TOWARDS ADAPTIVE TRAFFIC SIGNAL CONTROL THROUGH FOUNDATION MODELS AND REINFORCEMENT LEARNING

Lukas Klein, Arthur Müller, Magnus Redeker
Fraunhofer IOSB, IOSB-INA Lemgo
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation
lukas.klein@iosb-ina.fraunhofer.de

ABSTRACT

Traffic Signal Control (TSC) is pivotal for managing urban traffic flow and enhancing intersection safety. Traditional TSC systems are rule-based and tailored to specific intersections, requiring substantial training and resources, which restricts their flexibility. This paper proposes a novel adaptive, scalable solution utilizing Foundation Models (FM) and Reinforcement Learning (RL), designed to handle diverse urban intersections efficiently without extensive retraining. The approach leverages advanced neural network architectures, including attention mechanisms, to improve generalization capabilities across different intersection topologies. A safety control mechanism aligned with traffic regulations ensures the safe operation of traffic signals, significantly enhancing the system's reliability. By systematically classifying intersection types, the method tailors the control strategies to specific traffic scenarios, further reducing implementation times and expertise requirements. This FM- and RL-based approach not only reduces resource demands but also promises more efficient traffic flow and improved safety in various urban settings.

Keywords Traffic Signal Control · Foundation Models · Reinforcement Learning · Rapid Application in Heterogeneous Environments · Safe and Adaptive Traffic Flow Optimization

1 Introduction

The optimization of traffic flow is essential for ensuring the efficient utilization of urban road networks, reducing congestion, minimizing travel delays, and improving overall traffic system performance. Traffic signal control (TSC) systems face several challenges, including the integration of modes of transport, which involves considering the needs of pedestrians, cyclists, cars and truck drivers, and public transport to create a comprehensive traffic management system. The escalating traffic volumes and resulting congestion significantly extend waiting times for commuters and the logistics industry, thereby escalating operational costs and reducing overall efficiency. Safety aspects are crucial, focusing on ensuring the safety of both pedestrians and drivers to minimize the risk of accidents at intersections.

Reinforcement learning (RL) presents a promising AI-driven approach for optimizing TSC by leveraging vast amounts of real-time data. Advances in traffic sensing technologies, such as high-resolution cameras, radar sensors with a high range and accuracy, and the associated access to high-quality data provided by vehicle-to-everything (V2X) technologies, enable data at scale, facilitating the development of more adaptive and efficient RL-based TSC.

RL and other Artificial Intelligence (AI)-based techniques for TSC have already been extensively studied [1, 2, 3] and have even been evaluated in real-world settings, including Cologne,

Essenbach, Hamm, Ingolstadt, and Lemgo in Germany [4, 5]. However, challenges remain, particularly in terms of generalizability since most existing approaches require specific training of RL-agents for each individual intersection, leading to high computational costs, long implementation times, and a demand for specialized AI and traffic engineering expertise — resources that many municipalities, traffic engineering firms, and research institutions lack.

In order for AI-based TSC optimization to gain widespread adaptation, it must not only match or exceed the performance of traditional rule-based methods but also reduce implementation complexity and expertise requirements. To address these challenges, this paper introduces a Foundation Model (FM)-based approach that enhances adaptability, scalability, and efficiency. Key aspects of this approach include *topology-agnostic adaptability*, enabling the model to manage diverse intersections types without extensive retraining, and the creation of a *robust training dataset* that integrates both synthetic and real-world traffic data. By leveraging these advancements, the proposed method aims to provide a more flexible, cost-effective, and broadly applicable solution for modern traffic management.

Beyond optimizing traffic flow, ensuring safety remains a crucial aspect of AI-driven TSC. Implementing safety control mechanisms that comply with traffic regulations is necessary to guarantee the safe operation of traffic signals and prevent hazardous situations at intersections. Furthermore, a *systematic classification of intersection types* based on relevant characteristics will facilitate a deeper understanding and modeling of diverse traffic scenarios, ultimately contributing to more effective traffic management solutions. Additionally, leveraging *advanced neural network architectures*, such as attention mechanisms, can enhance the performance and generalization capabilities of the RL-based system, allowing it to better adapt to new and unseen traffic conditions and intersection topologies.

By addressing these aspects, AI-driven TSC solutions will become more efficient, adaptable, and accessible, ultimately improving urban traffic management systems.

This paper is structured as follows: Section 2 provides an overview of the state-of-the-art in RL-based TSC. Section 3 conceptualizes the proposed foundation model approach. Finally, Section 4 discusses the applicability and generalizability of the results and outlines directions for future research and development.

