Embedding Spatial and Semantic Contexts for Geo-Entity Typing in Smart City Applications

Basel Shbita
USC Information Sciences Institute
Marina del Rey, California, USA
shbita@isi.edu

Binh Vu
USC Information Sciences Institute
Marina del Rey, California, USA
binhvu@isi.edu

Fandel Lin
USC Information Sciences Institute
Marina del Rey, California, USA
fandelli@isi.edu

Craig A. Knoblock
USC Information Sciences Institute
Marina del Rey, California, USA
knoblock@isi.edu

Abstract

Geospatial data are critical for urban planning and smart city applications, yet understanding and classifying geo-entities in diverse datasets remains challenging. Accurate representation and classification of geo-entities are essential for tasks such as geo-entity typing and linking, enabling better map understanding and informed decision-making. This paper presents a self-supervised learning approach to classify geo-entities by embedding their geometric, spatial, and semantic neighborhood contexts, creating robust representations for geo-entity typing. Using *OpenStreetMap* (OSM) data, our method links geo-referenced entities to *Wikidata* classes and OSM tags with high performance, achieving an F_1 score of approximately 0.85. Beyond the technical contribution, our method addresses Responsible AI challenges, including transparency, and data standardization on the Web, aligning with sustainable smart city development.

CCS Concepts

• **Computing methodologies** → **Knowledge representation and reasoning**; **Neural networks**; • **Information systems** → **Geographic information systems**.

Keywords

Semantic typing; Geospatial data integration; Representation learning; Open data; Web technologies; Digital twin; Smart cities

ACM Reference Format:
Basel Shbita, Binh Vu, Fandel Lin, and Craig A. Knoblock. 2025. Embedding Spatial and Semantic Contexts for Geo-Entity Typing in Smart City Applications. In *Companion Proceedings of the ACM Web Conference 2025 (WWW Companion '25), April 28-May 2, 2025, Sydney, NSW, Australia.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3701716.3718325

1 Introduction

The increasing availability of digitized geospatial data is transforming urban development, governance, and public services in modern smart cities, as well as advancing research in the social and natural sciences [3, 9, 18, 41]. Advances in technology have enabled the extraction of structured vectorized geospatial information from open data sources such as *OpenStreetMap* (OSM) and scanned maps [11, 28, 38, 40]. These datasets provide a rich foundation for digital twin technologies, automated city planning, and AI-powered decision-making in smart urban environments [2].

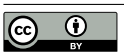
Despite these advancements, accurately classifying, linking, and integrating geo-entities remains a significant challenge, mainly when dealing with heterogeneous, large-scale urban and spatial data. Understanding the spatial and semantic contexts of entities such as roads, buildings, and natural features is crucial for applications ranging from infrastructure monitoring to environmental sustainability [8, 39]. Figure 1 illustrates different types of vectorized geo-entities extracted from urban datasets, emphasizing the variation in their shapes and spatial footprints.

A major challenge in geospatial AI applications is the lack of standardized entity classification methods incorporating spatial and semantic contexts [30]. This challenges data integration systems that require automatic understanding, such as those that involve digitized maps and remote sensing data [25, 35, 39].

The Web plays a pivotal role in the sharing and standardizing geospatial knowledge. Studies are exploring the abundant geospatial information available online for improved understanding of data and linking of entities with geospatial entities on the Web [5, 25, 34]. *OpenStreetMap*[1] (OSM), emerging as a significant open knowledge base on the Web, houses an expansive repository of crowd-sourced geospatial data obtained through collective efforts of an extensive network of contributors. OSM provides structured yet non-ontologized tagging systems, offering valuable insights into urban features and land use through community-driven annotations. Figure 2 shows an example of a geo-instance on OSM with the user-assigned tags `natural=water` and `water=reservoir`.

To address the challenge of fair, transparent, and scalable geo-entity classification, we introduce a self-supervised representation learning approach that embeds geometric, spatial, and semantic contexts of geo-entities. Our method leverages OSM data and links geo-referenced entities to structured knowledge bases, such as *Wikidata* [42], ensuring semantic consistency and interoperability. Unlike conventional GIS methods that rely on predefined feature engineering, our embedding model learns latent geo-entity representations based on their spatial configurations and neighboring

[1] https://www.openstreetmap.org/



WWW Companion '25, Sydney, NSW, Australia

ACM ISBN 979-8-4007-1331-6/2025/04
https://doi.org/10.1145/3701716.3718325

(a) Bodies of water (b) Waterways (c) Residential buildings

Figure 1: Examples of geo-instance shapes and footprints, encoded as vector data and categorized by type.

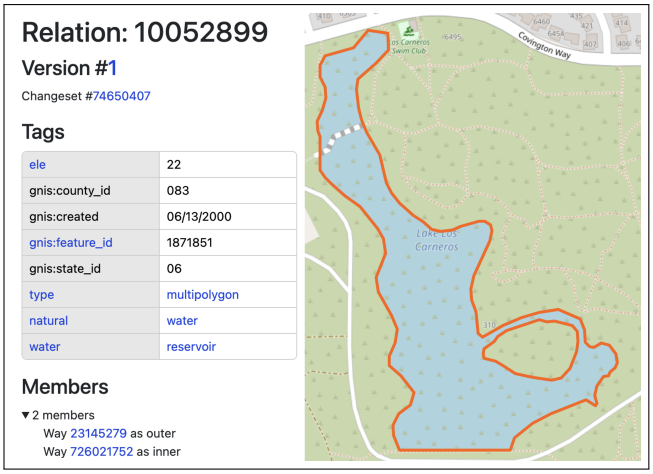


Figure 2: An *OpenStreetMap* instance depicting a geographic feature labeled with tags `natural=water` and `water=reservoir`, offering vital crowd-sourced information for data understanding, structuring, and integration.

features, reducing bias from manual feature selection and improving model interpretability.

The representation of a geo-entity faces challenges from its multi-dimensional nature, including its proximity to various other geo-entities. For example, the location of a building relative to other structures, roads, or green spaces and natural features like rivers can influence its function and size. However, AI-driven geospatial models in smart cities face critical fairness, accountability, and transparency challenges. Since OSM is community-driven, its data coverage and labeling quality vary across regions, potentially introducing biases in classification outcomes. Ensuring responsible AI in smart city applications requires methods to mitigate these biases, enforce consistency, and enhance interpretability. Our approach addresses these concerns by leveraging structured taxonomies to maintain semantic consistency and improve model interpretability through contrastive learning techniques.

