Delft University of Technology

An image embedding-based approach for classifying street networks

Garrido-Valenzuela, Francisco; Lange, Max; Herrera, Juan C.; Van Cranenburgh, Sander; Cats, Oded

DOI
10.1145/3748636.3762715

Licence
CC BY

Publication date
2025

Document Version
Final published version

Published in
33rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM SIGSPATIAL 2025

Citation (APA)
Garrido-Valenzuela, F., Lange, M., Herrera, J. C., Van Cranenburgh, S., & Cats, O. (2025). An image embedding-based approach for classifying street networks. In M. Mokbel, S. Shekar, A. Zufle, Y.-Y. Chiang, M. L. Damiani, & M. Youssef (Eds.), *33rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM SIGSPATIAL 2025* (pp. 120-123). (33rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM SIGSPATIAL 2025). ACM. https://doi.org/10.1145/3748636.3762715

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Check for updates

Latest updates: https://dl.acm.org/doi/10.1145/3748636.3762715

SHORT-PAPER

An image embedding-based approach for classifying street networks

FRANCISCO GARRIDO-VALENZUELA, Delft University of Technology, Delft, Zuid-Holland, Netherlands

MAX LANGE, Department of Transport & Planning, TU Delft, Delft, Zuid-Holland, Netherlands

JUAN C HERRERA, Pontifical Catholic University of Chile, Santiago, RM, Chile

SANDER VAN CRANENBURGH, Delft University of Technology, Delft, Zuid-Holland, Netherlands

ODED CATS, Department of Transport & Planning, TU Delft, Delft, Zuid-Holland, Netherlands

Open Access Support provided by:

Delft University of Technology

Department of Transport & Planning, TU Delft

Pontifical Catholic University of Chile

PDF Download
3748636.3762715.pdf
05 January 2026
Total Citations: 0
Total Downloads: 134

Published: 03 November 2025

Citation in BibTeX format

SIGSPATIAL '25: 33rd ACM International Conference on Advances in Geographic Information Systems
November 3 - 6, 2025
MN, Minneapolis, USA

Conference Sponsors:
SIGSPATIAL

SIGSPATIAL '25: Proceedings of the 33rd ACM International Conference on Advances in Geographic Information Systems (November 2025)
https://doi.org/10.1145/3748636.3762715
ISBN: 9798400720864

An image embedding-based approach for classifying street networks

Francisco Garrido-Valenzuela
f.garridov@uc.cl
CityAI Lab, Transport and Logistics Group, Delft University of Technology
Delft, The Netherlands

Max Lange
M.T.S.Lange@student.tudelft.nl
Department of Transport & Planning, Delft University of Technology
Delft, The Netherlands

Juan C. Herrera
jch@uc.cl
Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile
Santiago, Chile

Sander van Cranenburgh
S.vanCranenburgh@tudelft.nl
CityAI Lab, Transport and Logistics Group, Delft University of Technology
Delft, The Netherlands

Oded Cats
O.Cats@tudelft.nl
Department of Transport & Planning, Delft University of Technology
Delft, The Netherlands

Abstract

We present a method to classify street networks using only geotagged street-level imagery. By combining pre-trained image embeddings with unsupervised clustering, it produces visually coherent street typologies without supervised training or labeled data and requires only minimal data curation. The approach is lightweight, scalable, and, in principle, transferable across urban contexts. In a Delft (Netherlands) case study, we classify approximately 2,000 road sections using over 70,000 images. Our method recovers distinct street types such as residential, arterial, and historic ones. These results show that pre-trained visual embeddings alone can support effective street classification from visual inputs, offering a practical tool for urban planning, transport analysis, and mobility research.

CCS Concepts

• **Computing methodologies** → **Image representations**; **Cluster analysis**; *Dimensionality reduction and manifold learning*; Learning latent representations; *Spatial and physical reasoning*; • **Information systems** → *Geographic information systems.*

Keywords

Street classification, Embedding representation, Computer vision, Image processing, Road transportation

ACM Reference Format:
Francisco Garrido-Valenzuela, Max Lange, Juan C. Herrera, Sander van Cranenburgh, and Oded Cats. 2025. An image embedding-based approach for classifying street networks. In *The 33rd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '25), November 3–6, 2025, Minneapolis, MN, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3748636.3762715

This work is licensed under a Creative Commons Attribution 4.0 International License.
SIGSPATIAL '25, Minneapolis, MN, USA
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2086-4/2025/11
https://doi.org/10.1145/3748636.3762715

1 Introduction

Classifying streets is a core component of transport network design and urban planning. It supports applications ranging from defining operational guidance for automated vehicles and addressing safety concerns [2] to prioritizing infrastructure investment [5]. Beyond immediate transport uses, street hierarchies shape urban expansion, spatial development, and land-use patterns [7, 13]. These influence accessibility, business location, and housing demand. Overall, these roles underscore the need for meaningful and adaptable street classification systems.