2 Preliminaries

2.1 Traffic Signal Control

Traffic Signal Control (TSC) refers to the control of intersections within urban road networks to manage vehicular and pedestrian movement. Conventional TSC policies, embedded in traffic controllers, are designed by traffic engineers and rely on rule-based mechanisms to govern the operation of traffic signals. These control strategies are dynamic and adjust based on real-time traffic conditions, which are detected through various sensing technologies such as cameras, inductive loop detectors, and pedestrian buttons.

Traffic signals (or *traffic lights*) serve as visual signaling devices that provide guidance to road users by regulating the right-of-way for specific lanes or movements. Multiple traffic signals can be grouped into a *signal group*, which is defined as a set of traffic signals that control specific traffic flows simultaneously. This grouping ensures synchronized signaling for particular directions or lanes, optimizing traffic flow and enhancing intersection efficiency.

Traffic phases represent a specific combination of traffic signal states that define a stable traffic light configuration, during which the signal status remains unchanged. Each phase typically corresponds to a set of compatible traffic flows, such as vehicles traveling in the same or non-conflicting directions, ensuring smooth and coordinated movement through an intersection.

Phase transitions occur when the Traffic Light System (TLS) control mechanism initiates a shift from one traffic phase to another to regulate traffic flow effectively. These transitions follow a structured sequence, beginning with the traffic light switching from green to yellow and subsequently to red. Once a phase concludes, the corresponding signal groups of the next phase are activated by switching to green, ensuring a seamless and orderly traffic flow.

Phase-based control in traffic signal systems manages predefined combinations of signal groups as unified entities. These transitions between these phases are predetermined in control protocols and are specific to each intersection. Typically, only a subset of possible phase transitions is permitted to maintain a conflict-free and efficient operation. These transitions are often illustrated in *phase transition diagrams*, which outline the sequence of phase changes to uphold traffic safety. In contrast, *signal group-based control* operates individual signal groups independently, adjusting their timing separately while ensuring conflict-free traffic movement across the intersection.

In reinforcement learning (RL)-based TSC, it is essential to ensure that agents execute phase and signal group transitions in a manner that adheres to traffic safety regulations. Müller, Rangras, Ferfers, Hufen, Schreckenberg, Jasperneite, Schnittker, Waldmann, Friesen, and Wiering [5] propose a safety layer, a mechanism designed to evaluate and regulate the switching decisions made by RL agents. This layer allows phase transitions only if they meet predefined safety criteria, rejecting unsafe transitions through *action masking* to prevent violations of traffic regulations. The integration of such safety mechanisms can be implemented in various ways; therefore, the terms *safety layer* and *safety control system* are used to refer to any approach that enforces safety constraints within RL-based TSC.

2.2 Reinforcement Learning

Reinforcement learning (RL) agents interact with their environment with the objective of making optimal decisions and executing corresponding actions. At each time step t, an agent observes the current state of the environment, denoted as s_t. Based on this information, the agent selects the subsequent action a_t, to undertake within the environment. The agent's behavior is governed by a policy π, which defines a mapping from states to actions, expressed as $\pi(a_t|s_t)$. To assess the agent's performance in an RL framework, the environment provides a reward signal r_{t+1} which quantifies the effectiveness of the chosen action in achieving the specified objective. The fundamental goal of an RL agent is to maximize the expected cumulative reward over time.

This decision-making process is typically formalized using *Markov Decision Processes* (MDPs), a mathematical approach that simplifies sequential decision-making into a structured format. An MDP is defined as a tuple comprising the following elements [6]:

- **State space** S: A set of states representing the environment's current condition.
- **Action space** A: A set of possible actions from which the agent selects one at each time step t.
- **State transition function** $T(s_t, a_t, s_{t+1})$: A probability distribution over the successor state s_{t+1}, given the current state s_t and action a_t.
- **Reward function** $R(s_t, a_t)$: A function that assigns a reward r_{t+1} to the agent based on the action a_t taken in state s_t.
- **Discount factor** $\gamma \in [0, 1]$: A parameter that determines the importance of future rewards, with γ close to zero prioritizing immediate rewards.

A key property of an MDP is the *Markov property*, which states that the current state s_t contains all relevant information about the past states and actions, meaning that the successor state s_{t+1} depends solely on the current state s_t and the action a_t, independent of prior states and actions. MDPs can model both *continuous tasks*, which have no predefined endpoint, and *episodic tasks*, which terminate upon reaching a specific state. Consequently, MDPs can address a diverse array of decision-making scenarios within the RL domain.

