BEVLIGHT: GENERALIZABLE TRAFFIC SIGNAL CONTROL FROM UAV BIRD'S-EYE VIEWS

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning-based traffic signal control (TSC) has achieved promising performance, but most methods rely on structured traffic states that must be estimated from sensor observations in real deployments. Visual TSC learns signal decisions directly from images, yet existing methods depend on fixed roadside cameras and remain tied to specific intersection configurations and visual domains. UAVs offer on-demand bird's-eye-view (BEV) observations that jointly capture intersection topology and traffic dynamics, making them a promising sensing platform for visual TSC. We present BEVLight, a generalizable TSC framework that directly maps BEV observations to signal actions. BEVLight grounds visual features onto individual lanes and composes them along the junction topology into movement and phase representations, allowing a shared policy to score candidate phases across heterogeneous intersection layouts and signal plans. To train the policy, we develop a web-based virtual teleoperation platform and collect over 5K human control decisions together with their underlying traffic states. Simulator-provided traffic states first supervise lane- and movement-level representations, after which human phase selections and rankings supervise the phase scorer. Re-rendering the recorded trajectories under diverse appearances preserves both forms of supervision and encourages robustness to visual changes. Closed-loop experiments demonstrate zero-shot generalization across traffic demands, signal plans, and intersections: BEVLight reduces average waiting time by 7–9% relative to perception-to-control visual baselines and matches or outperforms structured-state RL controllers that receive ground-truth traffic states, achieving the lowest waiting time in 12 of 16 scenarios. An offline study further shows that BEVLight transfers to real UAV observations without target-domain adaptation.

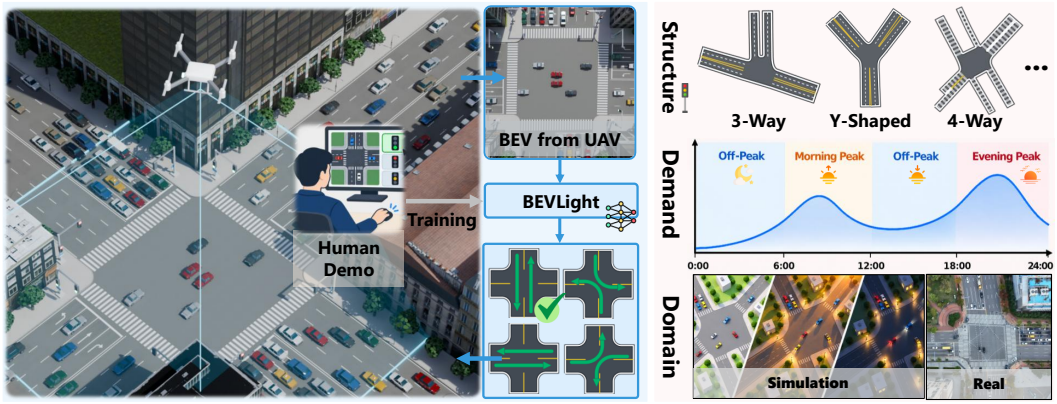
1 INTRODUCTION

Traffic signal control (TSC) requires decisions that adapt to evolving traffic conditions across intersections with different geometries. Recent controllers, from RL policies to LLM agents, have shown promising results, but most operate on structured traffic states such as vehicle counts, queue lengths, and vehicle positions (Wei et al., 2018; 2019a; Chen et al., 2020; Wang et al., 2024b; Lai et al., 2025; Yuan et al., 2026). These states are readily available in simulation but must be estimated in practice, requiring additional perception pipelines whose errors propagate into control.

Visual TSC instead learns signal decisions directly from images (Garg et al., 2022; Chen et al., 2024; Wang et al., 2025b), removing the need for explicit traffic-state estimation. Existing approaches, however, typically consider a fixed or limited set of intersections and signal plans, and obtain observations from roadside cameras that require site-specific installation and provide fixed coverage. UAVs offer an alternative: they can be deployed on demand without permanent roadside infrastructure (Menouar et al., 2017; Fonod et al., 2025), and their elevated viewpoints provide bird's-eye-view (BEV) observations that jointly capture intersection topology and traffic conditions, as illustrated in Figure 1(a).

A key challenge is therefore to learn a single UAV-based visual TSC policy that generalizes across deployment conditions. As illustrated in Figure 1(b), we consider three dimensions of generalization. Structure covers junction geometry and signal configuration, from different signal plans at the same intersection to unseen intersections with different lane topologies and feasible phase sets;

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) (b)

Figure 1: (a) UAV-based visual TSC: BEVLight maps BEV observations to signal phases and is trained from human demonstrations that select and rank feasible phases. (b) The three generalization dimensions considered in this work: structure, demand, and domain.

across these, the same action index can correspond to different movements, so the policy cannot rely on a fixed action layout. Demand reflects variations in traffic flow that produce different congestion and queue patterns at the same intersection. Domain captures appearance changes from weather, illumination, and the gap between simulated and real UAV imagery. Simulation can expose a policy to combinations of these variations that are impractical to collect at real intersections, provided the policy is not tied to any particular structure, demand, or appearance.

