GeoAI-Powered Lane Matching for Bike Routes in GLOSA Apps

Philipp Matthes
TU Dresden
Germany
philipp.matthes@tu-dresden.de

Daniel Jeschor
TU Dresden
Germany
daniel.jeschor@tu-dresden.de

Thomas Springer
TU Dresden
Germany
thomas.springer@tu-dresden.de

ABSTRACT

Dense urban areas like Hamburg strive to increase the attractiveness of cycling to overcome mobility-related problems such as space limitations, air pollution, and noise levels. To address this issue, smart mobility solutions can encourage more people to choose cycling. Green Light Optimal Speed Advisory for bikes (bike-GLOSA) can reduce the number of stops at red lights, allow smoother traveling, and convey a digital advantage to cyclists. However, the city-wide implementation of bike-GLOSA introduces new challenges, including the need for automated lane matching. Humans may employ spatial reasoning to determine the most logical sequence of lanes and associated traffic lights across intersection topologies for a given route. In this paper, we aim to replicate this spatial reasoning using Geospatial Artificial Intelligence (GeoAI). The proposed Machine Learning (ML) model not only overcomes limitations associated with location- and camera-based lane matching approaches. It also outperforms a previous route-based approach by 8%, with an F1 score of 92% on our test dataset. We critically examine our approach and real-world results to identify potential limitations and avenues for future research.

CCS CONCEPTS

• **Computing methodologies** → **Spatial and physical reasoning**; • **Information systems** → *Geographic information systems*; • **Applied computing** → Transportation.

KEYWORDS

Geospatial AI, Intelligent Transportation, Smart Cities, Spatial Modeling and Reasoning, Spatial Pervasive Computing

ACM Reference Format:
Philipp Matthes, Daniel Jeschor, and Thomas Springer. 2023. GeoAI-Powered Lane Matching for Bike Routes in GLOSA Apps. In *The 31st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23), November 13–16, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3589132.3625583

1 INTRODUCTION

Sustainable cities require innovative solutions to address the negative impacts of urban mobility, such as air and noise pollution and the inefficient use of space. One way to promote sustainability is to encourage non-motorized transportation, giving cyclists certain advantages in traffic flow [1]. A digital and non-invasive solution to this challenge is a bike-GLOSA app. Such an app advises speeds to cyclists to catch as many green lights as possible based on real-time signal data [2, 3, 4, 5].

However, some substantial challenges must be addressed to make bike-GLOSA a viable real-world application. One of these challenges is selecting upcoming traffic lights from far away to provide a speed advisory with enough reaction time to the user [6, 7]. This challenge includes two problems: Firstly, the GLOSA system needs to predict the cyclist's path. Secondly, the system must identify the most suitable traffic lights along that pathway. Unlike cars, it cannot be assumed that cyclists approach an intersection in a predictable, straightforward manner. The optimal route across an intersection may occasionally involve traffic lights shared with motor vehicles or pedestrians, often resulting in multiple ambiguous options for crossing an intersection.

In this paper, we aim to resolve this ambiguity and predict which sequence of signals a cyclist will traverse at an intersection. We propose an approach based on lane geometries of traffic lights, which model the turn shapes across an intersection. We utilize these