2.3 Foundation Models

Foundation Models (FMs) are large-scale, pre-trained neural networks designed to serve as adaptable platforms for a wide range of downstream tasks across various domains, including natural language processing (NLP), computer vision, and beyond. Prominent examples of FMs include models such as BERT [7], GPT [8], and Llama [9], which utilize vast amounts of data and advanced architectures to learn complex patterns and relationships inherent in the data.

The primary advantage of FMs is their ability to generalize across diverse tasks, significantly reducing the need for task-specific training. This generalization is facilitated through *transfer learning*, a paradigm in which models pre-trained on large and varied datasets are fine-tuned for specific applications [10]. FMs are typically built on the *transformer architecture* [11], which employs *self-attention mechanisms* to assign different weights to input tokens of varying lengths, thereby enabling the model to capture contextual relationships more effectively.

While transformer-based models generally excel at processing sequential data, it is important to note that, in the context of TSC, token sequences are not directly processed. Instead, the attention mechanism remains highly beneficial as it assists the model in identifying which elements within the input space correlate with specific traffic phases. Essentially, it calculates the relevance of various inputs (such as traffic flow data or sensor readings) to different phases of signal control, allowing the model to focus on the most pertinent information for each phase transition. Additionally, attention mechanisms facilitate massive parallelization during training, enabling FMs with transformer architecture to converge faster and overall improve their performance on large datasets.

2.4 Related Work

Several promising results are evident from research activities in TSC optimization:

- Ault and Sharon [12] present a toolkit for the development and comparison of various RL-based TSC systems and the benchmarking of scenarios based on realistic traffic conditions from the German cities of Cologne and Ingolstadt. While this work offers valuable insights, it primarily focuses on the training of only two.
- Zhao, Dai, Chen, Lin, Lv, and Wang [13] address the Sim2Real challenge by employing descriptive learning with a Wasserstein GAN to generate different traffic scenarios, utilizing transformer-based learning to discern traffic dynamics, and fine-tuning their FM through real-world applications. However, not addressed are structural and spatial features of intersections, which could provide valuable information for RL learning.
- Zang, Yao, Zheng, Xu, Xu, and Li [14] develop a metamodel that requires adaptation to specific traffic signals. However, it is not structure-agnostic, as it is specifically designed for 4-arm intersection topologies only. The approach is based on basic model-agnostic meta-learning (MAML), instead of improved and more robust later variants [15, 16]. Additionally, an outdated convolutional neural network (CNN) architecture is employed and not yet the current de facto standard of attention mechanisms in FMs.
- Zhang, Liu, Zhang, Zheng, and Yu [17] seek to enhance the generalization of TSC by training on diverse traffic flow data generated via a Wasserstein generative adversarial network (WGAN). Their approach integrates flow clustering with model-agnostic meta-learning to effectively adapt to various traffic conditions, demonstrating good performance across different traffic flows. However, the topology of the simulation models remains constant, which may restrict the FMs ability to generalize across topologically different intersections.
- Oroojlooy, Nazari, Hajinezhad, and Silva [18] propose a universal attention-based TSC-model that utilizes attention mechanisms to generalize across various intersection topologies. However, it lacks a dedicated safety mechanism to ensure only permissible actions are taken, which is crucial for a future deployment in real-world scenarios.

On the bottom line, research should focus on generalization capabilities across intersection types topology-agnostic TSC adaptability applying advanced neural network architectures, targeting straightforward effort-less and cost-effective TSC optimization.

3 Towards Adaptive Traffic Signal Control through Foundation Models and Reinforcement Learning

This section outlines the methodology for developing an FM for RL-based TSC, focusing on the integration of synthetic and real-world data to create a robust and adaptable dataset as depicted in Figure 1. By leveraging advanced techniques such as attention mechanisms and

systematic classification of intersection types, the approach aims to facilitate dynamic traffic signal management across diverse urban environments. The outlined framework not only targets the effective training of RL agents but also ensures compliance with essential safety regulations, ultimately contributing to more efficient, adaptive urban traffic systems.

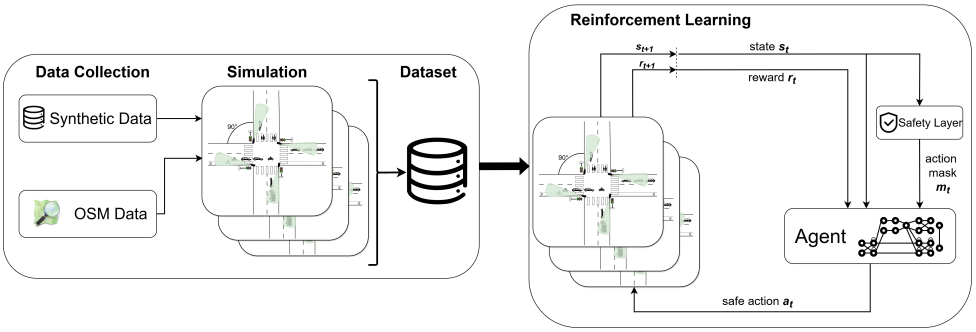

Figure 1: High-Level illustration of the data collection process and reinforcement learning workflow for training a Foundation Model with action masking. OpenStreetMap-Logo: © OpenStreetMap contributors, CC-BY-SA 2.0.

