional zones from VHR images using a multi-scale semantic segmentation network and object based approach. *Remote Sens. Environ.* **2021**, *261*, 112480. [\[CrossRef\]](#)
4. Hu, S.; He, Z.; Wu, L.; Yin, L.; Xu, Y.; Cui, H. A framework for extracting urban functional regions based on multiprototype word embeddings using points-of-interest data. *Comput. Environ. Urban Syst.* **2020**, *80*, 101442. [\[CrossRef\]](#)
5. Li, M.; Zhu, Y.; Zhao, T.; Angelova, M. Weighted dynamic time warping for traffic flow clustering. *Neurocomputing* **2022**, *472*, 266–279. [\[CrossRef\]](#)
6. Yang, H.; Peng, J.; Zhang, Y.; Luo, X.; Yan, X. Understanding the spatiotemporal impacts of the built environment on different types of metro ridership: A case study in Wuhan, China. *Smart Cities* **2023**, *6*, 2282–2307. [\[CrossRef\]](#)
7. Yang, X.; Bo, S.; Wang, J. Classifying urban functional zones by integrating the homogeneity and structural similarity of POIs. *J. Urban Plan. Dev.* **2024**, *150*, 04024052. [\[CrossRef\]](#)
8. Gao, Q.; Fu, J.; Yu, Y.; Tang, X. Identification of urban regions' functions in Chengdu, China, based on vehicle trajectory data. *PLoS ONE* **2019**, *14*, 0215656. [\[CrossRef\]](#)
9. Liu, X.; Tian, Y.; Zhang, X.; Wan, Z. Identification of urban functional regions in Chengdu based on taxi trajectory time series data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 158. [\[CrossRef\]](#)
10. Zhang, P.; Yang, M.; Wang, Y.; Yang, T.; Yu, H.; Yan, X. Integrating metro passenger flow data to improve the classification of urban functional regions using a heterogeneous graph neural network. *Int. J. Digit. Earth* **2024**, *17*, 2443468. [\[CrossRef\]](#)
11. Du, S.; Zhang, X.; Lei, Y.; Huang, X.; Tu, W.; Liu, B.; Meng, Q.; Du, S. Mapping urban functional zones with remote sensing and geospatial big data: A systematic review. *GISci. Remote Sens.* **2024**, *61*, 2404900. [\[CrossRef\]](#)
12. Chang, X.; Wu, J.; He, Z.; Li, D.; Sun, H.; Wang, W. Understanding user's travel behavior and city region functions from station-free shared bike usage data. *Transp. Res. Part F Traffic Psychol. Behav.* **2020**, *72*, 81–95. [\[CrossRef\]](#)
13. Zhao, J.; Fan, W.; Zhai, X. Identification of land-use characteristics using bicycle sharing data: A deep learning approach. *J. Transp. Geogr.* **2020**, *82*, 102562. [\[CrossRef\]](#)
14. Lee, J.; Yu, K.; Kim, J. Public bike trip purpose inference using point-of-interest data. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 352. [\[CrossRef\]](#)
15. Qian, Z.; Liu, X.; Tao, F.; Zhou, T. Identification of urban functional areas by coupling satellite images and taxi GPS trajectories. *Remote Sens.* **2020**, *12*, 2449. [\[CrossRef\]](#)
16. Hu, S.; Gao, S.; Wu, L.; Xu, Y.; Zhang, Z.; Cui, H.; Gong, X. Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. *Comput. Environ. Urban Syst.* **2021**, *87*, 101619. [\[CrossRef\]](#)

17. Niu, H.; Silva, E.A. Delineating urban functional use from points of interest data with neural network embedding: A case study in Greater London. *Comput. Environ. Urban Syst.* **2021**, *88*, 101651. [\[CrossRef\]](#)
18. Deng, Y.; He, R. Refined Urban Functional Zone Mapping by Integrating Open-Source Data. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 421. [\[CrossRef\]](#)
19. Jing, C.; Zhang, H.; Xu, S.; Wang, M.; Zhuo, F.; Liu, S. A hierarchical spatial unit partitioning approach for fine-grained urban functional region identification. *Trans. GIS* **2022**, *26*, 2691–2715. [\[CrossRef\]](#)
20. Jing, C.; Hu, Y.; Zhang, H.; Du, M.; Xu, S.; Guo, X.; Jiang, J. Context-aware matrix factorization for the identification of urban functional regions with POI and taxi OD data. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 351. [\[CrossRef\]](#)
