A Scalable Methodology for Identifying and Mapping Missing Cycling Infrastructure in OpenStreetMap through Street-Level Imagery and Community Validation

Simon Metzler*[1] and Dominik Ziemke[1]

[1]Institute of Transport Research, German Aerospace Center (DLR), Berlin, Germany

Short summary

High-quality and complete representations of cycling infrastructure are essential for transport research, planning, and publicly funded projects. In many countries, OpenStreetMap (OSM) serves as a primary open data source for modelling and decision support in this domain. At the same time, limitations of OSM data completeness – especially for cycling infrastructure – are frequently reported. Detecting infrastructure that is entirely missing from OSM remains particularly challenging, as such gaps cannot be identified through internal data consistency checks alone, while comprehensive authoritative reference datasets are often unavailable. This paper presents a methodology for systematically detecting such gaps by combining automated object detections from crowdsourced street-level imagery with spatial analysis of the existing OSM road network. Traffic signs and road markings serve as indicators of cycling infrastructure not yet represented in OSM. Candidate locations are filtered through spatial, temporal, and contextual criteria and translated into community validation tasks via MapRoulette. Applied across Germany, the approach generated over 1,200 validation tasks with false-positive rates below 20% among completed tasks. Through the resulting campaigns, more than 500 km of previously unmapped cycling infrastructure were added to OpenStreetMap.

Keywords: cycling infrastructure, data completeness, OpenStreetMap, street-level imagery, crowdsourcing, volunteered geographic information (VGI), community mapping

1 Motivation and Background

OpenStreetMap (OSM) (OSM contributors, 2026) has become a widely used spatial data foundation for transport research and applications (Lißner & Huber, 2021; Boeing, 2017; Ziemke et al., 2017). At the same time, limitations of OSM data completeness, especially for cycling infrastructure, are frequently reported (Vierø et al., 2023).

In recent years, significant progress has been made in analyzing and refining existing cycling infrastructure in OSM. In Germany, platforms such as *TILDA radinfra.de* (FixMyCity and OSM Verkehrswende Community, 2026) support this process by interpreting OSM tagging schemes and visualizing cycling infrastructure through rule-based representations of the data. Beyond visualization, the platform also provides interactive tools that support users in inspecting and improving OSM tagging, including direct links to relevant OSM wiki documentation and workflows for contributing edits. In addition, radinfra.de integrates MapRoulette challenges that help the community systematically refine and complete existing cycling infrastructure attributes in OSM.

However, these approaches primarily focus on improving the quality and completeness of infrastructure that is already represented in OSM. They cannot reveal cycling infrastructure that exists on the ground but is entirely absent from the database. A key limitation therefore remains: cycling infrastructure that exists in reality but is completely missing from OSM cannot be detected through internal data consistency checks alone. Until now, detecting such gaps has been requiring extensive local knowledge or manual field surveys, which do not scale to larger regions.

This paper proposes a methodology to systematically identify such missing infrastructure by combining street-level imagery, automated detection of cycling-related signs and markings, and community-based validation workflows within the OSM ecosystem.

Previous research has extensively examined the completeness and usability of OSM for representing cycling infrastructure. Several studies compare OSM with authoritative or municipal datasets in order to evaluate data quality. For example, Hochmair et al. (2014) assess the completeness

of bicycle trails and lanes in OSM by comparing them with official planning data, while Ferster et al. (2019) analyze OSM cycling infrastructure against municipal open data from several cities. More recently, the *BikeDNA* framework, introduced by Vierø et al. (2023), provides a systematic approach for assessing the quality of cycling infrastructure data in OSM. In their study, the authors apply this framework in a nationwide comparison between OSM and the Danish GeoDanmark dataset and show that neither dataset alone provides a complete representation of cycling infrastructure.