Embedding-based methods have been successfully applied in natural language processing (e.g., word embeddings [29, 31], sentence embeddings [33]) and computer vision (e.g., CNNs for image classification [13, 23]). Inspired by these advances, we apply self-supervised contrastive learning [12, 24] to geospatial representation learning, optimizing geo-entity embeddings for semantic typing and AI-driven urban analytics.

In this paper, we make the following contributions:

(1) We introduce a novel self-supervised embedding method for geo-entities that combines geometric, spatial, and semantic contexts. We employ open data from the web, particularly OSM, to characterize the geo-entity context.
(2) We address responsible AI challenges in smart city applications by incorporating structured taxonomies to mitigate bias in geo-entity classification. Our taxonomy-aware contrastive learning framework enhances both classification accuracy and interpretability.
(3) We conduct extensive experiments on real-world datasets, demonstrating high performance in linking geo-referenced entities to *Wikidata* classes and OSM tags. We also make our source code and data publicly available[2] as a contribution to the broader research community.

Our approach contributes to the broader goal of creating web-based intelligent infrastructures that support the next generation of human-centric smart cities, where AI and automation enhance livability while preserving data accountability and ethical decision-making.

2 Embedding Method

The task at hand involves geospatial entity embedding and representation learning to enable geo-entity typing and classification. This paper focuses on classifying geo-referenced entities represented in Well-Known Text (WKT), a widely used format for encoding geometric objects [14]. However, our approach is not limited to WKT, as it can be extended to any digitized geo-referenced data format, as long as the spatial information is structured and available. Ultimately, we aim to classify these entities into a set of semantic types within a given dataset, leveraging their geometric, spatial, and contextual attributes for robust classification. Figure 3 shows a visualization of the embedding architecture operating over the input data in WKT format.

[2]https://github.com/basels/GeoEntityContextNet

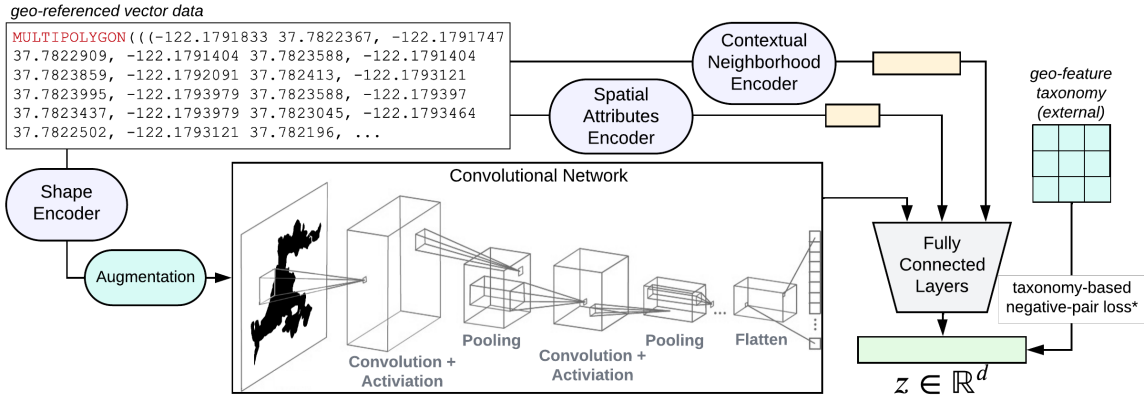


Figure 3: Illustration of the geo-entity encoding and embedding architecture, integrating shape, spatial, and neighborhood information, with auxiliary components depicted in blue and the resulting output, representing the latent vector, in green.

2.1 Representation Learning Model

Our architecture encompasses three main components: the shape encoder, the spatial attributes encoder, and the contextual neighborhood encoder, as depicted in Figure 3. The shape encoder aims to capture the geometric characteristics of the geospatial object, while the spatial attributes encoder extracts measurable attributes, such as area and length, employing standard spatial computational methods. Conversely, the neighborhood encoder generates a feature vector based on the semantic types of the neighboring geo-entities relative to the entity under consideration. To comprehensively represent the geospatial entity, we utilize the output from each encoder to train the primary embedding model.

2.1.1 Extracting Geometric and Spatial Features. Our approach holistically encodes the geo-entity's shape information, ensuring that it is not constrained by memorizing the positions of training examples. Addressing the heterogeneity and variable length of vector data, we generate a "footprint" outline for each entity to learn its shape characteristics. The WKT representation is discretized into a fixed 200×200 binary two-dimensional array, serving as a single-channel binary raster, and identified as the minimum resolution that adequately depicts lines and multi-lines, rendering visually perceptible. As depicted in Figure 3, we incorporate augmentation during training by applying various transformations, such as resizing, sharpness adjustments, rotations, and flips, to enhance model robustness and enable generalization across diverse shapes.

Simultaneously, spatial attribute encoding extracts and integrates size and length properties of geo-entities using established geospatial tools.[3] This step is crucial for the final embedding model, as the geometric shape encoding component captures only relative form while ignoring absolute scale. By incorporating shape and spatial dimensions, our approach ensures a more comprehensive representation of geo-entities, preserving critical distinctions.

2.1.2 Neighborhood Contextual Semantic Encoding. To materialize the neighborhood context of a specified geo-entity, our encoder embeds the relative positions of each neighboring feature for the target entity, ensuring comprehensive encapsulation of the data.

[3]https://shapely.readthedocs.io/en/stable/

Figure 4 provides a visual illustration, showing an anchor feature (e.g., `school`, highlighted in orange), and its neighborhood context — a collection of geo-features surrounding it at varying distances and with different type labels shown in different colors. We employ a "bag-of-features" vector encoding to capture the spatial relationships among the geospatial entities, using a distance-based encoding method to generate a "bag-of-distances" feature vector. This feature vector encodes the relative shortest distance to every recognized geo-type within the neighborhood, preserving relative distance and directionality information between entities. Along with the geometric and spatial features, it serves as an additional input to the model.