21. Qin, Q.; Xu, S.; Du, M.; Li, S. Identifying urban functional zones by capturing multi-spatial distribution patterns of points of interest. *Int. J. Digit. Earth* **2022**, *15*, 2468–2494. [\[CrossRef\]](#)
22. Yang, M.; Kong, B.; Dang, R.; Yan, X. Classifying urban functional regions by integrating buildings and points-of-interest using a stacking ensemble method. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102753. [\[CrossRef\]](#)
23. Liu, T.; Cheng, G.; Yang, J. Multi-scale recursive identification of urban functional areas based on multi-source data. *Sustainability* **2023**, *15*, 13870. [\[CrossRef\]](#)
24. Luo, G.; Ye, J.; Wang, J.; Wei, Y. Urban functional zone classification based on POI data and machine learning. *Sustainability* **2023**, *15*, 4631. [\[CrossRef\]](#)
25. Wang, Z.; Bai, J.; Feng, R. A multi-feature fusion method for urban functional regions identification: A case study of Xi'an, China. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 156. [\[CrossRef\]](#)
26. Zhang, Y.; Xu, Y.; Gao, J.; Zhao, Z.; Sun, J.; Mu, F. Urban functional zone identification based on multimodal data fusion: A case study of Chongqing's central urban area. *Remote Sens.* **2025**, *17*, 990. [\[CrossRef\]](#)
27. Wu, Y.; Li, Y. "Hot street" of crime detection in London borough and lockdown impacts. *Geo-Spat. Inf. Sci.* **2023**, *26*, 716–732. [\[CrossRef\]](#)
28. Liu, S.; Zhou, C.; Rong, J.; Bian, Y.; Wang, Y. Concordance between regional functions and mobility features using bike-sharing and land-use data near metro stations. *Sustain. Cities Soc.* **2022**, *84*, 104010. [\[CrossRef\]](#)
29. Campbell, A.A.; Cherry, C.R.; Ryerson, M.S.; Yang, X. Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. *Transp. Res. Part C Emerg. Technol.* **2016**, *67*, 399–414. [\[CrossRef\]](#)
30. She, B.; Zhu, X.; Ye, X.; Guo, W.; Su, K.; Lee, J. Weighted network Voronoi Diagrams for local spatial analysis. *Comput. Environ. Urban Syst.* **2015**, *52*, 70–80. [\[CrossRef\]](#)
31. Okabe, A.; Satoh, T.; Furuta, T.; Suzuki, A.; Okano, K. Generalized network Voronoi diagrams: Concepts, computational methods, and applications. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 965–994. [\[CrossRef\]](#)
32. Kang, C.; Qin, K. Understanding operation behaviors of taxicabs in cities by matrix factorization. *Comput. Environ. Urban Syst.* **2016**, *60*, 79–88. [\[CrossRef\]](#)
33. Kang, C.; Shi, L.; Wang, F.; Liu, Y. How urban places are visited by social groups? Evidence from matrix factorization on mobile phone data. *Trans. GIS* **2020**, *24*, 1504–1525.
34. Lee, C.K.H.; Leung, E.K.H. Spatiotemporal analysis of bike-share demand using DTW-based clustering and predictive analytics. *Transp. Res. Part E Logist. Transp. Rev.* **2023**, *180*, 103361. [\[CrossRef\]](#)
35. Zhang, J.; Li, X.; Yao, Y.; Hong, Y.; He, J.; Jiang, Z.; Sun, J. The Traj2Vec model to quantify residents' spatial trajectories and estimate the proportions of urban land-use types. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 193–211. [\[CrossRef\]](#)
36. Yang, X.; Yang, Y.; Zheng, X. Classifying urban functional zones by integrating POIs, Place2vec, and LDA. *J. Urban Plan. Dev.* **2023**, *149*, 04023034. [\[CrossRef\]](#)
37. Yang, X.; Jiao, H.; Wang, J. Classifying Urban Functional Zones by Integrating Place2Vec and GCN. *J. Urban Plan. Dev.* **2025**, *151*, 04025008. [\[CrossRef\]](#)
38. Yan, B.; Janowicz, K.; Mai, G.; Gao, S. 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, Redondo Beach, CA, USA, 7–10 November 2017; Hoel, E., Newsam, S., Rawada, S., Tamassia, R., Trajcevski, G., Eds.; Association for Computing Machinery: New York, NY, USA, 2017; Volume 35, pp. 1–10.
39. Zhao, Y.; Ke, S.; Lin, Q.; Yu, Y. Impact of land use on bicycle usage. *J. Transp. Land Use* **2020**, *13*, 299–316. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.