At the same time, recent work has highlighted the potential of crowdsourced street-level imagery for deriving infrastructure-related information from visual observations. Platforms such as *Mapillary*, a crowdsourced street-level imagery service, provide large-scale visual data that can be used for infrastructure analysis. For example, Kapp et al. (2025) use Mapillary (2026) images to analyze road surface type and quality and compare the resulting observations with OSM data.

In parallel, community-oriented platforms such as MapRoulette (2026) demonstrate how distributed microtasks can support targeted improvements of open geographic data. Majic et al. (2021) generate candidate errors using spatial reasoning to identify potentially inaccessible buildings in OSM and publish them as MapRoulette tasks for community review.

Our contribution connects these strands by using street-level image detections not only for analysis, but for the systematic identification and community-based validation of potentially missing cycling infrastructure in OSM. This is particularly relevant in countries like Germany, where comprehensive authoritative datasets for cycling infrastructure are not publicly available and comparison-based approaches therefore cannot be applied nationwide.

2 Observing Infrastructure through Street-Level Imagery

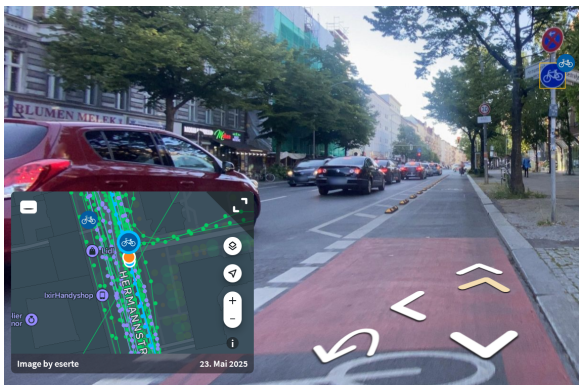
A central challenge in assessing the completeness of cycling infrastructure data is the lack of reliable ground truth. Verifying the presence of infrastructure through direct field surveys would require extensive local knowledge and large-scale observation efforts, which are difficult to achieve consistently across larger regions.

Street-level imagery provides an alternative form of observational evidence. While commercial platforms such as *Google Street View* or *Apple Look Around* offer extensive image coverage, their licensing terms do not permit the use of imagery for direct contributions to OSM OSM contributors (2026). In contrast, crowdsourced platforms such as Mapillary (2026) explicitly support OSM-based mapping workflows and provide open access to machine-learning-based object detections derived from the imagery.

Mapillary hosts large collections of georeferenced images documenting road environments at different locations and points in time. In addition to the imagery itself, Mapillary provides precomputed object detections generated by its machine-learning pipeline[1]. These detections identify objects such as traffic signs (Figure 1a) and road markings (Figure 1b) and can be accessed as georeferenced features.

In the context of cycling infrastructure, such detected objects can serve as indicators of the likely presence of cycling infrastructure in the surrounding road environment. Rather than relying solely on existing OSM data, these observations make it possible to infer candidate locations where cycling infrastructure may exist, but is not yet represented in the database. These observations form the conceptual basis for the methodology described in the following section.

[1] Cf. `https://www.mapillary.com/developer/api-documentation?#traffic-sign-tiles`, last accessed 26 March 2026.



(a) Traffic sign detection (traffic sign at the right edge, source: Mapillary).



(b) Bicycle marking detection (source: Mapillary).

Figure 1: Examples of cycling-related object detections in street-level imagery.

3 Methodology

The proposed workflow identifies candidate locations of potentially missing cycling infrastructure by combining automated detections from street-level imagery with the existing OSM road network.

1. Detection data preparation

   Automatically detected traffic signs and road markings related to cycling infrastructure are extracted from the Mapillary detection dataset. These detections provide spatial indicators of cycling-related infrastructure in the road environment.

2. Filtering of detections

   To reduce noise and false positives, detections are filtered based on the following criteria:

   - *Temporal recency:* Only detections with a timestamp newer than January 1, 2025 are considered.
   - *Multiple independent observations:* At least two observations of the same object are required. These observations must be separated by a minimum time interval of nine months to ensure that temporary objects (e.g., construction-related signs) are excluded.
   - *Contextual constraints:* In the context of German motorways, several blue-and-white signs visually resemble bicycle signage, which can lead to false positives. Therefore, all detections within a radius of 30 m from motorways are excluded.