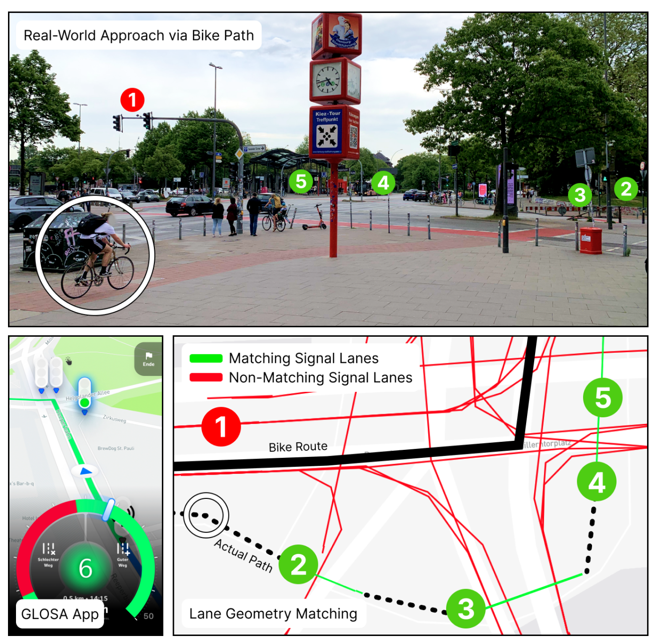

Figure 1: Signal 1 is a close match to the route but does not correspond to the actual bike path. Spatial reasoning is required to determine the most likely sequence of bike signals (2, 3, 4, 5). The final matching result is shown on the left.

SIGSPATIAL '23, November 13–16, 2023, Hamburg, Germany

ACM ISBN 979-8-4007-0168-9/23/11…$15.00
https://doi.org/10.1145/3589132.3625583

turn shapes to compare which traffic lights match a user-defined route best. However, this process involves spatial reasoning, as the generated route may not follow the cyclist's actual path across an intersection (see Figure 1). The goal is to teach spatial reasoning to a GeoAI model through human-generated examples, obtaining a model that can accurately preselect all traffic lights along a route.

The rest of this work is structured as follows. We review related work in Section 2, present our GeoAI model in Section 3, evaluate its performance in Section 4, and discuss implications and future research in Section 5.

2 STATE OF THE ART

Lane matching is tightly related to vehicle perception on roads and intersections for (semi)autonomous driving. In this problem domain, approaches with cameras and advanced sensor fusion represent the state-of-the-art [8, 9]. Similarly, Koukoumidis et al. (2011) have demonstrated signal matching for GLOSA with a windshield-mounted smartphone and its camera [10]. Nonetheless, alternative approaches must be explored as cyclists may either attach their smartphones to their handlebars or prefer using a GLOSA system through acoustic or vibrotactile interaction from their pocket.

In the car-centric field of GLOSA, a recurring approach is matching signals through the vehicle's location, as presented by Bernais et al. (2016) [11], Hao et al. (2018) [6], Stahlmann et al. (2018) [12] and Bhattacharyya et al. (2022) [13]. One reported challenge is inaccurate geopositioning and subsequent lane mismatching [13]. However, aside from general usage experiences, methodological details or benchmarks on the matching processes are sparse. Thus, comparison is difficult, and it is not clear how the proposed approaches would perform in a bike context. As highlighted in Figure 2, our largest concern is that short (median = 1m) bike lane geometries may delay an unambiguous matching until the user is located a few meters away from the signal, diminishing the speed advisory's effectiveness.

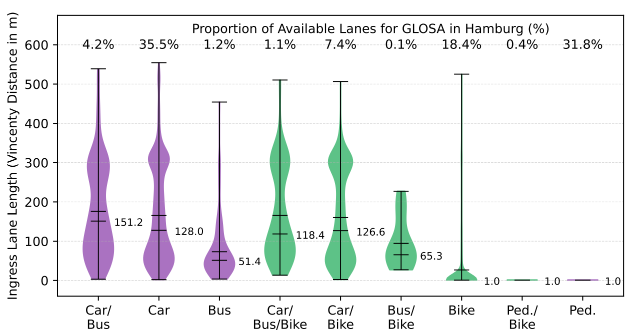

Figure 2: Distribution and medians of ingress lane lengths in Hamburg. Green lane types are relevant for bike-GLOSA.

Another open question is how to match multiple consecutive traffic lights for multi-segment GLOSA [14, 15, 16] without predicting upcoming turns. Although advanced map-matching methods have shown the ability to accurately match scattered user trajectories to a map reference [17, 18, 19], they still face the challenge of predicting the matched trajectory across upcoming intersections and road forks.