Existing approaches include manual rule-based schemes and data-driven techniques. Traditional systems rely on predefined categories linked to road function, traffic intensity, or access levels [14]. These systems often requiring on-site surveys and focusing on a limited set of transport attributes [13]. Among other approaches, computational and data-driven methods can extract multiple layers of information such as surface type via image processing [11] and traffic dynamics or accidents records via deep learning [12, 16]. More recently, representation learning on street network data has been applied—for example, to categorize street zones using OpenStreetMap infrastructure and points of interest [8]. More broadly, urban representation learning frameworks (e.g., *urban2vec* [15]) learn embeddings to characterize urban contexts which also support street classification, but they typically require training on large datasets to yield city- or domain-specific representations.

Despite these advances, widely used street classification methods either depend on predefined categories or require supervised training with large, well-distributed labeled datasets. This limits their adaptability across diverse urban contexts and their transferability, especially in data-scarce regions. On the one hand, focusing on a defined set of attributes can make schemes less adaptable to urban change, while reliance on field surveys and expert input increases costs. On the other hand, more flexible machine-learning approaches mitigate some issues yet often require training and computational demands. This motivates a scalable, adaptable, and data-efficient alternative that captures street complexity using widely available visual data.

We address this gap with a data-driven approach that combines pre-trained image embeddings and unsupervised clustering to classify streets using only street-level images (SLI). Rather than training new models or defining manual rules, we leverage generic image embeddings to encode visual appearance (textures, shapes, spatial layout) and group street segments by similarity. Because the method reuses pre-trained computer vision backbones, it is lightweight and transferable across cities. Additionally, the growing coverage of SLI platforms (e.g., Google Street View, Mapillary, Apple Look Around) further supports applicability where structured road datasets are scarce. We implement a four-step workflow: image-to-road matching, embedding extraction, street section aggregation, and clustering. We demonstrate it in Delft, the Netherlands, recovering coherent street typologies from visual patterns alone. Together, these elements provide a simple, training-free path to scalable street classification and a practical basis for street networks analysis.

2 Background

Image embeddings map images to fixed-length vectors so that visually similar images are nearby in feature space. The concept generalizes word embeddings from NLP [10] to vision by using pre-trained backbones (e.g., ResNet [6] or ViT) without the final classification/task-specific layer. Without this layer, models can produce embeddings suitable for clustering and retrieval. We leverage such generic embeddings to describe street-level imagery without additional training, enabling unsupervised grouping of street segments by visual appearance. Figure 1 illustrates this principle.

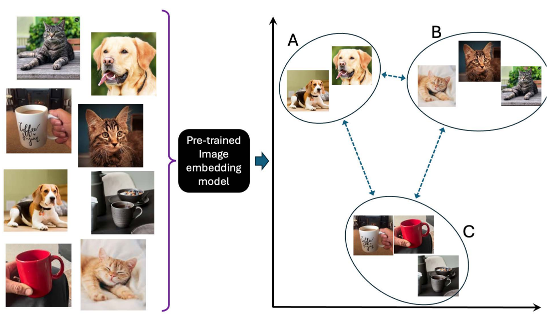

Figure 1: Illustrative mapping of images onto a 2D space using an image embedding model. Similar images appear closer to each other in the projected 2D space.

Figure 1 presents a diverse set of images, including dogs, cats, and cups processed with a pre-trained image embedding model. This model maps them into a high-dimensional space. The result is a structured representation where visually and semantically similar images are positioned closer together. This is the principle we use to classify streets.

3 Method

This section describes the input data and outlines our method for classifying the street network based on street-level imagery. To illustrate the method, we include examples from its implementation in Delft, the Netherlands. Delft is a historical city in the western part of the Netherlands, covering approximately $24km^2$. Figure 2 provides an overview of the pipeline employed in the classification process.