3. OSM network preparation

   The current road network is extracted from OSM data via Geofabrik extracts (Geofabrik, 2026) using the Osmium toolkit (Topf, 2026) and processed to obtain cycling-relevant road segments. Existing cycling-related tags are identified in order to distinguish already mapped infrastructure from potential gaps.

4. Spatial association and gap inference

   Detected objects are spatially associated with nearby road segments in the OSM network. For each detection, all road segments within a predefined search radius (25–30 m, depending on the task configuration) are identified, accounting for positional uncertainties in both the detection data and the OSM geometries.

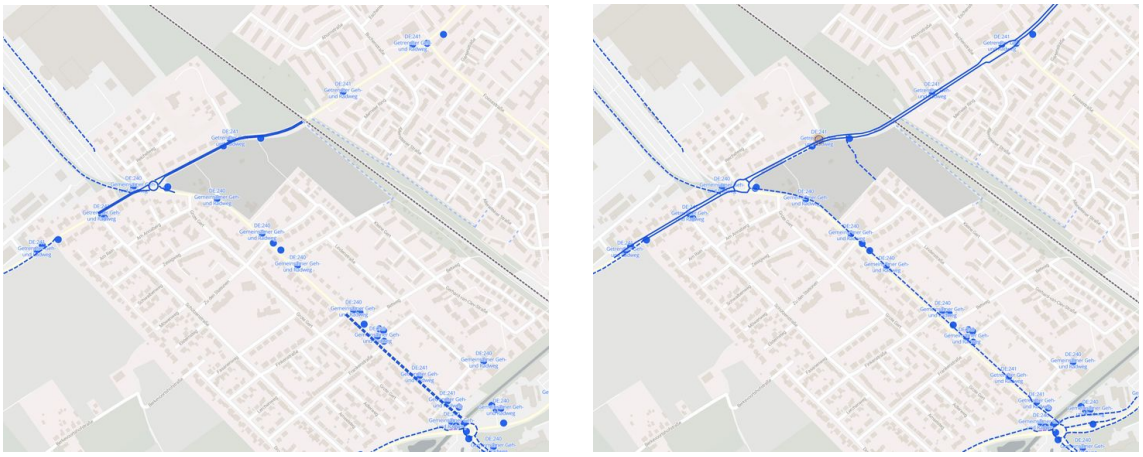
   The associated road segments are then examined for the presence of cycling-related tags (e.g. `highway=cycleway`, `cycleway=*`, or `bicycle=designated`).

   If at least one such segment exists, the detection is assumed to correspond to already mapped cycling infrastructure and is discarded. Otherwise, the detection is retained as a candidate mapping gap.

   This conservative heuristic prioritizes a low false-positive rate over exhaustive detection and provides a reliable basis for subsequent community validation. While some edge cases may therefore remain undetected, the resulting candidate locations provide a reliable basis for community validation tasks.

5. Operationalization via community tasking

   The resulting candidate locations are translated into microtasks and published as MapRoulette challenges[2]. Human mappers inspect the associated imagery and local context in order to verify the observation and update the OSM data if necessary.



(a) Situation before mapping.

(b) Cycling infrastructure added to OSM after validation.

Figure 2: Example of a candidate mapping gap identified by the workflow and the corresponding OSM update. (Source: TILDA radinfra.de)

As such, the workflow is semi-automated: detection, filtering, and task generation are automated, while final validation and mapping decisions remain with human contributors.

4 Implementation and Supporting Tools

The proposed workflow is implemented through a set of open-source tools that support the different stages of the pipeline, including the processing of street-level imagery detections and the generation of community validation tasks:

Processing of Mapillary detections.