3.1 Core Conception Principals

The following subsections describe the core design principles. These include data collection and preparation, the simulation framework, safety mechanisms, and ensuring agnosticism towards intersection and environment topology.

3.1.1 Data Collection and Preparation

In order to create a diversified dataset, simulation models of individual intersections are created using both synthetically generated networks and real-world data from OpenStreetMap (OSM, [19]). The generation of synthetic traffic networks enables the mapping of a variety of traffic scenarios and conditions, thereby providing a valuable basis for training an FM. The generation of networks based on real intersection situations from OSM enriches the dataset with real-world realistic data. The combination of synthetic and real data facilitates the generalization of the developed FM.

Furthermore, the dataset is augmented with multiple phase diagrams for each specific intersection, enabling the mapping of diverse switching logics for the same topology. These phase diagrams are generated using a tool to be developed from the respective network topology and traffic engineering expert knowledge in order to automatically generate a safety control system. Such a safety control system represents a pivotal component in the outlined methodology, and is instrumental in training of the FM to comply with the mandatory safety regulations governing the switching of traffic lights in the field.

To ensure that the dataset can be employed as a benchmark for the training and evaluation of RL for TSC, it is essential to classify different isolated intersection types based on relevant characteristics such as topology, participant groups (e.g., motorized and non-motorized traffic), public transport, traffic patterns (e.g., partial load and full load), and speed limits. The objective is to develop a systematic description of the various intersection types, which will aid in understanding of intersection design and serve as a foundation for the synthetic generation of simulation models for the dataset. To align the dataset with the actual distribution of intersections in the real world, a minimum number of intersections per type will be selected for inclusion.

3.1.2 Simulation Framework

A systematic classification of intersection types will be undertaken based on their topological characteristics as described in OpenStreetMap and informed by expert knowledge. Subsequently, the OSM Web Wizard, a tool developed by the *Deutsches Zentrum für Luft- und Raumfahrt*

(DLR), will be employed to generate a simulation scenario in the microscopic traffic simulation tool SUMO based on a selection of an OSM map section [20, 21] as in Figure 2. This will be followed by a post-processing phase with the objective of refining the resulting scenarios. The synthetic simulation models will then be generated to correspond with the identified intersection types, covering a range of traffic volumes, specifically full-load, partial-load and light-load, with calibration aimed at full-load conditions.

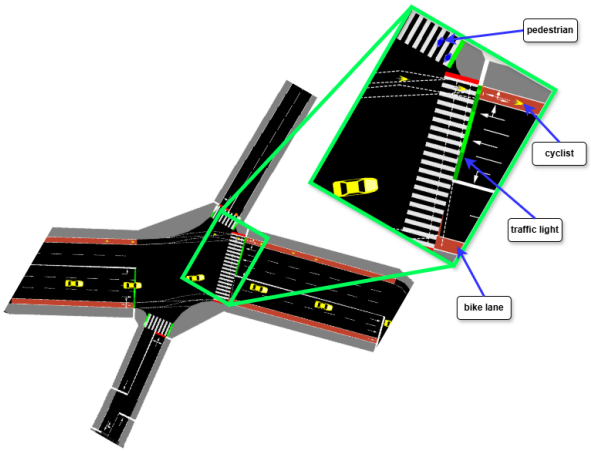

Figure 2: Illustration of an intersection simulated in SUMO, highlighting key traffic elements such as pedestrians, cyclists, bike lanes, and traffic lights.

In addition, different groups of road users and public transport modalities will be defined for each intersection to ensure a comprehensive and representative simulation framework. Each simulation model per intersection will also be equipped with a safety control mechanism, referred to as the *safety layer*, which will adhere strictly to the most essential guidelines for the design and operation of traffic signal systems in Germany [22].