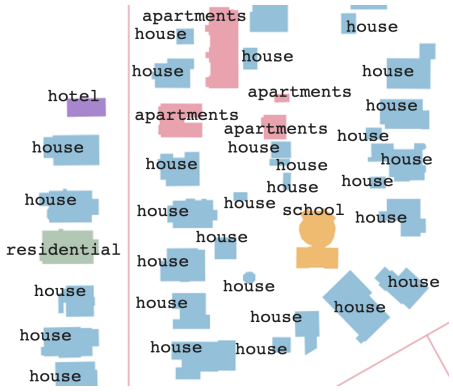


Figure 4: Illustration of a neighborhood, with anchor entity (`school`) in orange with surroundings features `house` in blue; `apartments` in red; `residential` in green; `hotel` in purple.

A spatial knowledge base or database are essential to comprehensively construct a neighborhood encoding. The knowledge base merely fetches the entities in the context "window". In this work, we specifically utilized OSM to retrieve neighboring geo-instances — including nodes, ways, and relations — within a defined distance threshold (a model hyperparameter) from the geo-referenced center of the entity.

2.2 Taxonomy-Guided Contrastive Learning

Integrating taxonomic information about geo-feature types into the learning framework provides an auxiliary tool for encoding rich semantic knowledge. This taxonomic knowledge, often structured as an ontology, not only enables systematic classification but also enhances model generalization. By distinguishing between semantically related geo-features, such as `commercial` vs. `residential` buildings or `motorway` vs. `primary` highways, our approach improves contrastive learning by identifying meaningful "negatives". Furthermore, recognizing hierarchical relationships, such as `beach` being a subtype of `natural`, refines vector representations by incorporating semantic similarities and dissimilarities at different levels of granularity.

Navigating the hierarchical maze of OSM tags presents a unique challenge due to their inconsistent granularity and dynamic nature. To address this, rigorous filtering is required to select meaningful labels for taxonomy-guided self-supervised training. We build upon methodologies proposed by Dsouza et al. [10] and Shbita et al. [34] to develop a structured taxonomy of OSM labels. The taxonomy organizes geo-feature categories into a multi-level hierarchy, mapping them to corresponding *Wikidata* classes.

We incorporate the taxonomy weights into the Normalized Temperature-scaled Cross Entropy loss function [7, 36], which we define as follows. For each anchor entity e_q in a given batch, the taxonomy-aware loss is calculated with respect to the positive and negative samples in the set, and is given by:

$$L_q = -\log \frac{\exp(sim(e_q, e_+)/\tau)}{\sum_{i=0}^{K} \exp(sim(e_q, e_i) \cdot w_{q,i}/\tau)} \tag{1}$$

where e_+ is a positive sample, $sim(e_i, e_j)$ is the cosine similarity between the normalized embeddings of entities e_i and e_j. The temperature τ scales the similarity scores. The sum is over one positive and K negative samples. $w_{q,i}$ is the weight representing the taxonomic distance of labels between e_q and a negative sample e_i. The taxonomic weight $w_{i,j}$ is defined by the relative distance within the taxonomy tree as:

$$w_{i,j} = \frac{d_{\text{tree}} - d_{i,j}}{d_{\text{tree}}} \tag{2}$$

where d_{tree} is the depth of the taxonomy tree, and $d_{i,j}$ is the depth of the common ancestor of entities i and j. This normalization ensures that weights adjust the influence of negative samples in the loss function to reduce the penalty of misclassifying entities to similar but still incorrect classes.

In the grand scheme of the embedding model, three distinct data inputs are combined to train a mapping function. This function learns to differentiate various geo-instances in a low-dimensional vector space based on their respective types in the taxonomy. The resultant embeddings can drive a classifier that effectively discriminates between target semantic types, as demonstrated in Section 3.

3 Evaluation

We evaluate the effectiveness of our proposed geo-entity embedding approach by training a model under various settings of our methodology and comparing it to two baselines, including the state-of-the-art (SotA) in geo-entity embedding. Using two distinct datasets, each model was evaluated through a classification and semantic typing task. The objective is to explore how different types of information affect the performance of our approach as an ablation study and to test our best-performing model against other systems, aiming to gain insights into the generalizability of the model and its proficiency in the overall task of semantic typing.

3.1 Experiment Setup

3.1.1 Data. Consistently across all settings, our model was trained using the same data, which encompassed 200,000 OSM instances from the California OSM snapshot[4]. We utilized linear and polygonal features whilst excluding discrete point-based features.[5] Currently, this comprehensive dataset encapsulates around 150 million instances, of which about 10 million contain at least one tag. Instances were tagged with 1 to 16 labels, resulting in an average of 2.3 tags per instance. While the dataset originally featured over 3,000 unique OSM tags, this was filtered down to 75 following the process described in Section 2.2.

The classification test datasets utilized were crafted by separately sampling from OSM. We ensured that the geo-instances in the test datasets were not present in the training data. The first dataset, WD-2k, comprises 2,146 instances with direct mapping to their *Wikidata* classes (based on the *Wikidata* instance labeled by OSM users), covering 11 distinct classes. The second dataset, OSM-16k, consists of 16,059 instances that span 18 OSM "classes" (most fine-grained tag per instance). Both datasets are publicly available via our repository.

The resulting embedding was tested using Support Vector Classification, which rendered the best results compared with other classifiers like Random Forest, K-Nearest Neighbors, and Logistic Regression. Model evaluation was measured in precision, recall, and F_1 scores, utilizing 8-fold cross-validation to divide the data into mutually exclusive subsets (87.5% training; 12.5% testing).

3.1.2 Experimental Settings. We evaluate our model performance under varying conditions using four variant settings. The first setting focused solely on shape information, excluding any neighborhood information or spatial attributes, while the second setting incorporated both shape and spatial data, adding a spatial encoder to include its area and length. We included shape, spatial, and contextual neighborhood data in the third setting but did not consider taxonomic relations. The fourth and final setting extended the third setting by adding taxonomic data to further enhance model performance.

3.1.3 Model Training. We determined the hyperparameters of the model systematically through an iterative process of experimentation. We found that within our dataset, the optimal neighborhood size was around 15" degrees (equivalent to approximately 450 meters or 1,500 ft) for the task of semantic typing. We chose a learning rate of 10^{-5} and weight decay of 0.05 to enable model stability throughout training. To accommodate the computational

[4] https://download.geofabrik.de/north-america/us/california.html

[5] Point features were excluded as they lack geometric or spatial value due to their zero-dimensional nature, making it impossible to measure length, area, or shape. Unlike linear and polygonal features, which represent physical entities with spatial extent, points typically denote locations or place names, making them unsuitable for our evaluation.