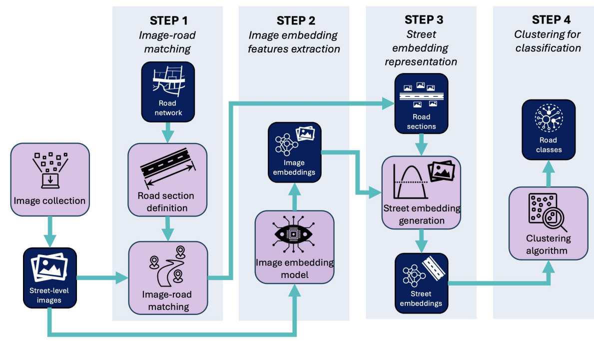

Figure 2: Pipeline of the four-step method.

Data. Our classification method utilizes street-level images (SLI) to extract visual street information and a Geographic Information System (GIS) file to represent the road network. SLIs are panoramic photographs captured at ground level, providing detailed visual and structural information about street surroundings. Several providers, including Google Street View (GSV), Mapillary [9], and Apple Look Around [1], offer these images worldwide. For this study, we use imagery from GSV [4], employing the image ID collection method described by [3]. This approach systematically gathers geo-tagged image IDs across a study area, with each ID corresponding to a specific image and its geographical coordinates. We then filter the images to retain only those facing forward and backward along the streets, ensuring that the road and its surroundings are visually captured. In addition to the SLIs, the GIS road network file is available through OpenStreetMap and it provides the coordinates needed to map and identify each street in the study area. For the city of Delft, around 70 thousands SLIs were collected covering approximately 300 km of road network.

3.1 Step 1: Image-road matching

The first step involves defining and associating the road section units with the SLIs. A road section unit is defined as a spatial line segment that represents either an entire street or a portion of it. These road sections serve as the minimal spatial units for street classification in our method. The definition of a section can vary depending on the application, such as the segment between two intersections, sections based on street names, or fixed-length segments (e.g., 100 meters). Then, the SLIs are spatially associated with the road sections.

In our application for Delft, we define road sections using intersection nodes from the GIS network file. These nodes represent points where streets intersect or experience strong deviations serving as natural boundaries for segmenting the road network. This process yields a total of 3,429 road sections averaging around 130 meters each. Once the road sections are defined, we create 20-meter buffers around them to match SLIs from their surroundings. If an image corresponds to multiple sections, it is assigned to the closest one. As a result, 1,914 road sections are matched with at least one image, and 68,178 SLIs are used in total.

3.2 Step 2: Image embedding features extraction

The second step is designed to extract features from the road sections' SLIs by using a pre-trained image embedding model. This model transforms street-level images into vectors (aka embeddings). These embeddings capture essential visual characteristics of the images, such as textures, colors, structural patterns, and composition, while reducing the dimensionality of the data. Several pre-trained models can be used for this purpose, including convolutional neural networks (CNNs) and Vision Transformers (ViTs). Once images are represented as vectors, visually similar images will be positioned closer together in the multidimensional space, as illustrated in Figure 1. The distances among images in the multidimensional space facilitate the street clustering based on shared visual characteristics.

We employ ResNet152, a CNN-based model pre-trained on ImageNet [6] with *IMAGENET1K_V2* PyTorch weights for processing Delft's SLIs. We choose this model because of its ability to extract high-quality features from images across diverse contexts, its ease of implementation, and its manageable computational requirements. For each road section, we process the associated images through this model to generate the embeddings (i.e., one vector per image). ResNet152 generates a $2,048$-dimensional feature vector for each image, preserving various visual attributes relevant for classification.

3.3 Step 3: Street embedding representation

The third step involves aggregating the image embeddings associated with each road section to create a unified street vector representation. To achieve this, we compute the mean of the embedding dimensions for all images linked to a given road section. By averaging, we derive a comprehensive representation that encapsulates the overall visual content of the section, balancing the contribution of each image while reducing noise from individual ones. This approach ensures that the resulting vector reflects the collective visual characteristics of the road section rather than relying on any single image.