To enhance efficiency, the interim time matrices and corresponding blocking times are automatically generated. As the interim times depend on the number of traffic streams at an intersection and the lengths of the entry and exit paths, it is possible to automate the calculation of these interim times based on the intersection's topology. The interim time matrix provides a representation of all interim times at a given intersection, facilitating comparison and reading across all traffic streams. In the domain of traffic engineering, the term *interim time* refers to the duration that must elapse between the end of the green time of one traffic stream and the start of the green phase of another, partially or fully conflicting, incoming traffic stream and are a crucial safety-related aspect, as they ensure that incoming and outgoing traffic flows can be switched without conflict, thus preventing potential collision scenarios.

3.1.3 Safety Mechanisms

The safety layer will be implemented natively in Python, eliminating the need for additional traffic engineering workstations such as *LISA*, [23]. This will ensure accessibility to a wider range of user groups while improving overall performance.

3.1.4 Intersection and Environment Topology Agnosticity

Previous studies on reinforcement learning for traffic signal control have primarily focused on phase-based switching, where RL agents operate within predefined traffic phases and phase transition plans. While this approach has demonstrated feasibility, it limits the system's adaptability to dynamic traffic conditions. To enable a more flexible and responsive control strategy, ongoing research aims to develop a Foundation Model (FM) that supports both phase-based and signal group-based switching, allowing for more granular and adaptive traffic signal management.

In order to achieve this, specific phase diagrams must be automatically generated for each intersection. A comprehensive topological analysis must be conducted, extracting key structural characteristics such as the number of arms, available lanes, and prevailing traffic flows. This analysis enables the modeling of traffic movement and the determination of the minimum number of required phases. By systematically analyzing potential conflicts between these phases, a basic phase sequence diagram can be generated while ensuring compliance with essential safety regulations.

For signal group-based switching, a safety control layer is required to prevent RL agents from activating conflicting signal groups simultaneously. This layer ensures that only non-conflicting traffic movements are permitted, thereby maintaining safe and efficient traffic operations.

To integrate RL-based control within a simulation environment, an environment class is designed to facilitate seamless interaction between RL agents and the traffic simulation. Built upon the Gymnasium framework [24], this implementation allows for the creation of environment instances that reflect specific intersection topologies, traffic volumes, public transportation modalities, and user groups. Users can define scenario parameters, enabling the automated instantiation of tailored environments from a pre-existing dataset. Each unique parameter combination results in a corresponding environment instance that accurately represents the defined traffic scenario.

Furthermore, the environment class provides a standardized interface for interacting with simulation components while incorporating the safety control layer to enhance reliability and ensure regulatory compliance. This design promotes modularity, making it easier to integrate RL agents with diverse traffic environments and facilitating the extensibility of RL frameworks, such as Ray RLlib. By enabling structured and safe RL training, this approach supports the development of scalable, real-world-applicable AI-driven traffic signal control systems.

3.2 Leveraging Reinforcement Learning and Foundation Models in the Concept of Adaptive Traffic Signal Control

To enable low-effort, cost-effective, and adaptive TSC, the development of intersection topology-agnostic neural network architectures is a critical focus of research. A topology-agnostic approach ensures that the same model architecture can flexibly adapt to different intersection types without requiring structural modifications or retraining for each new scenario. This adaptability extends not only to dynamic input representations that reflect the unique characteristics of each intersection but also to variable output structures, corresponding to the number of signal groups or predefined traffic phases at a given location.

To achieve this level of generalizability, advanced techniques such as attention mechanisms and Graph Neural Networks (GNNs) play a crucial role in RL-based TSC. GNNs are particularly well-suited for modeling traffic networks, where lanes, traffic movements, and intersections can be represented as nodes, and their relationships—such as connectivity, priority rules, and turning restrictions—can be represented as edges. Unlike traditional neural networks, GNNs allow flexible representations of varying traffic intersection topologies, making them an ideal choice for RL-based adaptive signal control.

Additionally, attention mechanisms further enhance the adaptability of RL-based TSC by dynamically weighting the relevance of different inputs within the graph. This is particularly useful in variable-sized inputs, where the model must determine which traffic flows, lanes, or intersections should be given priority based on real-time conditions. Unlike traditional feedforward architectures, which treat all input features equally, attention mechanisms enable the model to focus on the most relevant traffic features at any given moment, improving decision-making robustness.

To ensure effective integration into the RL framework, the design of these neural architectures must adhere to Markov Decision Process (MDP) principles, ensuring a well-structured formulation of states, observations, actions, and rewards. A key component of this structure is the incorporation of historical traffic data, allowing RL agents to learn from past traffic patterns rather than relying solely on instantaneous states. By recognizing temporal dependencies, agents can anticipate congestion patterns and peak-hour trends, enabling more proactive and adaptive

traffic signal control. By combining GNN-based representations with attention mechanisms, RL policies can flexibly handle input and output variations, enhancing the scalability and efficiency of AI-driven TSC Foundation Models.