Table 1: Summary of results for semantic-type classification in all experimental settings, across both datasets

		WD-2k			OSM-16k		
Setting		Precision	Recall	F_1	Precision	Recall	F_1
1	$\text{Ours}_{\texttt{shape}}$	0.497	0.506	0.501	0.473	0.512	0.492
2	$\text{Ours}_{\texttt{shape+spatial}}$	0.506	0.545	0.525	0.491	0.536	0.513
3	$\text{Ours}_{\texttt{full}}$	**0.850**	0.823	0.836	**0.877**	0.725	0.794
4	$\text{Ours}_{\texttt{full w/taxonomy}}$	0.849	**0.852**	**0.850**	0.858	**0.854**	**0.856**
	GPT-3.5-Turbo	0.198	0.209	0.121	0.145	0.063	0.026
	GeoVectors [37]	0.819	0.834	0.826	0.833	0.815	0.824

constraints of the available hardware resources, which included four NVIDIA GeForce RTX 2080 Ti GPUs and an Intel i7 CPU, providing 4,352 cores and 11 GB DDR6 memory per GPU, the batch size was established at 32. The model was trained for 100 epochs. Additionally, the hyperparameter d, representing the dimensionality of the latent vector, was set to 300, to enable fair evaluation with the SotA model of this dimensionality.

3.1.4 Baselines. To establish robust baselines for our study, we include two additional settings. First, we utilize GeoVectors [37] as a baseline, a pre-trained corpus of OSM embeddings, given its standing as the nearest SotA model trained to navigate analogous challenges of embedding geo-entities. GeoVectors was trained by leveraging two models: a neural location model for spatial relations and a pre-trained word embedding model to encode semantic similarities based on tags. Furthermore, we explore the capabilities of Large Language Models (LLMs) in a zero-shot classification setting to assess their performance with geographic data. Specifically, we used natural language queries to provide the transformer-based model, GPT-3.5 Turbo [1], with classification candidates and their descriptions alongside the geo-referenced input vector in its source WKT format to generate an answer regarding the semantic type.

3.2 Results and Discussion

We present the results of our experiments across the settings described above and discuss their implications for the effectiveness of our proposed method for semantic typing.

3.2.1 Overall Performance. Table 1 shows the results for each setting across both datasets. In our baseline, Setting 1, we solely relied on geometric shape data for classification, which resulted in F_1 scores of 0.501 for WD-2k and 0.492 for OSM-16k. Introducing the spatial attribute encoder in Setting 2, the scores elevated to 0.525 and 0.513, respectively.

Remarkably, performance was significantly boosted when both contextual neighborhood data and geo-entity type taxonomy were incorporated (Settings 3 and 4). Setting 4, which combines all these inputs, yielded the most impressive results: F_1 scores of 0.850 for WD-2k and 0.856 for OSM-16k. Interestingly, the peak precision was observed in Setting 3, where taxonomic data was omitted. This phenomenon suggests that in our non-guided contrastive learning, treating all negatives uniformly — as opposed to a weight-based approach in Setting 4 — results in finer distinctions between all entity types. This could be explained by the higher total (negative) loss per epoch, as observed in Setting 3 compared to Setting 4. These findings highlight how incorporating diverse data sources improves embedding quality and enhances classification performance.

A comparison with the SotA model shows that our model outperforms on both WD-2k and OSM-16k datasets. Notably, our model achieved better results on OSM-16k, where classification was aligned with OSM tags — a logical outcome given the model's training on this data source. This distinction is even more significant considering the added complexity in the OSM-16k task, with 18 classes versus 11 in WD-2k. However, the GPT-3.5 Turbo, in a zero-shot setting, scored lower with an F_1 score of 0.121 on the WD-2k dataset and only 0.026 on the OSM-16k dataset, highlighting challenges in adapting LLMs to spatial semantic tasks without a domain-specific and tailored training.

It is important to note that our model was constructed without embedding direct semantic information or OSM tags about the geo-entity in the self-learning process, focusing solely on geometric, spatial, and neighborhood contexts. In contrast, GeoVectors incorporated such semantic data, including *Wikidata* connections, subtly giving them an advantage. Ultimately, the results show our method's advantage.

3.2.2 Analysing the Optimal Setting. Figure 5 shows the per-class confusion matrix results for the WD-2k dataset utilizing our method's optimal setting (Setting 4). An initial analysis indicates that our model exhibits exceptional performance across most classes, notably achieving the highest scores for `light rail line`, `limited-access road`, and `stream`. This implies that our model more adeptly distinguishes linear features than polygon-based features. Various factors could contribute to the fact that `light rail line` secured the highest recall score among other linear features. This could be due to the distinctive geometric and spatial characteristics of light rail lines, which often display a "twisting", elongated shape and inhabit distinct environments compared to other linear features in urban areas.

There were some challenges in differentiating between particular classes. For instance, 47.8% of `school` features are misclassified as `high school`, and 27.6% as `park`. While `high school` (Q9826) and `school` (Q3914) are distinct in the labeling scheme, a human annotator might perceive one as a subclass of the other, rendering the task potentially redundant. The model's capability to classify and meaningfully capture numerous fine-grained `high school` instances is noteworthy, lending qualitative confidence to its ability to differentiate between unique types that may exist under a broader,

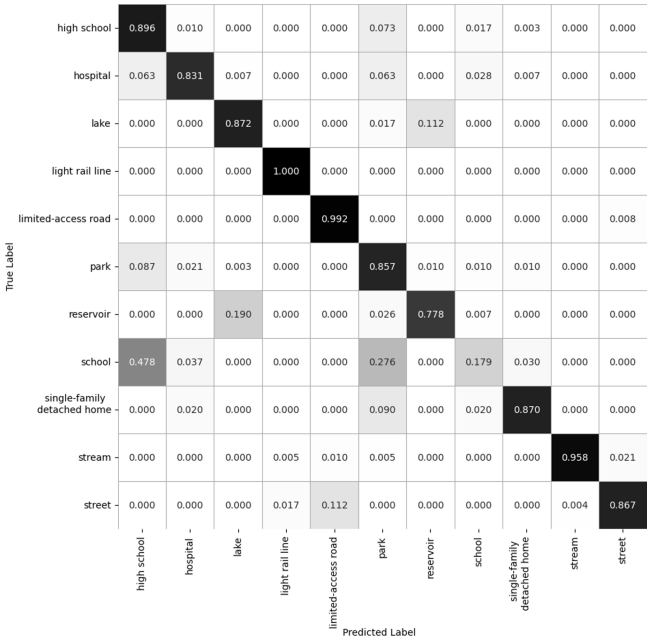


Figure 5: Confusion matrix illustrating classification results of geo-entities to *Wikidata* types using the WD-2k dataset, employing the model derived from the optimal setting (Setting 4). The matrix aggregates results across all mutually exclusive subsets of tests.

shared geo-feature taxonomy. The similarity between `school` and `park` may originate from their shared attributes (e.g., similar shape footprints and neighborhood environment), posing a challenge to accurate classification without further entity-specific knowledge.