Automated detections of traffic signs and road markings are retrieved and processed using dedicated scripts[3] developed for this project. The tool extracts cycling-related detections from the Mapillary dataset and applies the filtering procedures described in the methodology section (cf. sec. 3), including temporal filtering and aggregation of repeated observations. The processing pipeline is designed to run on a regular basis (e.g., weekly), ensuring that newly available detections from street-level imagery are incorporated into the workflow. The set of considered bicycle-related traffic signs and their corresponding Mapillary labels is summarized in Table 1.

[2] MapRoulette challenges: Traffic Signs, Road Markings, last accessed 30 March 2026.

[3] Cf. `https://github.com/vizsim/mapillary_trafficsigns`, last accessed 30 March 2026.

Table 1: Overview of bicycle-related traffic signs used in the analysis ('TS Code' is traffic sign code according to German traffic law).

TS Code	Description	Traffic Sign	Mapillary Label
DE:237	Bicycle path		`regulatory-bicycles-only-g1`
DE:240	Shared pedestrian and bicycle path		`regulatory-shared-path-pedestrians-and-bicycles-g1`
DE:241	Segregated pedestrian and bicycle path	 	`regulatory-dual-path-pedestrians-and-bicycles-g1` `regulatory-dual-path-bicycles-and-pedestrians-g1`
DE:244.2	End of bicycle road		`regulatory-end-of-bicycles-only-g2`
DE:1022-10	Bicycles permitted		`complementary-except-bicycles-g1`
DE:1000-33	Contraflow cycling permitted		`complementary-bike-route-g1`

Generation of MapRoulette challenges.

Candidate locations identified by the workflow are automatically converted into microtasks for the MapRoulette platform. Separate tools generate tasks for different types of infrastructure indicators:

- traffic sign detections[4]
- road marking detections[5]

Each task links the detected object to the corresponding street-level imagery and the nearby OSM road segment, allowing mappers to quickly inspect the visual evidence and update the map if necessary.
All tools used in this workflow are openly available and may be adapted for related infrastructure mapping tasks in other regions.

5 Results and Evaluation

The effectiveness of the proposed workflow is evaluated along two dimensions: first, the plausibility of the generated candidate tasks and the resulting validation outcomes, and second, the measurable additions of previously unmapped cycling infrastructure to OpenStreetMap.

Task validation through MapRoulette.