Once the street embeddings are generated, we address the curse of dimensionality, which can negatively affect clustering performance (in step 4). High-dimensional embeddings can result in sparse data distributions, making difficult the identification of meaningful groups. To mitigate this, we apply Principal Component Analysis (PCA) for reducing dimensionality while preserving at least 80% of the variance. This reduction enhances clustering stability and ensures classification results remain interpretable. After PCA, the embeddings are reduced from $2,048$ to 87 dimensions, which balances computational efficiency with information retention.

3.4 Step 4: Clustering for classification

The final step involves classifying the streets by applying clustering techniques to the aggregated street embedding vectors. The goal is to group road sections based on shared visual patterns captured from street-level imagery. We use the elbow method to determine the appropriate number of clusters, which evaluates the variance explained as a function of the number of clusters and identifies the point of diminishing returns. This approach allows us to balance classification detail with interpretability.

For the clustering implementation, we apply hierarchical agglomerative clustering, which is well-suited to the nested structure commonly found in urban road networks. This technique constructs a tree-like hierarchy of clusters, aligning with how streets are often organized, from major roads to local residential streets. We also test alternative clustering methods to evaluate the robustness of our results, including K-means and Gaussian Mixture Models (GMM). These methods are implemented using the same embedding representations. All three approaches yield consistent classifications, with the core urban typologies remaining stable across methods. Based on this consistency and the interpretability of hierarchical outputs, we select agglomerative clustering for our final analysis.

4 Results

This section presents the results of applying our embedding-based classification method to the street network of Delft.

4.1 Clustering results

We apply hierarchical clustering over the embeddings to identify distinct street types in Delft based on their visual appearance. The elbow method determines that six clusters provide the optimal balance between compactness and interpretability. Figure 3 shows the resulting classification, with Delft's roads color-coded by cluster.

Figure 3: Delft's streets colored by clusters. Cluster names are assigned based on the visual characteristics of the images.

The model successfully differentiates a variety of urban typologies. The city center, marked in light blue (cluster 2), is characterized by narrow streets, canals, and historic, closely built houses with minimal vehicle presence. Our model accurately identifies the natural boundaries of this area. Notably, the light blue (cluster 2) has two sub-regions separated by the arterial blue road. The smaller one is composed of some streets located west of the train tracks—outside the traditional boundaries of the old town— that are also grouped in this cluster, sharing similar visual features such as street width, facade style, and limited vehicular traffic. On the other hand, a few streets located within the area of the old town (right side of the major light blue region) are assigned to different clusters, likely due to distinctive characteristics such as wider streets or modern architecture. Additionally, the two major motorways passing through Delft are clearly distinguished in yellow (cluster 3). While the clustering results effectively segment different street types, qualitative validation is necessary to ensure these clusters align with meaningful visual distinctions. To support this, we conduct a qualitative review of randomly sampled images from each cluster.

4.2 Visual interpretation of clustered streets

We conduct a qualitative analysis to better understand the visual differences between clusters and validate their coherence. To do so, we sample random images from each group. This exploratory process complements the cluster definitions by analyzing visual features in the images. Figure 4 presents a set of random images from the six clusters.

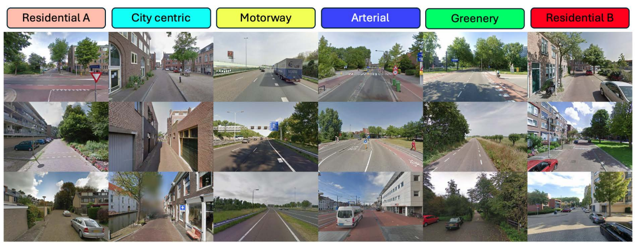

Figure 4: Randomly sampled images from each cluster.

We labeled the clusters based on characteristics observed in the sampled images and their spatial distribution of the streets in Figure 3. Cluster 2 (light blue) is labeled *City Centric* due to its narrow streets and historic buildings. This area closely aligns with the well-known historic center of Delft. Clusters 1 and 6 (pink and red) are labeled *Residential A* and *Residential B*, respectively, characterized by residential houses, open areas, and the presence of trees on the streets. Residential B encircles the city center, maintaining a built environment similar to the historic core. In contrast, Residential A, located in the southern outskirts, exhibits wider streets, newer buildings, and more green spaces, reflecting a more modern residential layout. Cluster 3 (yellow) is identified as *Motorway*, encompassing the main motorways in Delft, and visible from the images showing wide, higher-speed roads. Cluster 4 (blue), labeled *Arterial*, features wider streets with multiple lanes, reflecting their role as major traffic arteries connecting different city areas. Lastly, Cluster 5 (green), labeled *Greenery*, includes roads surrounded by more rural landscapes and vegetation. While similar to Arterial roads in their connectivity function, the visual cues of abundant greenery distinguish this cluster. These labels serve as a preliminary interpretation of the clusters, which can be further refined by incorporating additional data and expert knowledge.