Route-based matching approaches may circumvent all aforementioned problems. However, an ongoing challenge of routing foundations such as OpenStreetMap is that bike paths are not captured consistently [20, 21]. Aside from the option to improve the map's accuracy [22], a more refined matching can be applied that compares the geometric properties between routes and lanes. In initial work, we proposed a matching procedure that excludes nearby lanes not matching the geometric properties of the route [23]. This approach works with inaccurate OpenStreetMap bike routes but applies its thresholds equally across all situations, limiting its robustness against edge cases.

In summary, route-based lane matching approaches are currently the most promising approach to fulfill the practical constraints of bike-GLOSA, compared to location-based and camera-based methods. However, the main limitation is routing inaccuracies. Hence, it is crucial to find more advanced models to improve the error-robustness of route-based lane matching.

3 GEOAI MODEL FOR LANE MATCHING

To find an ML model that reasonably decides between matching and non-matching lane geometries along a route, we require a ground truth dataset and a suitable benchmark. Through a systematic search for models and features on our dataset, we identify and propose the most suited ML pipeline for our task's characteristics.

3.1 Ground Truth and Benchmark

To create our ground truth dataset, we label matching lanes by hand along randomly generated routes in Hamburg, our deployment area. The routes are generated through OpenStreetMap and GraphHopper's `bike2` routing profile. To reduce the impact of personal preference on the decision process beyond objective spatial reasoning, we adhere to a fixed set of labeling rules:

(1) Consideration is given to traffic flow and the feasibility of safely maneuvering through the intersection.
(2) The intersection traversal must be entirely legal according to turn restrictions and street laws.
(3) Preference is given to dedicated bicycle lanes or paths indicated by distinct markings or signage.
(4) The overall continuity of the route is maintained to ensure a smooth and logical progression.
(5) Lanes on the wrong roadside must be avoided.
(6) Lanes should never be in an opposing direction to the route.

With this ruleset, we match 1043 lane geometries to 149 generated routes and note down routing errors. For additional 5397 lane-route constellations, the decision "no match" is made. We eliminate 3092 duplicates out of 6440 samples from routes that pass through intersections in the same way, avoiding false impressions that the model works well on unseen data. Artificial augmentation (scaling, translating, and rotating) of lane geometries is not used due to unsatisfactory preliminary test results.

The benchmark for our model requires accounting for the imbalanced nature of the dataset. Hence, we utilize stratified k-fold cross-validation to prevent our results from depending on a relatively hard or easy test split. All parts of the model pipeline, including feature scalers, are only fitted on the training proportion, ensuring that no test data leaks into the training process. The F1 score is selected as the core evaluation metric, as it is not boosted by the expected high number of true negatives when excluding far-away traffic lights from the route.

3.2 Model Architecture

Based on the task's characteristics and the available dataset size, we focus on traditional ML architectures as provided by the Scikit-Learn framework: k-NN, Decision Tree, Random Forest, MLP, AdaBoost, Gaussian Naive Bayes, QDA, and SVM. After a grid search, the best F1 score could be achieved with the MLP.

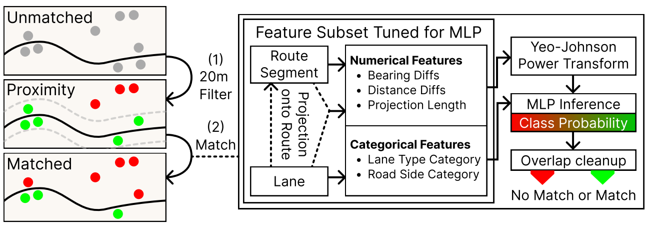

Figure 3: Data flow diagram of our final ML approach.