3.2.1 State space

In RL, the state space is defined as a set encompassing all potential states of the environment. At each time step t during the RL training process, a state s_t is provided as the input to the training framework's policy. Generally, this state represented as a vector containing information about the environment's current state, such as the current phase of the traffic lights at the intersection or the queue length on each lane.

To facilitate the learning of a foundation model with the greatest possible breadth of information, it is essential to construct the state space meticulously to ensure its generalizability across different intersection topologies. The agent requires an awareness of the current signal phase and the average waiting time of road users at each incoming lane to accurately assess the prevailing traffic conditions and make optimal decisions to improve traffic flow. Additionally, to enhance generalization capabilities, it is essential to differentiate between varying traffic load levels (e.g., full load, partial load) and incorporate temporal information, such as the duration of the current signal phase or the average lane occupancy over a predefined time window (e.g., the last ten minutes). Beyond traffic dynamics and temporal information, capturing spatial information is fundamental, ensuring the RL agent learns the topological structure of an intersection. The agent must recognize key attributes of the intersection, such as the number of lanes, the number of arms, and the topological layout (e.g, the angles between lanes). This enables the agent to tailor its decision based on the structural characteristics of the traffic environment, ensuring optimized TSC strategies across varying intersection designs.

For a fixed intersection topology, the state space can be represented as a fixed-length vector, simplifying neural network training by ensuring a uniform input size at each time step t. However, this approach becomes problematic when training an FM across multiple intersections, as input size variations arise from differences in intersection structures. Therefore, it is crucial to develop and utilize an appropriate policy architecture, such as an artificial neural network, that accommodates variable input sizes.

A promising approach to address variable-size inputs is to define a fixed-length vector that applies for each lane at an intersection. This vector encapsulates all the relevant spatial, temporal, and contextual information for that particular lane. In cases where certain lanes provide less information, zero-padding can be applied to maintain a uniform vector size, enabling an information-size-agnostic representation. To further enhance the topology-agnostic nature of the state space, the entire intersection state can be modeled as a graph, where nodes represent lanes and edges capture the logical dependencies between lanes (e.g., connectivity, conflicting movements, or priority rules). This graph-based representation allows the RL agent to adapt dynamically to different intersection layouts while preserving crucial spatial relationships, ultimately improving its ability to generalize and optimize traffic signal control across diverse urban environments.

3.2.2 Action space

Analogous to the state space, the action space is defined as a set of all possible actions that an agent can perform in the environment, serving as the output of the policy during the RL process. When training an RL agent for a specific intersection, the output layer is typically designed with a fixed dimension corresponding to the number of switchable phases or signal groups, thereby defining a discrete action space. The logits—i.e., the probability distribution over possible actions—are then sampled to determine the next action.

When training an FM across different intersections, the output size of the policy varies based on the number of defined phases or signal groups. To achieve output agnosticism, several techniques can be employed. One straightforward approach involves dynamically adjusting the output layer to match the action space size of each intersection during training. However, this approach necessitates fine-tuning for each new intersection. A more scalable and generalizable

approach leverages attention mechanisms. Here, an embedding representation is computed for each lane-specific feature vector, capturing relevant lane-specific information. An attention mechanism then assigns weights to these embeddings, determining the policy's output. A downstream translation mechanism subsequently maps the attention score of an embedding back to their corresponding lanes, associating them with the correct signal group or phase.

Furthermore, the security layer not only verifies the switching intentions of the agents to ensure compliance with legal and safety regulations, but also indicates which phases or signal groups can be switched in the subsequent time step. Following [25], a masking vector is generated from these permitted phases or signal groups, constraining the action space of the agents and enabling only allowed phase or signal group transitions. This action masking technique must be adapted to the network architecture implemented.

3.2.3 Reward function

The formulation of a reward function aimed at optimizing traffic flow is a crucial step in enabling a foundation model to make well-informed TSC decisions. This process is guided by the *Handbuch für die Bemessung von Straßenverkehrsanlagen* (HBS) [26], which provides standardized metrics for the evaluation of traffic flow efficiency. The proposed reward function may serve as a benchmarking metric, leveraging the concept of *Level-of-Service* (LOS) as a foundational element. LOS is a discrete metric that assesses roadway segments and traffic directions based on the average waiting time of road users in seconds. It differentiates between motorized and non-motorized users, enabling the weighting of various groups of traffic participants. The overall LOS of an intersection is determined by the worst-performing traffic direction, ensuring that congestion in any direction is appropriately penalized. By incorporating LOS into the reward function, the FM can be trained to prioritize traffic efficiency while considering the needs of all road users.