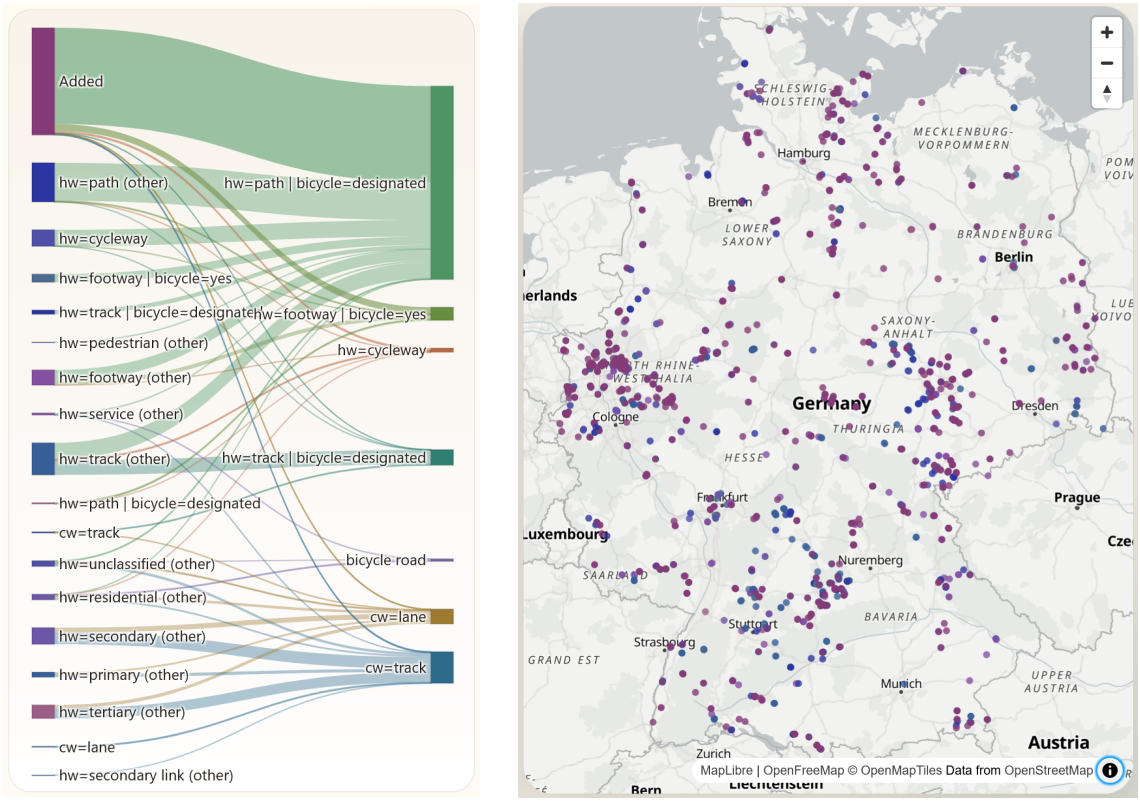
The first evaluation dimension concerns the plausibility of the generated candidate tasks. Task outcomes from the MapRoulette platform provide a direct indication of whether the automated detections correspond to real mapping gaps. Contributors can either confirm the task by updating OSM or mark it as "not an issue".
Two large-scale MapRoulette challenges (as described in sec. 4) were generated, focusing on traffic signs and road markings respectively. The proportion of confirmed fixes serves as an approximate indicator of the false-positive rate of the detection pipeline.

[4] Cf. Task generator for traffic signs on GitHub, last accessed 30 March 2026.
[5] Cf. Task generator for road markings on GitHub, last accessed 30 March 2026

For the traffic-sign based challenge (669 tasks), 478 tasks resulted in confirmed mapping fixes, while 103 tasks (17% of completed tasks) were marked as "not an issue". In 16 additional cases, the infrastructure had already been mapped before the task was completed. For the road-marking challenge (579 tasks), 210 tasks resulted in confirmed mapping fixes, while 36 tasks (14% of completed tasks) were classified as false positives. All reported statistics reflect the status of the challenges on 30 March 2026 and may change as additional tasks are completed.

Mapping impact on OSM.

The second evaluation dimension concerns the overall effect of the workflow on the OSM dataset. To assess this impact, edits triggered by the mapping campaigns are analyzed using dedicated changeset hashtags. Each MapRoulette challenge automatically pre-fills a specific hashtag in the OSM editor when contributors open a task. The relevant hashtag for this study (i.e. the MapRoulette challenges described in sec. 4) are `#missing-cw_mapillary-signs` for traffic-sign based tasks and `#missing-cw_mapillary-feature` for tasks related to road markings. These hashtags allow edits resulting from the challenges to be reliably traced in the global OSM history.
Using the *ohsome-planet* framework[6], all changesets containing these hashtags were extracted and analyzed using our implementation[7].
This enables the reconstruction of the corresponding mapping activities and the derivation of aggregated statistics on newly added cycling infrastructure. In total, more than 500 km of previously unmapped cycling infrastructure were added to OpenStreetMap in Germany.



(a) Comparison of infrastructure mapping before (left) and after (right) the campaign.

(b) Segments modified during the campaign, colored according to their category prior to the campaign.

Figure 3: Cycling infrastructure added to OSM through the mapping campaign (identified via changeset hashtag `#missing-cw_mapillary-signs`).