The model's processing pipeline visible in Figure 3 consists of two steps: First (1), we *filter* out all lanes more than 20 meters from the route. Second (2), we *match* the remaining lane geometries to the route. By projecting each lane onto the route geometry, we obtain one sample per lane geometry with the corresponding section of the route. Based on recursive feature elimination and iterative model fine-tuning, the following features were found to be the most effective out of an initially larger feature set. They are calculated for each sample:

- *Lane Type:* One-hot-encoded type of lane extracted from the signal's metadata: Bike, Bike/Pedestrian, Bike/Car, Bike/Car/Bus, and Bike/Bus.
- *Road Side:* Side of the lane geometry relative to the route section, based on the side of the closest point on the route to the first lane geometry point. One-hot encoded as: "left," "right," or "no_side" if in-between.
- *Bearing Diffs:* Maximum and minimum difference of each lane segment's angle to its corresponding segment on the route section. The difference between the angles of the route section's first and last segment is also included as a feature.
- *Distance Diffs:* Distances between the lane's first and last coordinates and their corresponding points on the route section. Another added feature is the shortest distance between both geometries.
- *Projection Length:* Length of the corresponding route section.

All non-categorical features are subjected to a *Yeo-Johnson power transformation* [24] to standardize the values and optimize model performance by symmetrizing the value distribution. Finally, the preprocessed values are piped into the trained MLP model. In the output layer, the neural activations are combined to generate a probability for "match" or "no match". This process is repeated for each signal along the route within the initial 20m radius, resulting in a list of matched signals. To avoid overlapping selections, as highlighted in Figure 4, we perform an *overlap cleanup* step. If projected lane geometries overlap, the lane with higher activation in the model's output neuron for "match" is selected.

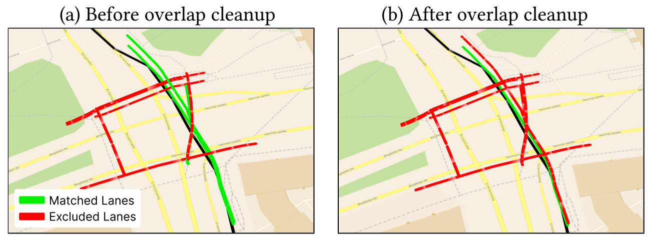

Figure 4: Overlap cleanup using the MLP's class probability.

4 RESULTS

To ensure our model's generalization across various scenarios, it is crucial to have a diverse dataset encompassing a wide range of route-lane combinations and lane shapes. As shown in Figure 5, the generated routes cover a broad area within the city, reaching nearly all of the 356 intersections with overall 2414 bike traffic lights. Each lane in the ground truth is typically only associated with one (323) or two (105) routes, indicating low dataset redundancy in repeatedly traversing the same lanes. Overall, 73.7% (1780) of all 2414 bike lanes were considered at least once in the ground truth, ensuring that the model learns from and is tested on a substantial portion of the available bike lanes.

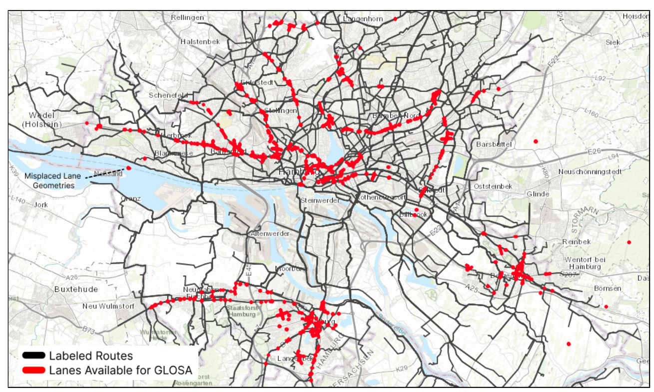

Figure 5: Labeled routes in our ground truth.

4.1 Benchmark and Real-world Performance

Overall, our model achieves a test F1 score of 92% (training: 94%). This marks a considerable improvement over our previous approach [23], which only achieved 84% on our dataset.