A crucial consideration in the design of a suitable reward function involves addressing the temporal constraints of real-world traffic systems. In a naive simulation of a TSC, the agent receives an observation and reward at each time step t and can execute an action at any given moment, resulting in an equidistant interaction between the agent and the environment. However, in real-world traffic systems, actions cannot be executed at arbitrary time steps due to regulatory constraints such as minimum green times and mandatory phase transitions. Consequently, the intervals between successive decisions of RL-agent's are not equidistant, meaning that the environment continues to evolve even when the RL agent cannot take an action. To address this, it is essential to employ an appropriate reward aggregation strategy, ensuring that all rewards accumulated during non-decision periods are incorporated into the learning process rather than being lost. This approach guarantees that the agent accurately evaluates the impact of its actions over time.

Furthermore, to enhance the adaptability of the FM, its architectural framework must support signal group-based control, allowing agents to execute simultaneous actions that modify the states of multiple signal groups concurrently. To validate these concepts, a proof-of-concept prototype will be developed and trained on a subset of the dataset. This functional validation phase is crucial for assessing the model's applicability across diverse intersection topologies, demonstrating its feasibility and effectiveness in real-world urban traffic management. By integrating advanced AI techniques with well-established traffic engineering principles, this approach aims to enhance the efficiency, adaptability, and robustness of next-generation RL-based TSC systems.

4 Conclusion and Outlook

This paper introduced the concept of a Foundation Model (FM) for Reinforcement Learning (RL)-based Traffic Signal Control (TSC), designed to generalize across diverse intersection topologies by integrating attention mechanisms and graph-based representations. By addressing the challenge of variable input and output sizes, the proposed FM eliminates the need for training individual models for each intersection, significantly reducing computational demands and implementation effort. The topology-agnostic approach, supported by techniques such as

Graph Neural Networks and attention mechanisms, ensures that the model can dynamically adapt to different intersection structures, traffic patterns, and regulatory constraints.

A key contribution of this work will be the development of a benchmark dataset, derived from a combination of synthetic and real-world traffic networks, sourced from OpenStreetMap (OSM). This dataset will facilitate the training, evaluation, and comparison of RL-based TSC methods in a standardized manner, enabling municipalities, traffic engineers, and researchers to implement AI-driven traffic management solutions without requiring extensive AI and traffic engineering expertise or substantial computational resources. Furthermore, the integration of traffic simulations allows for realistic scenario modeling, enhancing the applicability and robustness of the proposed approach.

To ensure compliance with traffic regulations and safety standards, the model incorporates a safety control mechanism, extending the work of Müller, Rangras, Ferfers, Hufen, Schreckenberg, Jasperneite, Schnittker, Waldmann, Friesen, and Wiering [5], which enforces constraints on signal group switching and prevents unsafe phase transitions. This regulatory framework ensures that RL-based TSC solutions remain legally and functionally viable for real-world deployment.

The FM proposed in this paper lays the foundation for future research, focusing on:

- Extending the dataset to cover a broader range of urban environments and intersection types, ensuring greater generalization and real-world applicability.
- Refining the neural network architectures, particularly in optimizing the integration of adaptive mechanisms, such as attention mechanisms and graph-based representations, to enable RL agents to efficiently process and respond to different intersection topologies without requiring intersection-specific retraining.
- Validating the model in large-scale simulations and real-world deployments, bridging the gap between theoretical advancements and practical AI-based traffic management.

Making the benchmark dataset open-source will enable ongoing improvements in RL-based TSC by providing researchers with a standardized, accessible platform for testing and development. The targeted generalizing FM will significantly reduce implementation time, removing the need for costly, intersection-specific retraining, thereby democratizing access to AI-driven traffic optimization for municipalities, researchers, and companies with or without limited resources.

Acknowledgment

This work is part of the KISLEK project (19F1185A) and was supported by the German Federal Ministry for Digital and Transport (BMDV). The authors are responsible for the content.

References

[1] W. Miao, L. Li, and Z. Wang. “A Survey on Deep Reinforcement Learning for Traffic Signal Control”. In: *33rd IEEE CCDC*. IEEE.

[2] K.-L. A. Yau. “A survey on reinforcement learning models and algorithms for traffic signal control”. In: *ACM Computing Surveys* (2017).

[3] H. Wei, G. Zheng, V. Gayah, and Z. Li. “Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation”. In: *ACM SIGKDD Explorations Newsletter* 22.2 (2021).

