Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 5 Issue 4 (2025)

Bridging the Adaptability Gap in Machine Learning Driven Road Safety Systems: A Research Roadmap Toward Safer Streets

Mr. Rohit Kapoor,

Asst. Professor, Department of Computer Science, Lucknow Public College of Professional Studies,

Email Id: r_kapur1982@yahoo.co.in

Dr Sanjeev Verma,

Assistant Professor, Institute of Management Sciences (IMS), University of Lucknow,

Email Id: verma_sanjeev@lkouniv.ac.in

Mr. Nilesh Khare,

Assistant Professor, School of Computer Applications, BBD University Lucknow

Email Id: nileshkhare81@gmail.com

Abstract

ML-based road safety systems also have the potential of reducing traffic accidents and improving pedestrian safety. However, widespread adoption is a major barrier: rigidity. Urban environments are complex and volatile, with changing traffic volume patterns, evolving infrastructure, diverse weather and unpredictable human activities. Modern machine learning models sometimes have difficulty generalising across spatiotemporal variations, resulting in reduced performance when applied outside of controlled or static testing scenarios. In this work we first introduce the adaptation gap that quantifies the discrepancy of a model's performance in training and its robustness to operational variability in the real world. We are aiming for an ambitious agenda of research filling the gap by advancing generalized models, supporting continual learning and managing context, as well as incorporating human-in-the-loop approaches. The paper puts forward a multilayer perspective for future responsive, resilient and just road safety. It accomplishes that by combining the latest developments in machine learning, computer vision, edge computing and transportation systems engineering. We conclude with ethical considerations and policy implications, advocating interdisciplinary collaboration for improving the safety of urban mobility.

Keywords—Machine learning, road safety, adaptability, continual learning, generalization, computer vision, intelligent transportation systems, model robustness

I. Introduction

Road traffic injuries remain an important cause of death throughout the world, with an estimated 1.19 million deaths annually by the World Health Organization (WHO). Fast urbanisation and increasing number of vehicles have led to increased need for advanced road safety technologies. Machine learning (ML), especially deep learning models as those used in computer vision and sensor fusion, have been increasingly becoming an important element in predictive analytics, collision avoidance, real-time hazard identification and related tasks.

While positive results have been reported under laboratory and controlled pilot conditions, many machine learning based systems face barriers in fitting the real-world scenarios due to a lack of robust performance. A model trained to work in the city with good weather may not be able to detect pedestrians in extreme environments, such as at night or during poor weather. Similarly, changes in the environment - including temporary work zones or recent installation of traffic signs — may not be considered on static models. This vulnerability, characterized as the adaption gap, undermines the dependability and the public trust of intelligent transportation systems.

The adaptation gap is no longer a simple byproduct of these technologies but rather a rich problem, accentuating the discrepancy between training data distributions and real operating conditions. Here, we lay out this adaptability gap, its origins and a research agenda to develop ML systems that can operate robustly, adaptively and equitably across heterogeneous urban environments.

II. The Adaptability Gap: Definition and Manifestations

The gap in adaptability is defined as follows:-

1) **Variability of Environment:** It includes variation in illumination, meteorology, occlusion and sensor degradation (e.g. fading) that have profound influence on input data quality. CNNs are trained using daytime images and thus usually mis-classify objects under foggy or night darkness conditions [3].

2) **Behavioural Dynamics:** Road users present high degree of variability in their behaviour. There are local variations in jaywalking behaviour, cycling studies, and driver rage. Models developed in one geographical area may not be generalizable to another/geographical area [4].

3) **Evolution of infrastructure:** the infrastructure in a city evolves over time, caused by construction works, temporary signposts or event management. Instead, we can expect the dog to learn the smell because their robot will be trained to do so and statically deployed until re-trained or changed.

4) **Lack of Data in Edge Cases:** Rare, and high-cost events, such as toddlers running across the street or sudden vehicle evasive [19], are poorly captured in training data with model ignorance at important time [5].

5) **Temporal Drift:** Models trained on historical data are no longer reliable if traffic patterns shift. Seasonal variation in traffic demand, expansion of shared mobility systems, and changes in the urban infrastructure contribute to a degradation of model accuracy over time.