The main source of improvement is cases with imperfect routing, which occur in 77% of all samples, as tagged during the dataset generation process. In 72% of cases, cyclists were forced by the route onto the road despite the availability of a bike lane. In 5% of the samples, the route forced cyclists to the wrong roadside. While our ML model correctly identified 221 out of 257 cases with routing errors, the previous approach [23] only identified 200 cases. Hence, our ML model has a higher chance of mitigating a routing error with 86% compared to 78%. Cases without routing errors are mismatched with a slightly higher chance of 4.85% with our model compared to 3.80%. In 474 such cases, the ML model correctly recognized 451 lanes, compared to 456 lanes with the previous approach.

To validate the calculated F1 score in the real world, we recorded four test drives in Hamburg with camera and screen recordings, finding that 51 out of 56 traffic lights were correctly matched. This aligns with the measured results on our ground truth.

4.2 Discussion

Our results demonstrate that the proposed model can match lane geometries of bike signals with high accuracy, even in the presence of routing errors, taking less than a second for a typical route. This ensures scalability and responsiveness in our bike-GLOSA app. Our approach maximizes activation distance independently of lane length or geolocation accuracy, enabling multi-segment and multimodal speed advisory.

However, there are also limitations. Foremost, we cannot guarantee that all ground truth examples always correspond to the most likely path chosen by users. Similarly, some users will inevitably choose a different path than proposed by our model. We assume that this case is sufficiently rare. Also, we cannot yet underline if our model applies outside of its intended application area (Hamburg). Model fine-tuning may be necessary to reach the reported score with changing intersection and route characteristics depending on the city. Optimistically, reaching higher scores in other cities or with a higher accuracy routing foundation may also be possible.

5 CONCLUSIONS AND FUTURE WORK

Route-based lane matching is a key solution to establishing bike-GLOSA apps on the city scale, a measure to motivate more environmentally friendly street traffic. We demonstrate a GeoAI model that aims to match the most suitable sequence of bike signals at an intersection, even if the route does not accurately follow the bike path. Our model achieves an F1 score of 92% on our ground truth dataset, improving on a previous approach by over 8%. Investigations of our ground truth dataset and field tests indicate that our approach generalizes well on a city scale and in real-world usage.

Addressing the lane matching challenge is an important avenue toward operating bike-GLOSA within or above the city scale. Only if the lane matching challenge is overcome will such a system be able to impact today's society and mobility behavior. Now that a technical foundation exists, multi-segment approaches and multimodal interaction methods should be reconsidered. The ultimate goal is to make bike-GLOSA systems practical to find out whether they really motivate people to cycle. Our route-based approach may be worthwhile for cars as well, although it may require additional considerations. Our implementation is available under GitHub.com/priobike/priobike-sg-selector.

Acknowledgments: This work is funded by the Federal Ministry for Digital and Transport (BMDV) project PrioBike-HH (FKZ: 16DKVM001B). Map tiles for figures are provided by Mapbox, OpenStreetMap, and CARTO.

REFERENCES

[1] Francesco Paolo Appio, Marcos Lima, and Sotirios Paroutis. 2019. Understanding smart cities: innovation ecosystems, technological advancements, and societal challenges. *Technological Forecasting and Social Change*, 142.

[2] Irina Tal, Bogdan Ciubotaru, and Gabriel-Miro Muntean. 2015. Vehicular-communications-based speed advisory system for electric bicycles. *IEEE Transactions on Vehicular Technology*, 65.

[3] Sven Fröhlich, Thomas Springer, Stephan Dinter, Sebastian Pape, Alexander Schill, and Jürgen Krimmling. 2016. Bikenow: a pervasive application for crowdsourcing bicycle traffic data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct.*

[4] Stephen Fickas, Marc Schlossberg, et al. 2019. Fast Track: Allowing Bikes to Participate in a Smart Transportation System. Tech. rep. National Institute for Transportation and Communities (NITC).

[5] Azita Dabiri, Andreas Hegyi, and Serge Hoogendoorn. 2020. Optimized speed trajectories for cyclists, based on personal preferences and traffic light information-a stochastic dynamic programming approach. *IEEE Transactions on Intelligent Transportation Systems*, 23.