Additional observations indicate that the model occasionally misinterprets `lake` instances as `reservoir`, a plausible error given the similar footprints and environmental roles of these water bodies. Likewise, `street` is confused with `limited-access road` 11.2% of the time, a mistake potentially stemming from geometric similarities and proximities to analogous geo-feature types.

Figure 6 shows the per-class confusion matrix results for the OSM-16k dataset utilizing our method's optimal setting (Setting 4). The results suggest that the model effectively captures the defining characteristics and contexts of almost all 18 geo-feature types, with particularly strong true positive rates indicated by the high scores (dark shades) along the diagonal. Certain classes, such as `track_leisure`, `beach`, and `golf_course`, show a high degree of predictive accuracy. However, some classes like `commercial_landuse` and `parking_amenity` demonstrate significant confusion with other amenity and building types, often being misclassified as `retail_building` and `retail_landuse`, respectively. This could indicate an overlap in the feature space or insufficient differentiation between these feature types.

The dashed outlines around entity clusters in the confusion matrix in Figure 6 represent groups with a common tag "ancestor", highlighting the taxonomic hierarchy. Notably, confusion between entities is more frequent within these clusters than between them, indicating the model's proficiency in distinguishing general tags (buildings vs. natural features) from closely related tags (building types). Future work could incorporate satellite imagery or aerial data to enhance land use differentiation and refine the classification of closely related geo-entities and features.

3.2.3 Visualizing the Latent Space. Furthering our understanding of the model's performance, we employed t-Distributed Stochastic Neighbor Embedding (t-SNE) to plot the embeddings of 10,000 geo-entities from OSM, as depicted in Figure 7. Figure 8 provides a comparative view to the detailed tags illustrated in Figure 7, showcasing labels of the identical data points at the highest level of the OSM tag taxonomy. Each figure displays notable separation among various classes. Evaluating ground truth labels at a higher taxonomy level unveils noteworthy clustering, supplying additional qualitative evidence supporting the model's generalizability. The t-SNE plot shows that the clusters representing distinct classes have minimal overlap, affirming the model's capacity to discern inherent patterns in the data. However, it is vital to note that t-SNE, a two-dimensional representation suited for visualizing high-dimensional datasets through dimensionality reduction, may incur some information loss during projection. Nonetheless, visualization serves as a valuable tool for assessing the quality of the embeddings and gaining insight into the interrelations among diverse classes. In summary, our proposed method for semantic embedding utilizing multi-faceted learning has yielded encouraging results, adeptly capturing spatial, geometric, and neighborhood information about geo-entities. The precision, recall, and F_1 scores, together with the confusion matrix and t-SNE visualization, show the strengths and potential areas for refinement within our method.

3.2.4 Interoperability and Smart City Applications. Our approach aligns with existing web-based geospatial standards, including OGC GeoSPARQL [4], by enabling interoperability with OSM and *Wikidata* vocabularies (which follows RDF and Web-compatible schemas). This alignment enhances cross-platform integration, ensuring that semantic representations are reusable across smart city services and enabling seamless data exchange between heterogeneous smart city infrastructures. As a result, our method supports AI-driven geospatial analytics in applications such as urban planning, mobility optimization, and digital twin systems.

Beyond geospatial data classification, our approach has broader practical implications for real-world smart city applications. Accurate geo-entity classification can improve urban mobility analysis, environmental monitoring, and disaster response planning. For instance, better classification of transportation infrastructure can inform traffic optimization models, while detecting green spaces and water bodies can aid in climate adaptation strategies. Additionally, our method can support city governance systems by enabling smarter urban zoning decisions, infrastructure maintenance, and land use forecasting. By integrating our embeddings into real-time smart city platforms, urban planners and policymakers can make data-driven decisions that promote sustainability and livability.

