

Advancing Digital Twins for Urban Mobility in Sub-Saharan Africa: Feature-Based Road Classification.

* 04601-A: ICT in Africa Seminar

Nchofon Tagha Ghogomu
College of Engineering
Carnegie Mellon University Africa
 ntaghagh@andrew.cmu.edu

Abstract—The potential of Digital Twins in driving sustainable digital transformation is increasingly gaining attention. DTs play a crucial role in simulation, enabling effective planning and decision-making for large-scale systems. The global digital twin industry, valued at USD 16.75 billion in 2023, is projected to grow at an annual rate of 35.7% until 2030. However, the adoption of DTs for urban planning in Sub-Saharan Africa remains limited, primarily due to perceived cost barriers. Additionally, urban mobility dynamics in SSA differ significantly from those in the Global North due to variations in road network structures, vehicle types, and driving behaviors. Researchers at Carnegie Mellon University Africa have been investigating the use of DTs to enhance urban mobility in Rwanda and across Africa. Previous studies have focused on the temporal analysis of trips in Kigali using various transportation modes and have proposed a computer vision-based methodology for modeling the mobility patterns of moto-taxis in SSA. Leveraging and fine-tuning a YOLOv12n object detection model, we explore classifying roads, sidewalks, and crosswalks using an open-source dataset augmented with custom-labeled images from Kigali. Through a three-iteration training pipeline involving augmentation and data relabeling, we show that computer vision can be used for road feature extraction towards a context-based DT for urban mobility. This study yielded a final average mAP@0.5 of 0.723 for all classes. This work serves as a practical baseline for future digital twin modeling efforts in emerging regions.

Index Terms—digital twin, urban mobility, road modeling, computer vision, deep learning, feature extraction.

I. INTRODUCTION

Digital Twins (DTs), broadly defined as virtual representations of physical objects or systems that dynamically emulate their real-world counterparts [1], are emerging as a transformative tool for sustainable digital transformation. The global DT market, valued at USD 16.75 billion in 2023, is projected to grow at a compound annual rate of 35.7% through 2030 [2]. While predictive maintenance and industrial optimization in the Global North dominate current applications [2], the potential of DTs to address infrastructure and mobility challenges in under-served regions, particularly SSA, remains underexplored. Barriers such as high implementation costs, limited digital infrastructure [3], and a shortage of skilled expertise hinder adoption in SSA, yet the urgency of improving

urban planning and transportation systems in rapidly growing African cities underscores the need for scalable, context-sensitive DT frameworks.

Recent advancements in DT development highlight their versatility. For instance, the Clifton Suspension Bridge (UK) case study demonstrated a workflow for retrofitting existing infrastructure with sensor networks to create real-time structural health monitoring systems [4]. Similarly, DTs have been applied to detect impaired driving via deep learning models analyzing vehicle dynamics [5] and to model road networks using edge computing for real-time environmental data [6]. However, these applications largely reflect Global North contexts, relying on assumptions about standardized road infrastructure, driving behaviors, and regulatory frameworks that diverge sharply from SSA's realities. Tools like Simulation of Urban Mobility (SUMO), while valuable for traffic modeling [7], lack the granularity to capture informal road networks, mixed-traffic patterns like with moto-taxis, and adaptive driving practices prevalent in African urban centers.

To arrive at the desired results, this project answers the following question:

- i. What is the state of the art of DTs
- ii. What features generally define a road stretch in the context of SSA?
- iii. How can deep learning be used for the real-time extraction of these features?
- iv. How performant are deep learning models is feature extraction?

Focusing on Kigali, Rwanda—a representative SSA city with rapid urbanization and moto-taxi-dominated mobility, we propose a framework leveraging computer vision to classify roads using dashcam footage. To do this, this project leverages a deep learning model for real-time road feature extraction.

II. LITERATURE REVIEW

The global DT market exhibits sustained growth, yet adoption rates remain disproportionately low in Sub-Saharan Africa (SSA) [2], reflecting infrastructural and technological disparities. Recent scholarly efforts have prioritized methodological frameworks for DT implementation. For instance,

Buuveibaatar et al. introduced a five-layered architecture for road DT planning, encompassing data acquisition, processing, modeling/algorithms, analysis/control, and service layers [3]. Similarly a workflow for retrofitting DTs into existing built assets, demonstrated through a proof-of-concept case study on Bristol’s Clifton Suspension Bridge was proposed [4]. While these frameworks advance procedural rigor, their focus remains narrowly technical, with limited exploration of DT applications in mobility optimization or real-world operationalization.