[4] ZEIT ONLINE GmbH. *Wie KI-Ampeln Fußgänger bevorzugen*. URL: `https://www.zeit.de/mobilitaet/2024-12/strassenverkehr-ampeln-kuenstliche-intelligenz-fussgaenger-sicherheit/komplettansicht`.

[5] A. Müller, V. Rangras, T. Ferfers, F. Hufen, L. Schreckenberg, J. Jasperneite, G. Schnittker, M. Waldmann, M. Friesen, and M. Wiering. “Towards Real-World Deployment of Reinforcement Learning for Traffic Signal Control”. In: *20th IEEE ICMLA*. 2022. URL: `https://10.1109/icmla52953.2021.00085`.

[6] R. S. Sutton. "Reinforcement learning: An introduction". In: *A Bradford Book* (2018).

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2018. DOI: 10.48550/ARXIV.1810.04805. URL: https://arxiv.org/abs/1810.04805.

[8] OpenAI et al. *GPT-4 Technical Report*. 2023. DOI: 10.48550/ARXIV.2303.08774. URL: https://arxiv.org/abs/2303.08774.

[9] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample. *LLaMA: Open and Efficient Foundation Language Models*. 2023. DOI: 10.48550/ARXIV.2302.13971. URL: https://arxiv.org/abs/2302.13971.

[10] L. Torrey and J. Shavlik. "Transfer learning". In: *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. "Attention Is All You Need.(Nips), 2017". In: *arXiv preprint arXiv:1706.03762* 10 (2017), S0140525X16001837.

[12] J. Ault and G. Sharon. "Reinforcement learning benchmarks for traffic signal control". In: *35th Neural Information Processing Systems Datasets and Benchmarks Track*. 2021.

[13] C. Zhao, X. Dai, Y. Chen, Y. Lin, Y. Lv, and F.-Y. Wang. "Parallel Learning Based Foundation Model for Networked Traffic Signal Control". In: *26th IEEE ITSC*. 2023.

[14] X. Zang, H. Yao, G. Zheng, N. Xu, K. Xu, and Z. Li. "Metalight: Value-based meta-reinforcement learning for traffic signal control". In: *Proceedings of the AAAI conference on artificial intelligence*. 2020.

[15] C. Finn, P. Abbeel, and S. Levine. "Model-agnostic meta-learning for fast adaptation of deep networks". In: *International conference on machine learning*. 2017.

[16] J. Beck, R. Vuorio, E. Z. Liu, Z. Xiong, L. Zintgraf, C. Finn, and S. Whiteson. "A survey of meta-reinforcement learning". In: *arXiv preprint arXiv:2301.08028* (2023).

[17] H. Zhang, C. Liu, W. Zhang, G. Zheng, and Y. Yu. "Generalight: Improving environment generalization of traffic signal control via meta reinforcement learning". In: *29th ACM international conference on information & knowledge management*. 2020.

[18] A. Oroojlooy, M. Nazari, D. Hajinezhad, and J. Silva. "Attendlight: Universal attention-based reinforcement learning model for traffic signal control". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 4079–4090.

[19] OpenStreetMap Foundation. *OpenStreetMap (OSM)*. URL: https://www.openstreetmap.org.

[20] Eclipse Foundation. *Eclipse Simulation of Urban MObility (SUMO)*. URL: https://eclipse.dev/sumo/.

[21] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner. "Microscopic Traffic Simulation using SUMO". In: *21st IEEE ITSC*. 2018. URL: https://elib.dlr.de/124092/.

[22] FGSV Verlag GmbH. *Richtlinien für Lichtsignalanlagen - Lichtzeichenanlagen für den Straßenverkehr (RiLSA)*. 2010.

[23] SCHLOTHAUER & WAUER. *LISA – Software-Paket zur Planung, Bewertung und Optimierung von LSA*. URL: https://www.schlothauer.de/software-lisa.

[24] M. Towers, A. Kwiatkowski, J. Terry, J. U. Balis, G. De Cola, T. Deleu, M. Goulão, A. Kallinteris, M. Krimmel, A. KG, et al. "Gymnasium: A Standard Interface for Reinforcement Learning Environments". In: *arXiv preprint arXiv:2407.17032* (2024).

[25] A. Müller and M. Sabatelli. "Safe and psychologically pleasant traffic signal control with reinforcement learning using action masking". In: *25th IEEE ITSC*. 2022. URL: https://ieeexplore.ieee.org/abstract/document/9922306.

[26] FGSV Verlag GmbH. *Handbuch für die Bemessung von Straßenverkehrsanlagen. Technisches Regelwerk. Teil S - Stadtstraßen*. 2015.