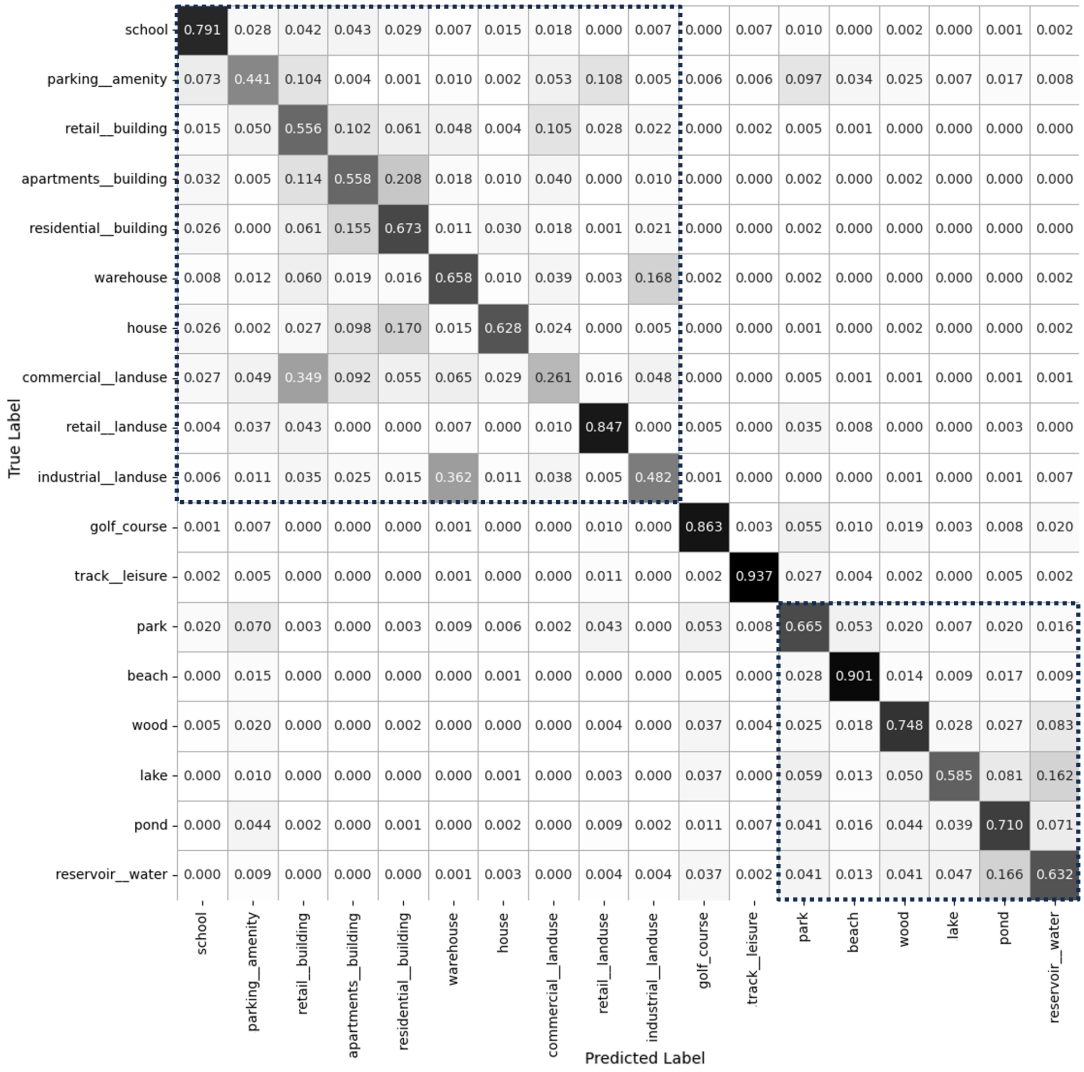


Figure 6: Confusion matrix illustrating classification results of geo-entities to *OpenStreetMap* types using the OSM-16k dataset, employing the model derived from the Setting 4. The matrix aggregates results across all mutually exclusive subsets of tests.

4 Related Work

The semantics of geospatial information is a rich domain that demands special attention within the web. Although GIS interoperability research has addressed fundamental issues regarding the geometry of geospatial features, recent surveys indicate that current approaches do not effectively address the utilization of specific semantics by users for performing tasks that leverage geospatial data [15–17]. Despite these challenges, research on geospatial semantics has seen significant growth in recent years.

Web-based geospatial data presents significant challenges in interoperability, standardization, and data reliability. While frameworks like Linked Open Data (LoD), and OGC GeoSPARQL [4] have improved cross-platform data sharing, challenges remain in integrating heterogeneous data sources. Our work contributes to this field by bridging structured and unstructured geospatial knowledge, providing semantic representations that can be integrated into web-based urban intelligence systems.

The use of machine learning for geospatial data classification has gained significant attention, with convolutional neural networks (CNNs) being a popular approach. Castelluccio et al. [6] proposed a CNN-based approach for land use classification using remote sensing images, and Li et al. [27] developed a CNN-based framework for automatic recognition of building footprints. Dsouza et al. [10] proposed a neural architecture that capitalizes upon a shared latent space for tag-to-class alignment for OSM entities. Klemmer et al. [22] developed a GNN-based approach for context-aware vector encoding of geographic coordinates, Kaczmarek et al. [20] proposed a GNN-based method for spatial object classification using topology, and Xu et al. [44] used a GCN-based approach that incorporates spatial context and aggregates information of adjacent nodes within

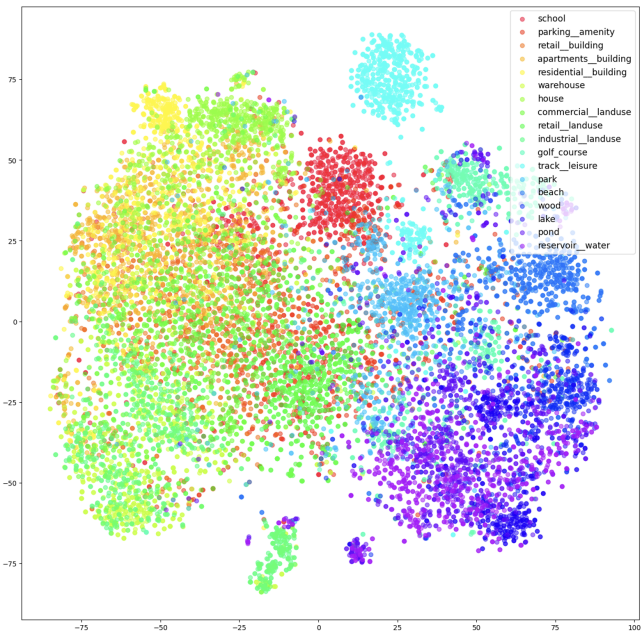


Figure 7: t-SNE visualization of embeddings derived from a 10k sample of OSM geo-entities from the California snapshot, generated using our model, and labeled according to the most fine-grained OSM tag. The colors signify the ground-truth labels attributed to each instance.

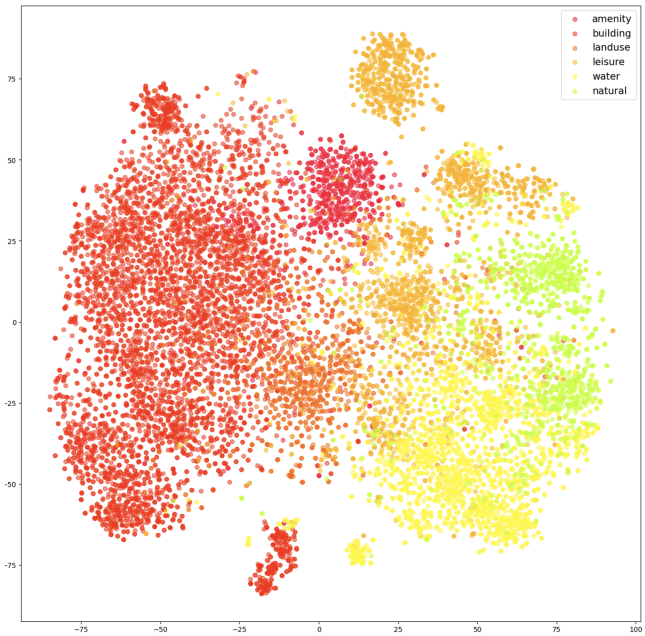


Figure 8: t-SNE visualization utilizing the same data in Figure 7, showcasing 10k OSM samples. Here, entities are labeled according to the highest-level OSM tags. Different colors distinctly categorize the respective high-level ground-truth labels assigned to each instance.

the graph for urban land-use classification. Yan et. al. [45] developed an approach that combines multiple features extracted from the boundary of a geospatial object to obtain a cognitively compliant shape encoding. Our work is concerned with a learning task that incorporates multiple sources of information for use in NNs, specifically open data, such as OSM, to improve geospatial data representation and classification.