Emerging applications demonstrate DTs’ potential in road infrastructure management. Recent studies leverage edge computing and deep learning for real-time impaired driving detection [8] and environmental monitoring [9]. Concurrently, CMU-Africa’s ongoing initiative to develop a DT of Rwanda’s road network represents a pioneering effort in SSA, though progress remains nascent, currently limited to temporal trip analysis in Kigali and mobility pattern mapping for moto-taxis. These works however, target particular aspects of mobility and do not capture road features for effective road modeling.

Road classification methodologies are foundational to DT accuracy. Remote sensing techniques have enabled automated road network extraction [9], while machine learning models, including hyperparameter-tuned deep autoencoders, classify road surfaces such as wet, icy, rough, or curvy [10]. However, these efforts remain siloed; classification outputs are seldom integrated into DT frameworks. Notably, hypergraph-based approaches [11] model high-order road network relationships but lack domain-specific adaptation to SSA’s unique road conditions (e.g., unpaved roads, informal pathways). A pressing need exists for regionally tailored road feature datasets to support DT-driven simulations in underrepresented contexts

III. METHODOLOGY

This research was broken down into milestones at the beginning as seen in Table I. In this section, we will explore the data used, the model, and the evaluation matrices.

#	Week	Status
1	Week 1: Data Collection & Annotation	Completed
1.1	Identify key roads in Kigali for data collection	Completed
1.2	Capture and store dashcam footage	Completed
1.3	Organize and structure the dataset	Completed
2	Week 2: Preprocessing & Data Augmentation	Completed
2.1	Annotate road features using RoboFlow	Completed
2.2	Resize images to YOLO input specifications	Completed
2.3	Apply augmentation techniques	Completed
3	Week 3: Model Training	Completed
3.1	Train YOLOv12 on annotated dataset	Completed
3.2	Optimize hyperparameters	Completed
3.3	Monitor training loss and refine model	Completed
4	Week 4: Model Validation & Final Evaluation	Completed
4.1	Evaluate model with mAP and F1-score	Completed
4.2	Validate curvature estimation with RMSE	Completed
4.3	Assess robustness in real-world SSA conditions	In progress

TABLE I
PROJECT TIMELINE WITH KEY MILESTONES

A. Data, Sources, and Preprocessing

The dataset utilized in this study is an open-source collection titled Sidewalk Dataset (v6) [12], curated and hosted on

Roboflow Universe. It contains 1,603 images annotated for instance segmentation, aimed at supporting the development of computer vision models for detecting and segmenting sidewalks across diverse urban and peri-urban settings. The dataset includes pre-labeled classes of road features aligned with the objectives of this research. Its high-quality, diverse imagery, featuring road characteristics relevant to Sub-Saharan environments, made it a suitable foundation for this project.

A summary of the original Sidewalk Dataset (v6) is provided in Table II. Based on insights from the initial two development iterations, a portion of the dataset was relabeled to better capture project-specific features, and an additional 52 images depicting roads in Kigali were incorporated to enrich contextual relevance. The final curated and labeled dataset is publicly accessible on Roboflow at: <https://app.roboflow.com/nchofon/road-bcmta/2>, with its composition detailed in Table III.

In the first iteration, no image preprocessing was applied, which led to poor crosswalk classification performance. To address this, the second iteration introduced targeted data augmentation, focused on crosswalk-containing images, using the Albumentations library. Despite these efforts, performance improvements were minimal. Consequently, in the final iteration, images were relabeled with more precise class boundaries, and an additional 10% of the dataset was supplemented with Kigali road images to enhance contextual relevance and model accuracy.

Attribute	Description
Total Images	1,603
Train Set	1,130 images (70.5%)
Validation Set	362 images (22.6%)
Test Set	111 images (6.9%)
Labeled Classes	Crosswalk, Sidewalk, Road

TABLE II
SUMMARY OF THE SIDEWALK DATASET.