A visualization of the corresponding changes is shown in Figure 3 and is also available online as an interactive application[8]. Figure 3a illustrates how infrastructure categories changed through the mapping campaigns, with source categories representing the relevant tags before the edit and

[6] Cf. ohsome-planet GitHub repository, last accessed 27 March 2026.
[7] Cf. `https://github.com/vizsim/osm_hashtag_analyse`, last accessed 30 March 2026.
[8] Cf. Interactive visualization of the mapping impact on GitHub, last accessed 30 March 2026.

target categories representing the resulting cycling infrastructure after the update. The accompanying map (Figure 3b) provides the spatial context by showing the segments modified during the campaign. Together, these visualizations allow a more detailed inspection of how missing infrastructure was added and how existing tagging was refined.

6 Beyond Mapping: Assessing Observability and Completeness

Beyond the direct improvement of OSM data, the proposed workflow also enables a more explicit assessment of the observability of cycling infrastructure. In many regions, the completeness of OSM cannot be evaluated directly because the underlying ground truth is unknown. However, street-level imagery introduces a new form of indirect evidence.

By combining image coverage density, temporal recency, and the reliability of detected infrastructure indicators, it becomes possible to estimate the degree to which cycling infrastructure in a given area is realistically observable and therefore mappable.

Importantly, this does not amount to a direct measurement of OSM completeness, nor does it provide a formal uncertainty estimate for the mapped cycling infrastructure. Rather, the proposed perspective is limited to observability: it indicates where the available street-level imagery provides a sufficient basis for detecting cycling-related infrastructure, and where such a basis is still missing. This distinction allows researchers and practitioners to differentiate between:

- data gaps caused by missing observations (i.e., areas where no recent imagery exists), and
- gaps caused by incomplete mapping despite sufficient visual evidence.

In other words, areas with dense and recent street-level imagery provide a much stronger basis for assessing data completeness than areas with sparse or outdated observations.

To support this analysis, we developed an interactive visualization tool that aggregates Mapillary coverage statistics at different administrative levels[9]. The tool summarizes the length and share of road segments that are covered by recent street-level imagery. Users can explore these statistics at state, district (Landkreis), or municipality level. The underlying coverage analysis determines for each road segment in the OSM network whether it is located within close proximity of recent Mapillary imagery, based on spatial overlap with buffered image sequences[10].

The analysis focuses on major roads derived from the current OSM network (up to tertiary highways), as these are the most relevant segments for cycling infrastructure detection. Figure 4 illustrates the resulting spatial patterns of coverage at the municipality level. Blue areas indicate a higher share of covered roads, while red areas represent locations where street-level imagery coverage remains limited.

[9] Cf. Interactive Mapillary Coverage visualization tool on GitHub, last accessed 30 March 2026.

[10] Cf. Coverage analysis implementation on GitHub, last accessed 30 March 2026.