5 Conclusions

We propose a lightweight, city-agnostic method for street classification that combines pre-trained image embeddings with unsupervised clustering over street-level imagery, requiring no supervised training and minimal data curation. In a Delft case study, the approach produced recognizable street typologies from visual patterns alone; qualitative inspection and consistency across clustering algorithms support the coherence of the groupings. Computationally, cost is dominated by single-pass embedding extraction and clustering (no training loop), which supports practical scalability on commodity hardware; scalability is governed mainly by SLI availability and the number of images/segments.

Limitations include dependence on SLI coverage and quality, lack of explicit handling of temporal/seasonal change, the inherent opacity of embeddings, and qualitative label assignment. While the design is transferable in principle, multi-city evaluations are needed to establish external validity and operational cost profiles. Future work should test the method across diverse urban contexts, develop practitioner-in-the-loop workflows for systematic labeling, compare backbone encoders (e.g., ResNet vs. ViT/CLIP), and integrate complementary modalities (e.g., satellite or traffic data).

Acknowledgments

This work was supported by the TU Delft AI Labs programme.

References

[1] Apple Inc. 2023. Apple Maps Look Around. https://www.apple.com/maps/ Accessed: 2024-04-12.

[2] Atze Dijkstra. 2011. En route to safer roads. How road structure and road classification can affect road safety. (2011).

[3] Francisco Garrido-Valenzuela, Oded Cats, and Sander van Cranenburgh. 2023. Where are the people? Counting people in millions of street-level images to explore associations between people's urban density and urban characteristics. *Computers, Environment and Urban Systems* 102 (2023), 101971.

[4] Google. 2023. Google Street View. https://www.google.com/maps/streetview/ Accessed: 2024-11-03.

[5] Christian Gorges, Kemal Öztürk, and Robert Liebich. 2018. Road classification for two-wheeled vehicles. *Vehicle system dynamics* 56, 8 (2018), 1289–1314.

[6] 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.

[7] Yiwen Hu and Chao Liang. 2024. Study on the spatial relationship between road network and the diversity of urban public facilities: the case of the central area of Changsha City. *Journal of Engineering and Applied Science* 71, 1 (2024), 156.

[8] Kacper Leśniara and Piotr Szymański. 2022. Highway2vec: Representing OpenStreetMap microregions with respect to their road network characteristics. In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. 18–29.

[9] Mapillary. 2023. Mapillary. https://www.mapillary.com/ Accessed: 2024-05-12.

[10] T Mikolov. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[11] M Mohammadi. 2012. Road classification and condition determination using hyperspectral imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39 (2012), 141–146.

[12] Madhar Taamneh, Salah Taamneh, and Sharaf Alkheder. 2017. Clustering-based classification of road traffic accidents using hierarchical clustering and artificial neural networks. *International journal of injury control and safety promotion* 24, 3 (2017), 388–395.

[13] Stefanos Tsigdinos, Georgios Salamouras, Ioannis Chatziioannou, Efthimios Bakogiannis, and Alexandros Nikitas. 2024a. A worldwide review of formal national street classification plans enhanced via an analytical hierarchy process: Street classification as a tool for more sustainable cities. *Cities* 154 (2024a), 105371.

[14] Jeffrey Tumlin. 2011. *Sustainable transportation planning: tools for creating vibrant, healthy, and resilient communities*. John Wiley & Sons.

[15] Zhecheng Wang, Haoyuan Li, and Ram Rajagopal. 2020. Urban2vec: Incorporating street view imagery and pois for multi-modal urban neighborhood embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 1013–1020.

[16] Jun Zhang, Xiao Chen, Yang Xiang, Wanlei Zhou, and Jie Wu. 2014. Robust network traffic classification. *IEEE/ACM transactions on networking* 23, 4 (2014), 1257–1270.