[6] Peng Hao, Guoyuan Wu, Kanok Boriboonsomsin, and Matthew J Barth. 2018. Eco-approach and departure (ead) application for actuated signals in real-world traffic. *IEEE Transactions on Intelligent Transportation Systems*, 20.

[7] Hironori Suzuki and Yoshitaka Marumo. 2020. Safety evaluation of green light optimal speed advisory (glosa) system in real-world signalized intersection. *Journal of Robotics and Mechatronics*, 32.

[8] NJ Zakaria, MI Shapiai, RA Ghani, MNM Yasin, MZ Ibrahim, and N Wahid. 2023. Lane detection in autonomous vehicles: a systematic review. *IEEE Access.*

[9] Yeongho Son, Elijah S Lee, and Dongsuk Kum. 2019. Robust multi-lane detection and tracking using adaptive threshold and lane classification. *Machine Vision and Applications*, 30.

[10] Emmanouil Koukoumidis, Li-Shiuan Peh, and Margaret Rose Martonosi. 2011. Signalguru: leveraging mobile phones for collaborative traffic signal schedule advisory. In *Proceedings of the 9th international conference on Mobile systems, applications, and services.*

[11] Benedikt Bernais, Andreas Lotz, and Hongjun Pu. 2016. Design and implementation of a traffic light assistance system based on c2x and tsi messages. In *AmE 2016-Automotive meets Electronics; 7th GMM-Symposium.* VDE.

[12] Rainer Stahlmann, Malte Möller, Alexej Brauer, Reinhard German, and David Eckhoff. 2018. Exploring glosa systems in the field: technical evaluation and results. *Computer Communications*, 120.

[13] Kinjal Bhattacharyya, Pierre-Antoine Laharotte, Arthur Burianne, and Nour-Eddin El Faouzi. 2022. Assessing connected vehicle's response to green light optimal speed advisory from field operational test and scaling up. *IEEE Transactions on Intelligent Transportation Systems.*

[14] Marcin Seredynski, Wojciech Mazurczyk, and Djamel Khadraoui. 2013. Multi-segment green light optimal speed advisory. In *2013 IEEE International Symposium on Parallel & Distributed Processing, Workshops and Phd Forum.* IEEE.

[15] Mahdi Sharara, Marc Ibrahim, and Gerard Chalhoub. 2019. Impact of network performance on glosa. In *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC).* IEEE.

[16] Lior Simchon and Raul Rabinovici. 2020. Real-time implementation of green light optimal speed advisory for electric vehicles. *Vehicles*, 2.

[17] Paul Newson and John Krumm. 2009. Hidden markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems.*

[18] Ilze Andersone. 2019. Heterogeneous map merging: state of the art. *Robotics*, 8.

[19] Zhenfeng Huang, Shaojie Qiao, Nan Han, Chang-an Yuan, Xuejiang Song, and Yueqiang Xiao. 2021. Survey on vehicle map matching techniques. *CAAI Transactions on Intelligence Technology*, 6.

[20] Anastassia Vybornova, Tiago Cunha, Astrid Gühnemann, and Michael Szell. 2022. Automated detection of missing links in bicycle networks. *Geographical Analysis.*

[21] David Wasserman, Alex Rixey, Xinyi Zhou, Drew Levitt, and Matt Benjamin. 2019. Evaluating openstreetmap's performance potential for level of traffic stress analysis. *Transportation research record*, 2673.

[22] Philipp Matthes, Thomas Springer, and Jeschor Daniel. (in press). Accurate bike routing for lane prediction in glosa apps via infrastructure reference models. In *2023 IEEE International Conference on Intelligent Transportation Systems (ITSC).* IEEE.

[23] Philipp Matthes and Thomas Springer. 2022. Matching traffic lights to routes for real-world deployments of mobile glosa apps. In *2022 IEEE International Smart Cities Conference (ISC2).* IEEE.

[24] In-Kwon Yeo and Richard A Johnson. 2000. A new family of power transformations to improve normality or symmetry. *Biometrika*, 87.