Geospatial embedding techniques have been explored for geospatial data analysis. Tempelmeier et al. [37] published GeoVectors, offering a pre-trained OSM embedding corpus we referenced earlier. Additionally, Jenkins et al. [19] proposed a method for unsupervised representation learning of spatial data via multimodal embedding. Another example is SpaBERT [26], a spatial language model that provides a general-purpose representation of geo-entities based on named neighboring entities in geospatial data, which can be helpful for geo-entity typing. Moreover, Qiu et al. [32] introduced a method that employs geospatial distance to optimize knowledge embedding for a Geographic Knowledge Graph (GeoKG) to help refine latent representations of geo-entities and geo-relations. In contrast to the approaches mentioned above, our work leverages geometric properties of geospatial features, including their shape, as part of the input signals and other information, to optimize the embedding process.

Incorporating open data, such as *OpenStreetMap*, for geo-entity representation has received limited attention. Woźniak and Szymański [43] proposed a method to embed OSM regions. This method is not directly comparable since it does not embed arbitrary OSM region entities; instead, it decomposes space and embeds each grid cell using the tags contained in it to learn vector representations.

5 Conclusion and Future Work

As digitized geospatial data becomes increasingly available, developing techniques for responsible, AI-driven urban intelligence is crucial. This work introduces a novel approach for self-supervised geo-entity embedding, leveraging geometric, spatial, and semantic neighborhood contexts to generate robust representations for geospatial applications such as smart cities. Our method enables seamless geo-entity typing and classification by using open web data, particularly *OpenStreetMap* (OSM), bridging the gap between urban geospatial data integration and AI-driven decision-making. Additionally, we implemented a taxonomy-aware contrastive learning framework, integrating hierarchical semantic relationships into the loss function to enhance geo-entity classification.

Future work could integrate explainability techniques to improve interpretability and user trust and investigate bias detection mechanisms to ensure responsible and equitable AI deployment in smart city environments. Incorporating pre-trained word embeddings and attention mechanisms could further refine representations, extending the model's understanding of urban knowledge beyond local spatial contexts. Additionally, leveraging textual knowledge from open knowledge bases, such as Yago2Geo [21], could enrich geo-entity representations, supporting tasks such as geo-entity linking, spatial decision-making, and digital twin applications.

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. GPT-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[2] Leonidas Anthopoulos and Panos Fitsilis. 2010. From Digital to Ubiquitous Cities: Defining a Common Architecture for Urban Development. In *2010 Sixth International Conference on Intelligent Environments*. 301–306. doi:10.1109/IE.2010.61

[3] Leonidas G Anthopoulos, Marijn Janssen, and Vishanth Weerakkody. 2015. Comparing Smart Cities with different modeling approaches. In *Proceedings of the 24th International Conference on World Wide Web*. 525–528.

[4] Robert Battle and Dave Kolas. 2011. Geosparql: enabling a geospatial semantic web. *Semantic Web Journal* 3, 4 (2011), 355–370.

[5] Christopher Bone, Alan Ager, Ken Bunzel, and Lauren Tierney. 2016. A geospatial search engine for discovering multi-format geospatial data across the web. *International Journal of Digital Earth* 9, 1 (2016), 47–62. doi:10.1080/17538947.2014.966164

[6] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. 2015. Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092* (2015).

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*. PMLR, 1597–1607.

[8] Yao-Yi Chiang, Muhao Chen, Weiwei Duan, Jina Kim, Craig A Knoblock, Stefan Leyk, Zekun Li, Yijun Lin, Min Namgung, Basel Shbita, et al. 2023. GeoAI for the Digitization of Historical Maps. In *Handbook of Geospatial Artificial Intelligence*. CRC Press, 217–247.

[9] Yao-Yi Chiang, Stefan Leyk, and Craig A. Knoblock. 2014. A survey of digital map processing techniques. *ACM Computing Surveys (CSUR)* 47, 1 (2014), 1–44. doi:10.1145/2557423

[10] Alishiba Dsouza, Nicolas Tempelmeier, and Elena Demidova. 2021. Towards Neural Schema Alignment for OpenStreetMap and Knowledge Graphs. In *International Semantic Web Conference*. Springer, 56–73.

[11] Weiwei Duan, Y Chiang, Crang A Knoblock, Stefan Leyk, and J Uhl. 2018. Automatic generation of precisely delineated geographic features from georeferenced historical maps using deep learning. In *Proceedings of the 22nd International Research Symposium on Computer-based Cartography and GIScience (Autocarto/UCGIS)*, Scott Freundschuh and Diana Sinton (Eds.). UCGIS.org, 59–63. https://www.ucgis.org/assets/docs/AutoCarto-UCGIS%202018%20Proceedings.pdf

[12] Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821* (2021).

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[14] John Herring et al. 2011. Opengis® implementation standard for geographic information-simple feature access-part 1: Common architecture [corrigendum]. (2011).

[15] Yingjie Hu. 2017. Geospatial semantics. *arXiv preprint arXiv:1707.03550* (2017).

[16] Krzysztof Janowicz, Song Gao, Grant McKenzie, Yingjie Hu, and Budhendra Bhaduri. 2020. GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science* 34, 4 (2020), 625–636. doi:10.1080/13658816.2019.1684500

[17] Krzysztof Janowicz, Simon Scheider, Todd Pehle, and Glen Hart. 2012. Geospatial semantics and linked spatiotemporal data–Past, present, and future. *Semantic Web* 3, 4 (2012), 321–332.