These problems are exacerbated in edge-computing scenarios, where the limited computing resources limit model complexity and update rate.

III. Root Causes of the Adaptability Gap

In order to develop effective treatments, it is important to look at the root reasons of the adaptation gap from both a technical and a systemic point of view.

a) Overfitting to Training Environments

The vast majority of machine learning models have been trained on datasets that were collected under certain conditions (for example, Cityscapes [6] and KITTI [7]), that may not accurately represent global variety. These datasets frequently have too many examples of certain cities (such European urban hubs), types of vehicles, and weather conditions. Because of this, models acquire features that are only useful in certain situations, which makes them perform poorly when they are out of distribution (OOD) [8].

b) The Static Training Paradigm

The lifespan of a traditional ML system is train-deploy-maintain. Once models are put into use, they are rarely changed unless there is a big decline in performance. This is different from how roads are always changing, with new dangers popping up all the time.

c) Not being able to think about things in context

A lot of the time, current systems treat inputs as separate observations instead than as part of a bigger picture. For instance, pedestrian recognition models might not be able to figure out what someone wants depending on the state of the traffic lights, the availability of pavements, or the paths of vehicles. These are all important pieces of information for making correct predictions [9].

d) Few Feedback Loops

ML models don't learn from their mistakes in actual time like people do. Without closed-loop feedback systems (such retraining on false positives and negatives), it's hard to keep becoming better.

e) Different types of sensors and data fusion

Sensor fusion (e.g., LiDAR, cameras, radar) makes systems more reliable, but problems with quality of data, calibration drift make systems less reliable, especially in bad weather [10].

IV. A Research Roadmap to Bridge the Adaptability Gap

A. Improving the Generalisability and Robustness of the Model

i) Generalisation Across Domains and Invariance

Through domain generalisation (DG) methods, models can be trained to achieve good performance on unseen domains by utilising the data from the target domain itself. You can make things more robust by methods such as adversarial training [13], meta-learning [12] and domain-invariant feature learning [11]. For example, in pedestrian recognition work of Javanmardi et al. [14] demonstrated that DANNs reduced the variance between cities by a factor of 23%.

ii) Adding Realistic Perturbations to Data

Multiple synthetic training scenarios such as different weather, rain, fog and snow can be created from simulators like CARLA [15] and SUMO [16]. But there remains a gap between simulation and reality. Such a gap could be addressed by employing recent advances in photorealistic rendering and domain randomisation [17]. In addition, a hybrid method is employed by mixing synthetic data with real-world perturbation -enhanced data to increase flexibility.

iii) Predictions with Concern for Uncertainty

Estimates of uncertainty are yielded by Bayesian neural networks (BNNs) [18] and Monte Carlo dropout, which computers can then use to determine when inputs do not adhere to learnt distributions. So with this sort of information, backup strategies can be formed like alerting a human operator or reducing the level of autonomy.

B. **Enabling Continual and Lifelong Learning**

1. Incremental Learning without Catastrophic Forgetting

Traditional retraining from scratch is prohibitively expensive for edge devices. New pattern learning and memory of old patterns are supported by the Cl approach, such as EWC [19], replay-based buffers [20] and generative replay [21].

Recent research by Lee et al. [22] introduced a CL based object detector that can assimilate new traffic devices (e.g., construction warnings) via small memory buffer, reducing the update latency by 85% relative to full retraining.

2. Federated and Edge Learning Frameworks

Federated learning (FL) enables the distributed updates of models at the level of sensors or vehicles in various geographical locations by avoiding transmission of raw data. This allows the user to in privacy change local conditions. In order to improve road safety, each traffic camera and/or autonomous car reports its updates to the global model as well as locally adjusting a local version of the model [23].

3. Event-Based Changes to the Model

Systems need to be event-driven (as opposed to being rolled out as planned updates) such as observing an increase in false negatives over time, or evidence of a change in the environment like first snowfall. These can start light-weight model fine-tuning or data collection efforts.