Attribute	Description
Total Images	513
Train Set	360 images (70%)
Validation Set	102 images (20%)
Test Set	51 images (10%)
Labeled Classes	Crosswalk, Sidewalk, Road
Preprocessing	Auto-orient, Resize to 640x640

TABLE III
SUMMARY OF THE ROAD DATASET BY AUTHOR

B. Model and Training

The model fine-tuned in this study is YOLOv12n, a lightweight object detection architecture from the Ultralytics YOLOv8 series, optimized for real-time performance in resource-constrained environments. YOLOv12n comprises 159 fused layers, approximately 2.56 million parameters, and operates at 6.3 GFLOPs, making it well-suited for deployment on modest hardware, such as mobile or embedded systems. The model was trained for 200 epochs using Ultralytics v8.3.113 on a Tesla P100 GPU (16GB VRAM) via the Kaggle environment, enabling efficient gradient propagation and batch-level optimization.

C. Evaluation Matrix

The performance of this model was evaluated using the standard matrices:

- 1) *Training/Validation Loss*: Comprising classification, objectness, and bounding box regression components.
- 2) *mAP@0.5 and mAP@0.5:0.95*: Mean Average Precision at different Intersection-over-Union (IoU) thresholds to assess both loose and strict localization accuracy.
- 3) *Precision-Recall Curves*: To visualize trade-offs across threshold values.
- 4) *Confusion Matrix*: The extent to which the model classifies or misclassifies an object.

IV. RESULTS.

There was progressive improvement in results over various iterations. The PR results for the first, second, and third iterations are summarized in Table IV. Also, to understand how good the prediction was, the confusion matrices of these three iterations were compared and summarized in Table V. The PR curve and confusion matrix of the final iteration is seen in Fig. 1 and 2 respectively. Finally, a summary of the model training for 200 epochs is seen in Fig. 3.

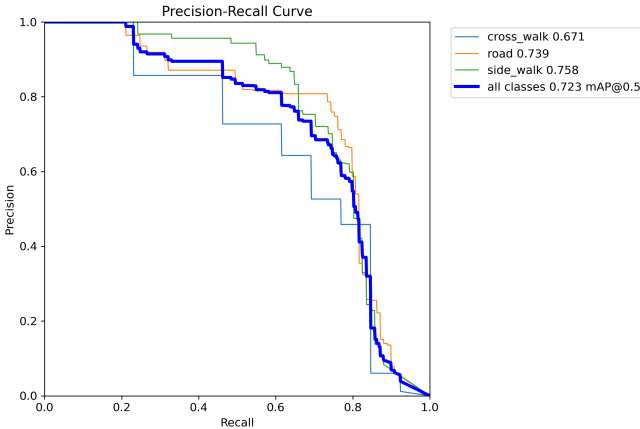


Fig. 1. PR Curve - 3rd, iteration.

V. DISCUSSION

The results from the three training iterations provide key insights into the complexities of road feature extraction in Sub-Saharan urban environments. While data augmentation contributed to modest performance improvements, particularly in crosswalk recall between the first and second iterations, the most notable performance gain emerged only after relabeling and refining the dataset in the third iteration.

This underscores the pivotal role of accurate and context-sensitive annotation over augmentation techniques. In the first two iterations, augmentations such as CLAHE, sharpening, and brightness adjustments introduced realistic variability, yet the model continued to underperform in detecting crosswalks, achieving average precisions (AP) of just 0.203 and 0.264, respectively. Following relabeling, crosswalk AP surged to

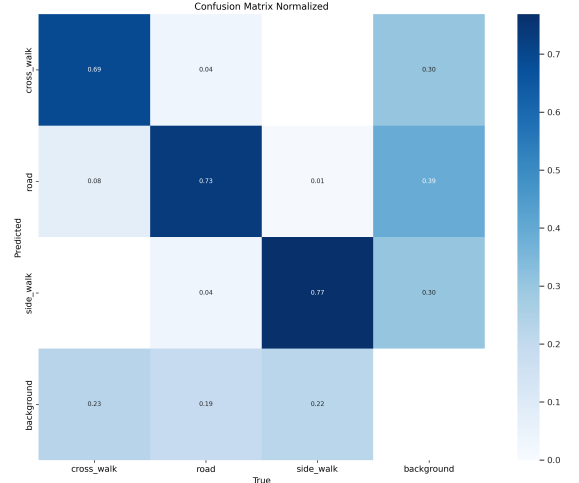


Fig. 2. Normalized Confusion Matrix of Final Model.

0.671, reflecting a threefold improvement. Similarly, class-wise accuracy in the confusion matrix improved from 0.27 to 0.69, affirming that enhanced labeling had a significantly greater impact on model performance than augmentation alone. This finding is consistent with the established understanding that augmentation can only amplify features that are already clearly and accurately annotated.