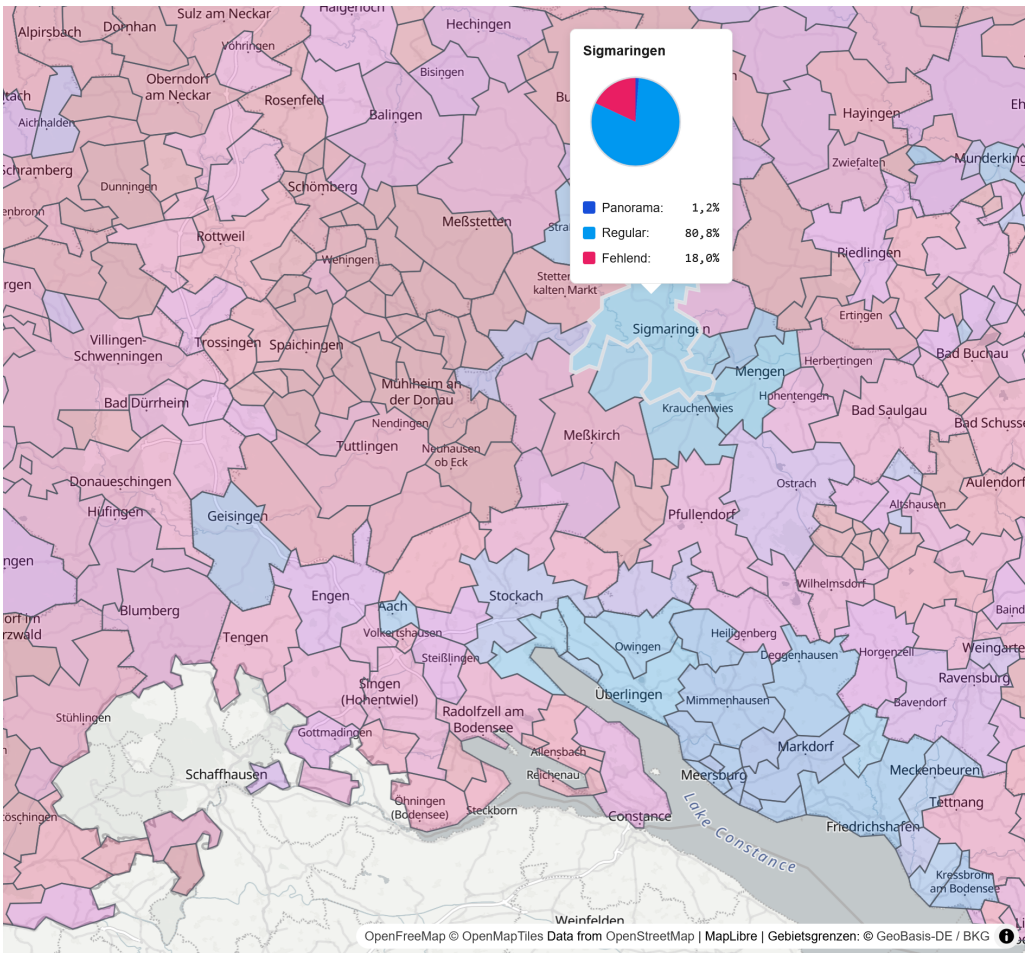


Figure 4: Aggregated coverage of the major road network at municipality level on Mapillary. Blue indicates a high share of covered roads, while red indicates low coverage.

Such observability metrics can help to contextualize future OSM completeness assessments and guide additional data collection efforts. For example, areas with low imagery coverage may require targeted image collection campaigns before any stronger statements can be made about the mapped infrastructure. Conversely, areas with high imagery coverage provide a more suitable basis for identifying potential mapping gaps. A formal assessment of completeness or uncertainty, however, lies beyond the scope of this paper.

7 Discussion and Transferability

The presented workflow demonstrates how automated street-level imagery detections and community-based mapping can be combined to systematically identify and address missing infrastructure in OpenStreetMap. The evaluation results show that the proposed filtering strategy leads to relatively low false-positive rates while generating a substantial number of validated mapping contributions. A key strength of the workflow is the integration of automated detection, spatial filtering, and community validation. Rather than directly modifying OSM data through automated processes, the pipeline translates detected indicators into targeted microtasks for human mappers. This ensures that final mapping decisions remain consistent with community practices and quality standards.
A relevant limitation of the approach is that the detection of potential infrastructure gaps depends on the availability and spatial coverage of recent street-level imagery. In regions with sparse or outdated imagery, the method cannot reliably identify missing infrastructure. In addition, not all cycling infrastructure is clearly marked by traffic signs or road markings, which may result in undetected infrastructure. This dependency directly constrains the observability of cycling infrastructure, as discussed in Section 6.
The developed workflow is highly transferable. The underlying principle, combining visual detections with crowdsourced validation tasks, can be applied to other types of transport infrastructure or geographic contexts. Potential applications include the identification of missing speed-limit

information, pedestrian crossings, or other traffic-related features that are visible in street-level imagery.

More broadly, the approach illustrates how computer vision outputs from large-scale imagery platforms can be integrated into participatory mapping workflows. By linking automated observations with community-driven validation, the pipeline provides a scalable framework for improving and assessing open transport datasets.