C. Integrating Context and Cognition

1. Multimodal Contextual Modelling

Models need to capture static as well as dynamic contexts. Static factors include urban design properties available in the road, and dynamic variables consists of traffic signal patterns, weather conditions and congestion perception at that time. Relationships between traffic elements and their environmental context could be modelled using graph neural networks (GNNs) [24].

2. Temporal Reasoning and Intent Prediction

Sequential models, such as Transformers and RNNs can predict intentions of traffic participants. A car that suddenly slows while approaching a school could mean there's a pedestrian walking, which can prompt action.

3. Knowledge Graphs for Situational Awareness

Integrating such structured area awareness information into knowledge graphs can guide model predictions. Hybrid neuro-symbolic models [26] provide a means to introduce reasoning and learning where, for example, learnt patterns are combined with explicit rules to improve interpretability and safety.

D. Human-in-the-Loop and Trust Calibration

a) Adaptive Interfaces for Human Supervision

Trust in machine learning systems should be adaptive and shaped by human operators when uncertainty is high or during novel circumstances. Active learning algorithms allow computer systems to select ambiguous samples for labeling and enhance a training set iteratively [27].

b) Clarification for Stakeholder Assurance

(XAI) tools, such as Grad-CAM [28] or SHAP values [29], can visualize model attention, and help both engineers and regulators to comprehend the reason of decisions. Such transparent systems build trust and are conducive for public acceptance and regulatory adherence.

c) Equity and Bias Remediation

Adaptability must not intensify unfairness. Such models may perform poorly for marginalized populations like people with disabilities and people wearing non-standard clothing. Assessing datasets for diversity and using fairness-aware learning are both important to achieve inclusive safety.

V. System Architecture for Adaptive Road Safety

A stratified system design may be employed to implement the research pathway

a) Perception Layer

• Integration of multiple sensors (cameras, LiDAR, radar)

• Real time based anomaly detection (e.g., sensor fault, blockage)

b) Adaptive Inference Layer
- Uncertainty measure and value opinioneer ensemble methods
- Gradual updating with CL-enabled modules

• Contextual Forecast in Graph Neural Networks and Temporal models

c) Learning and Update Framework

• A federated learning orchestrator

• Event-driven update scheduler

• A robust data pipeline for retraining

d) Cognition and Logic Layer
• Traffic regulations repository

• Repository of incident history

• Neuro-symbolic inference mechanism

e) Layer of Humans Interaction XAI dashboards for traffic management officers

• Mobile alerts for vulnerable road users (e.g., in the form of smartphone apps)

• Feedback for public reporting

This architecture allows its deployment at centralised (e.g., city level traffic management) and decentralised (e.g., vehicle-level) levels and is hence scalable.

VI. Challenges and Ethical Considerations

Despite advancements in technology, there are still a number of challenges to overcome:

- **Privacy**: Continuous surveillance for training data raises privacy concerns. Federated learning and edge processing mitigate but do not eliminate risks.
- **Security**: Adaptive systems may be vulnerable to adversarial attacks targeting model updates [31].
- **Regulatory Compliance**: Automated systems must comply with traffic laws and liability frameworks, which vary across jurisdictions.
- **Digital Divide**: High-tech solutions may only be deployed in affluent cities, widening safety disparities.

Ethical deployment requires participatory design, involving communities in system development and prioritizing equity in safety outcomes.

VII. Policy and Implementation Recommendations

In order to accelerate the application of flexible road safety interventions, the following process could be adopted.

i) Baseline Benchmarking: Establish general benchmarks to evaluate model transferability and usefulness in different contexts (e.g., climate, population, infrastructure).
ii) Data Sharing: An open, anonymised evidential information exchange of gathered data by municipalities and manufacturers supported by fair use legislations.
iii) Public-private partnerships: Unite cities, universities and private firms to prototype and deliver adaptive systems.
iv) Regulatory Sandboxes: Established under a regulatory framework, they enable testing and targeted learning in live traffic under conducive government supervision.

v) Global Safety Metrics: Standardization of metrics for the assessment in support with the WHO road safety targets, focusing on fatality reduction and vulnerable road user protection.