[18] Marijn Janssen, Paul Brous, Elsa Estevez, Luis S Barbosa, and Tomasz Janowski. 2020. Data governance: Organizing data for trustworthy Artificial Intelligence. *Government information quarterly* 37, 3 (2020), 101493.

[19] Porter Jenkins, Ahmad Farag, Suhang Wang, and Zhenhui Li. 2019. Unsupervised representation learning of spatial data via multimodal embedding. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 1993–2002.

[20] Iwona Kaczmarek, Adam Iwaniak, and Aleksandra Świetlicka. 2023. Classification of Spatial Objects with the Use of Graph Neural Networks. *ISPRS International Journal of Geo-Information* 12, 3 (2023), 83.

[21] Nikolaos Karalis, Georgios Mandilaras, and Manolis Koubarakis. 2019. Extending the YAGO2 knowledge graph with precise geospatial knowledge. In *The Semantic Web – ISWC 2019 (Lecture Notes in Computer Science)*, Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtěch Svátek, Isabel Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon (Eds.). Springer, Cham, 181–197. doi:10.1007/978-3-030-30796-7_12

[22] Konstantin Klemmer, Nathan S Safir, and Daniel B Neill. 2023. Positional encoder graph neural networks for geographic data. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 1379–1389.

[23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25 (2012).

[24] Phuc H Le-Khac, Graham Healy, and Alan F Smeaton. 2020. Contrastive representation learning: A framework and review. *Ieee Access* 8 (2020), 193907–193934.

[25] Zekun Li, Yao-Yi Chiang, Sasan Tavakkol, Basel Shbita, Johannes H Uhl, Stefan Leyk, and Craig A Knoblock. 2020. *An automatic approach for generating rich, linked geo-metadata from historical map images*. Association for Computing Machinery, New York, NY, USA, 3290–3298. doi:10.1145/3394486.3403381

[26] Zekun Li, Jina Kim, Yao-Yi Chiang, and Muhao Chen. 2022. SpaBERT: A Pretrained Language Model from Geographic Data for Geo-Entity Representation. *arXiv preprint arXiv:2210.12213* (2022).

[27] Zhichao Li, Shuai Zhang, and Jinwei Dong. 2022. Suggestive Data Annotation for CNN-Based Building Footprint Mapping Based on Deep Active Learning and Landscape Metrics. *Remote Sensing* 14, 13 (2022), 3147.

[28] Fandel Lin, Craig A Knoblock, Basel Shbita, Binh Vu, Zekun Li, and Yao-Yi Chiang. 2023. Exploiting Polygon Metadata to Understand Raster Maps-Accurate Polygonal Feature Extraction. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*. 1–12.

[29] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[30] Georgios Mylonas, Athanasios Kalogeras, Georgios Kalogeras, Christos Anagnostopoulos, Christos Alexakos, and Luis Muñoz. 2021. Digital twins from smart manufacturing to smart cities: A survey. *Ieee Access* 9 (2021), 143222–143249.

[31] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[32] Peiyuan Qiu, Jialiang Gao, Li Yu, and Feng Lu. 2019. Knowledge embedding with geospatial distance restriction for geographic knowledge graph completion. *ISPRS International Journal of Geo-Information* 8, 6 (2019), 254.

[33] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).

[34] Basel Shbita and Craig A Knoblock. 2024. Automatically Constructing Geospatial Feature Taxonomies from OpenStreetMap Data. In *2024 IEEE 18th International Conference on Semantic Computing (ICSC)*. IEEE, 208–211.

[35] Basel Shbita, Craig A Knoblock, Weiwei Duan, Yao-Yi Chiang, Johannes H Uhl, and Stefan Leyk. 2023. Building Spatio-Temporal Knowledge Graphs from Vectorized Topographic Historical Maps. *Semantic Web* 14, 3 (2023), 527–549. doi:10.3233/SW-222918

[36] Kihyuk Sohn. 2016. Improved deep metric learning with multi-class n-pair loss objective. *Advances in neural information processing systems* 29 (2016).

[37] Nicolas Tempelmeier, Simon Gottschalk, and Elena Demidova. 2021. GeoVectors: A Linked Open Corpus of OpenStreetMap Embeddings on World Scale. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 4604–4612.

[38] Johannes H. Uhl and Weiwei Duan. 2021. Automating information extraction from large historical topographic map archives: new opportunities and challenges. In *Handbook of Big Geospatial Data*, Martin Werner and Yao-Yi Chiang (Eds.). Springer, Cham, 509–522. doi:10.1007/978-3-030-55462-0_20

[39] Johannes H. Uhl, Stefan Leyk, Yao-Yi Chiang, Weiwei Duan, and Craig A. Knoblock. 2019. Automated extraction of human settlement patterns from historical topographic map series using weakly supervised convolutional neural networks. *IEEE Access* 8 (2019), 6978–6996. doi:10.1109/ACCESS.2019.2963213

[40] Johannes H Uhl, Stefan Leyk, Weiwei Duan, Zekun Li, Basel Shbita, Yao-Yi Chiang, and Craig A Knoblock. 2021. Towards the large-scale extraction of historical land cover information from historical maps. *Abstracts of the ICA* 3 (2021), 1–2.

[41] Johannes H Uhl, Stefan Leyk, Zekun Li, Weiwei Duan, Basel Shbita, Yao-Yi Chiang, and Craig A Knoblock. 2021. Combining remote-sensing-derived data and historical maps for long-term back-casting of urban extents. *Remote Sensing* 13, 18 (2021), 3672. doi:10.3390/rs13183672

[42] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.

[43] Szymon Woźniak and Piotr Szymański. 2021. Hex2vec: Context-Aware Embedding H3 Hexagons with OpenStreetMap Tags. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. 61–71.

[44] Yongyang Xu, Bo Zhou, Shuai Jin, Xuejing Xie, Zhanlong Chen, Sheng Hu, and Nan He. 2022. A framework for urban land use classification by integrating the spatial context of points of interest and graph convolutional neural network method. *Computers, Environment and Urban Systems* 95 (2022), 101807.

[45] Xiongfeng Yan, Tinghua Ai, Min Yang, and Xiaohua Tong. 2021. Graph convolutional autoencoder model for the shape coding and cognition of buildings in maps. *International Journal of Geographical Information Science* 35, 3 (2021), 490–512.