Acknowledgment

This work is part of the Collaborative Research Centre (CRC) AgiMo, which investigates data-driven approaches to responsible mobility planning and relies on high-quality geospatial data as a foundation for mobility modelling. The work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — TRR: 408/1 2025 — Project number: 531327426. We want to thank Tobias Jordans for inspiring discussions and for supporting the integration of relevant features into TILDA radinfra.de. We further acknowledge the contributions of the OpenStreetMap community, whose volunteered geographic data made this work possible.

References

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, *65*, 126–139. Retrieved from https://www.sciencedirect.com/science/article/pii/S0198971516303970 doi: 10.1016/j.compenvurbsys.2017.05.004

Ferster, C., Fischer, J., Manaugh, K., Nelson, T., & Winters, M. (2019). Using OpenStreetMap to inventory bicycle infrastructure: A comparison with open data from cities. *International Journal of Sustainable Transportation*, *14*(1), 64–73. Retrieved from http://dx.doi.org/10.1080/15568318.2018.1519746 doi: 10.1080/15568318.2018.1519746

FixMyCity and OSM Verkehrswende Community. (2026). *radinfra.de.* Retrieved from https://tilda-geo.de/regionen/radinfra (last accessed 30 March 2026)

Geofabrik. (2026). *OpenStreetMap Data Extracts.* Retrieved from https://download.geofabrik.de/ (last accessed 15 March 2026)

Hochmair, H. H., Zielstra, D., & Neis, P. (2014). Assessing the completeness of bicycle trail and lane features in OpenStreetMap for the United States. *Transactions in GIS*, *19*(1), 63–81. Retrieved from http://dx.doi.org/10.1111/tgis.12081 doi: 10.1111/tgis.12081

Kapp, A., Hoffmann, E., Weigmann, E., & Mihaljević, H. (2025). StreetSurfaceVis: a dataset of crowdsourced street-level imagery annotated by road surface type and quality. *Scientific Data*, *12*(1). Retrieved from http://dx.doi.org/10.1038/s41597-024-04295-9 doi: 10.1038/s41597-024-04295-9

Lißner, S., & Huber, S. (2021). Facing the needs for clean bicycle data – a bicycle-specific approach of GPS data processing. *European Transport Research Review*, *13*(1), 8. Retrieved from https://doi.org/10.1186/s12544-020-00462-2 doi: 10.1186/s12544-020-00462-2

Majic, I., Naghizade, E., Winter, S., & Tomko, M. (2021). There is no way! Ternary qualitative spatial reasoning for error detection in map data. *Transactions in GIS*, *25*(4), 2048–2073. Retrieved from http://dx.doi.org/10.1111/tgis.12765 doi: 10.1111/tgis.12765

Mapillary. (2026). *Mapillary.* Retrieved from https://www.mapillary.com (last accessed 30 March 2026)

MapRoulette. (2026). *MapRoulette - Microtasking platform for improving OpenStreetMap data.* Retrieved from https://maproulette.org (last accessed 23 March 2026)

OSM contributors. (2026). *OpenStreetMap contributors.* Retrieved from https://www.openstreetmap.org (Data from OpenStreetMap, licensed under the Open Database License (ODbL), last accessed 30 March 2026)

Topf, J. (2026). *Osmium Command Line Tool.* Retrieved from https://github.com/osmcode/osmium-tool (last accessed 15 March 2026)

Vierø, A. R., Vybornova, A., & Szell, M. (2023). BikeDNA: A tool for bicycle infrastructure data and network assessment. *Environment and Planning B: Urban Analytics and City Science*, *51*(2), 512–528. Retrieved from http://dx.doi.org/10.1177/23998083231184471 doi: 10.1177/23998083231184471

Ziemke, D., Metzler, S., & Nagel, K. (2017). Modeling bicycle traffic in an agent-based transport simulation. *Procedia Computer Science*, *109*, 923–928. Retrieved from https://www.sciencedirect.com/science/article/pii/S1877050917311043 doi: 10.1016/j.procs.2017.05.424