Furthermore, confusion with the background class remained a consistent challenge, particularly in early training phases. This is anticipated given the subtle visual differences between features like sidewalks and crosswalks and their surroundings in many Sub-Saharan road scenes, where markings may be worn, inconsistent, or absent. Despite this, the iterative improvement in precision and recall for sidewalks and roads suggests that the YOLOv12n model was able to progressively learn more robust and generalizable features when trained with better-labeled data.

VI. CONCLUSION

This study demonstrates the feasibility of using deep learning, specifically the YOLOv12n model, for extracting road features such as sidewalks, roads, and crosswalks in Sub-Saharan urban settings. The progressive improvement across training iterations, culminating in an mAP@0.5 of 0.723, highlights the importance of accurate, context-aware labeling. These findings confirm that computer vision can effectively support the development of region-specific digital twins, offering a scalable pathway to enhance urban mobility planning in emerging regions like Kigali.

VII. FUTURE WORK

For future work, the model will be extended to detect additional road features, including surface conditions and curvature, to enable a more comprehensive digital representation of roads. Enhancements will also focus on training with region-specific imagery to improve contextual accuracy and validating performance using real-world dashcam footage as originally planned.

TABLE IV
MODEL PERFORMANCE ACROSS TRAINING ITERATIONS

Training Iteration	mAP@0.5	Crosswalk AP	Road AP	Sidewalk AP
Iteration 1	0.546	0.203	0.723	0.712
Iteration 2	0.540	0.264	0.699	0.657
Iteration 3 (Final)	0.723	0.671	0.739	0.758

TABLE V
CLASS-WISE NORMALIZED ACCURACY FROM CONFUSION MATRICES ACROSS TRAINING ITERATIONS

Training Iteration	Crosswalk Accuracy	Road Accuracy	Sidewalk Accuracy	Background Accuracy
Iteration 1	0.27	0.67	0.68	0.26
Iteration 2	0.29	0.63	0.61	0.33
Iteration 3 (Final)	0.69	0.73	0.77	0.22

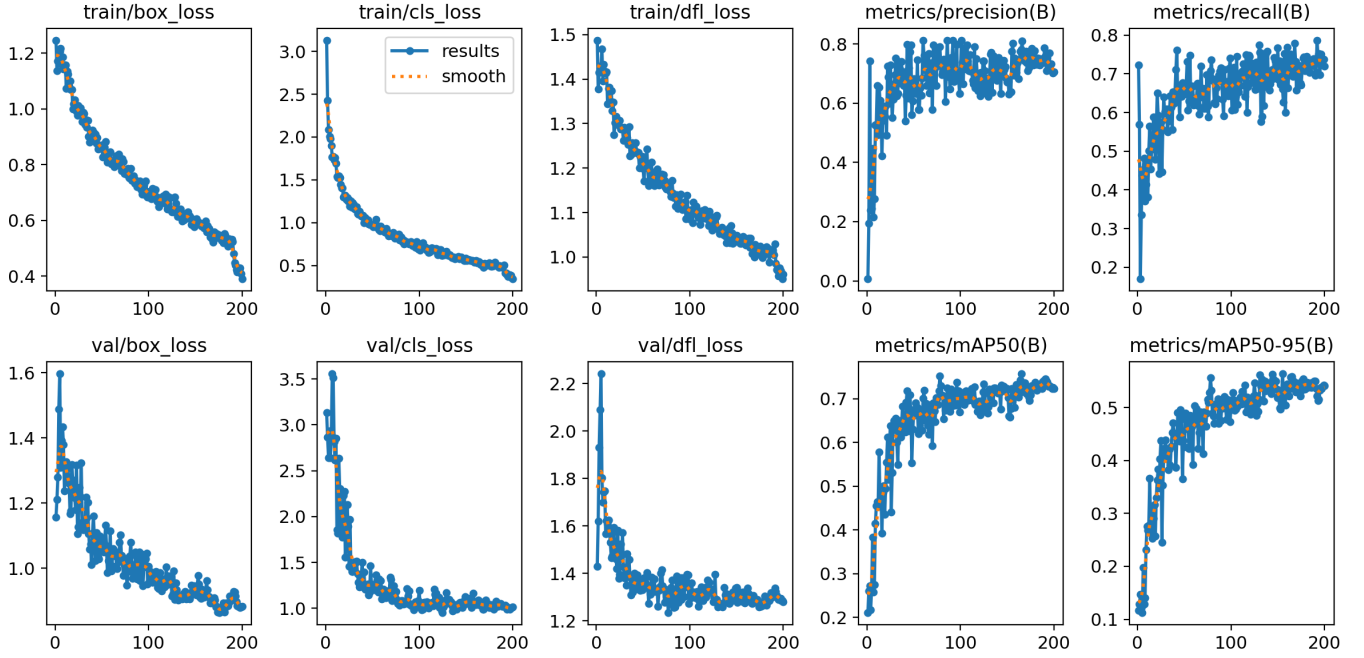


Fig. 3. Model Training Summary.