VIII. Conclusion

The adoption gap is the main obstacle for ML-based road security solutions. While there is optimism in next generation technology, environmental conditions and human behaviour, and infrastructure all conspire to make these systems less effective currently than they might at first glance seem. This difference calls for the shifting from the existing, "static" and non-connected paradigms to more extended, adaptive and self-improving solutions.

The article describes a complete process for studying the future train of generalisation, lifelong learning, common sense inference and human communication. Smart, flexible, equitable and trusted systems of safety can be ours if we aim for it.

Scalable interventions, long-term evaluations, and all-around policies should be the focus areas of future interventions. Only by cross-discipline coalitions can help to achieve safer streets for everyone.

References

[1] World Health Organization, "Global status report on road safety 2023," WHO, Geneva, 2023.

[2] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3354–3361.

[3] S. S. Salehi, H. Rajagopalan, and I. G. Sluiter, "Impact of weather conditions on deep learning-based object detection in autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3677–3687, 2021.

[4] Y. Liu et al., "Cross-dataset generalization in urban scene understanding: A benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 1521–1536, 2023.

[5] D. Amodei et al., "Concrete problems in AI safety," *arXiv preprint arXiv:1606.06565*, 2016.

[6] M. Cordts et al., "The Cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3213–3223.

[7] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[8] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Adv. Neural Inf. Process. Syst.*, 2017, pp. 6402–6413.

[9] N. Bhattacharyya et al., "Multi-modal trajectory prediction for autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 1, pp. 554–565, 2023.

[10] Y. Xiang, B. Choi, and D. Held, "MVA3D: Multiview attention for 3D object detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 11 092–11 099.

[11] K. Muandet, D. Balduzzi, and B. Schölkopf, "Domain generalization using causal inference," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 2421–2460, 2013.

[12] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1126–1135.

[13] Y. Ganin et al., "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.

[14] M. Javanmardi and T. Moon, "Domain adaptation for pedestrian detection: An application of adversarial networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1486–1495, 2020.

[15] A. Dosovitskiy et al., "CARLA: An open urban driving simulator," in *Proc. 1st Annu. Conf. Robot Learn.*, 2017, pp. 1–16.

[16] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO – Simulation of Urban MObility," *Int. J. Adv. Syst. Meas.*, vol. 5, no. 3&4, pp. 128–138, 2012.

[17] J. Tremblay et al., "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 969–977.

[18] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.

[19] E. Parisi, R. Kemker, J. L. Part, C. Kanan, and C. W. Bieser, "Continual lifelong learning with neural networks: A review," *Neural Netw.*, vol. 113, pp. 54–71, 2019.

[20] D. Lopez-Paz and M. A. Ranzato, "Gradient episodic memory for continual learning," in *Adv. Neural Inf. Process. Syst.*, 2017, pp. 6467–6476.

[21] S. Lee et al., "Generative adversarial replay for continual learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10 89–10 98.

[22] H. Lee, J. Park, and S. Kim, "Continual learning for real-time traffic sign recognition in smart cities," *IEEE Trans. Intell. Transp. Syst.*, early access, 2023, doi: 10.1109/TITS.2023.3287552.

[23] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.

[24] Y. Li et al., "Spatial-temporal graph convolutional networks for skeleton-based action recognition," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 3562–3569.

[25] A. Vaswani et al., "Attention is all you need," in *Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[26] A. d'Avila Garcez et al., "Neurosymbolic AI: The 3rd wave," *AI Mag.*, vol. 43, no. 2, pp. 51–66, 2022.

[27] B. Settles, "Active learning literature survey," *Univ. Wisconsin–Madison Tech. Rep.*, 1648, 2009.

[28] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, pp. 336–359, 2020.

[29] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neural Inf. Process. Syst.*, 2017, pp. 4765–4774.

[30] M. B. Zafar, I. Valera, M. G. Rodriguez, and K. P. Gummadi, "Fairness constraints: Mechanisms for fair classification," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2017, pp. 962–970.

[31] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in *Proc. IEEE Symp. Secur. Privacy*, 2017, pp. 39–57.