APPENDIX A CLASSIFICATION RESULTS

The prediction results of the first batch of test images is as shown in Fig. 4.

APPENDIX B RESPONSIBLE USE OF AI

I admit that AI was used in this project primarily to organize the presentation of ideas and results. The output was critically analyzed, compared with the original work, and corrected to meet the desired outcome. AI was also used to organize milestones and set reasonable timelines that was closely followed in the execution of this project.

REFERENCES

- [1] IBM. (2021, Aug.) What is a digital twin? Accessed: Mar. 28, 2025. [Online]. Available: <https://www.ibm.com/think/topics/what-is-a-digital-twin>
- [2] MarketsandMarkets. (2023, Jul) Digital twin market size, share, and trends. Accessed: Mar. 28, 2025. [Online]. Available: <https://www.marketsandmarkets.com/Market-Reports/digital-twin-market-225269522.html>
- [3] M. Buuveibaatar, I. Brilakis, M. Peck, G. Economides, and W. Lee, "A conceptual framework for planning road digital twins," *Buildings*, vol. 15, no. 3, 2025. [Online]. Available: <https://www.mdpi.com/2075-5309/15/3/316>
- [4] M. Pregnolato, S. Gunner, E. Voyagaki, R. De Risi, N. Carhart, G. Gavriel, P. Tully, T. Tryfonas, J. Macdonald, and C. Taylor, "Towards civil engineering 4.0: Concept, workflow and application of digital twins for existing infrastructure," *Automation in Construction*, vol. 141, p. 104421, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0926580522002941>
- [5] S. J. Siddiqi, S. Saleh, M. A. Jan, and M. Tariq, "Securing the vetaverse: Web 3.0 for decentralized digital twin-enhanced vehicle-road safety," *Future Generation Computer Systems*, vol. 164, p. 107555, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167739X24005193>
- [6] O. E. Marai, T. Taleb, and J. Song, "Roads infrastructure digital twin: A step toward smarter cities realization," *IEEE Network*, vol. 35, no. 2, pp. 136–143, 2021.
- [7] Eclipse Foundation. (n.d.) Eclipse SUMO - simulation of urban mobility. Accessed: Mar. 28, 2025. [Online]. Available: <https://eclipse.dev/sumo/>
- [8] P. Sharma, R. Kumar, M. Gupta *et al.*, "A critical analysis of road network extraction using remote sensing images with deep learning,"



Fig. 4. Road Feature Predictions - First Batch.

- Spatial Information Research*, vol. 32, pp. 485–495, 2024. [Online]. Available: <https://doi.org/10.1007/s41324-024-00576-y>
- [9] L. Cheng, X. Zhang, and J. Shen, "Road surface condition classification using deep learning," *Journal of Visual Communication and Image Representation*, vol. 64, p. 102638, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1047320319302597>
- [10] M. Ahmed Hamza, H. Alqahtani, D. H. Elkamchouchi, H. Alshahrani, J. S. Alzahrani, M. Maray, M. Ahmed Elfaki, and A. S. A. Aziz, "Hyperparameter tuned deep autoencoder model for road classification model in intelligent transportation systems," *Applied Sciences*, vol. 12, no. 20, 2022. [Online]. Available: <https://www.mdpi.com/2076-3417/12/20/10605>
- [11] L. Zhang and C. Long, "Road network representation learning: A dual graph-based approach," *ACM Trans. Knowl. Discov. Data*, vol. 17, no. 9, Jun. 2023. [Online]. Available: <https://doi.org/10.1145/3592859>
- [12] yolo, "sidewalk dataset," <https://universe.roboflow.com/yolos6mwf/sidewalk-6imhx>, jul 2024, visited on 2025-04-23. [Online]. Available: <https://universe.roboflow.com/yolo-s6mwf/sidewalk-6imhx>