To address these challenges, we propose BEVLight, a generalizable UAV-based visual TSC framework that maps BEV observations directly to signal actions. BEVLight grounds visual features onto individual lanes and composes them along the junction topology into movement and phase representations, so that a single shared scorer ranks candidate phases across different layouts and signal plans. Training such a scorer requires supervision on how competing phases should be prioritized, not only which phase is chosen. We therefore learn the policy from human demonstrations that rank all feasible phases at each decision, a preference structure that neither scalar RL rewards nor single-action labels provide. To collect them, we develop a virtual teleoperation platform in which participants control simulated intersections from a BEV view, selecting and ranking feasible phases while the platform records the underlying traffic trajectories. Simulator-provided traffic states first supervise lane- and movement-level representations, after which human selections and rankings supervise the phase scorer. Re-rendering the recorded trajectories under diverse appearances preserves both forms of supervision and encourages appearance-invariant representations.

The contributions of our work are summarized as follows:

- We propose BEVLight, a topology-grounded framework for TSC from UAV BEV observations. By composing visual lane features into movement and phase representations, a single policy handles different intersection layouts and signal plans without retraining.

- We learn signal decisions from human demonstrations that rank all feasible phases, providing preference supervision that a shared phase scorer exploits across signal plans with different phase sets. A virtual teleoperation platform supports collecting over 5K such ranked decisions and re-rendering them under varied visual appearances.

- We demonstrate zero-shot generalization across traffic demands, signal plans, and intersections. BEVLight reduces average waiting time by 7–9% relative to perception-to-control visual baselines, and matches or outperforms structured-state RL methods that receive ground-truth traffic states, achieving the lowest waiting time in 12 of 16 scenarios. An offline study further shows transfer to real UAV videos without target-domain adaptation.

2 PRELIMINARIES AND PROBLEM FORMULATION

2.1 TRAFFIC SIGNAL CONTROL PRELIMINARIES

Figure 2 illustrates a four-way intersection under two phase configurations.

Lane. Incoming lanes $\mathcal{L}^{\text{in}}$ carry vehicles toward the intersection and outgoing lanes $\mathcal{L}^{\text{out}}$ carry them away, with $\mathcal{L} = \mathcal{L}^{\text{in}} \cup \mathcal{L}^{\text{out}}$; in Figure 2, ℓ_1–ℓ_5 are incoming and ℓ_{21}–ℓ_{24} are outgoing.

Movement. A movement $m \subseteq \mathcal{L}^{\text{in}} \times \mathcal{L}^{\text{out}}$ groups admissible lane connections that serve the same directional flow and share signal control, e.g., the opposing through movements $m_1 = \{(\ell_2, \ell_{23}), (\ell_3, \ell_{22})\}$ and $m_5 = \{(\ell_{19}, \ell_6), (\ell_{18}, \ell_7)\}$. We denote the movement set by $\mathcal{M}$.

Phase. A phase $p \subseteq \mathcal{M}$ contains non-conflicting movements that receive the right of way simultaneously, and a phase configuration defines the feasible phase set $\mathcal{P} = \{p_1, \ldots, p_K\}$. Plan A and Plan B in Figure 2 group the same movements differently: Plan A releases $\{m_1, m_5\}$ as Phase-3, whereas Plan B pairs them with other movements.

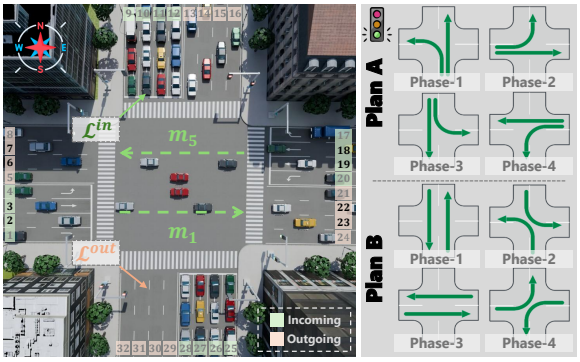


Figure 2: Illustration of a four-way intersection under two phase configurations, each comprising four traffic signal phases.

2.2 UAV BIRD'S-EYE-VIEW TRAFFIC SIGNAL CONTROL

We characterize a junction by its configuration

$$\mathcal{C} = (\mathbf{M}_{\mathcal{L}}, \mathcal{M}, \mathcal{P}), \qquad \mathbf{M}_{\mathcal{L}} = \{\mathbf{M}_\ell : \ell \in \mathcal{L}\}, \tag{1}$$

where $\mathbf{M}_\ell \in \{0,1\}^{H_{\text{BEV}} \times W_{\text{BEV}}}$ is the binary BEV mask of lane ℓ, $\mathcal{M}$ is the movement set, and $\mathcal{P}$ is the feasible phase set of the current signal plan.

At decision step t, the controller receives

$$o_t = \left(\mathbf{I}_t^{\text{BEV}}, \mathcal{C}, k_t, \tau_t\right), \qquad \mathbf{I}_t^{\text{BEV}} = \left(I_{t-H+1}^{\text{BEV}}, \ldots, I_t^{\text{BEV}}\right), \tag{2}$$

where $\mathbf{I}_t^{\text{BEV}}$ collects the latest H UAV images, $k_t \in \{1, \ldots, K\}$ indexes the current phase p_{k_t}, and $\tau_t \geq 0$ is its elapsed duration. The policy selects a feasible target phase $a_t = \pi_\theta(o_t) \in \mathcal{P}$. Our goal is a single policy π_θ that generalizes across junction layouts, signal plans, and traffic demands, and transfers from simulated BEV sequences to real UAV observations.

3 METHOD

BEVLight learns a shared visual policy across junction geometries and signal configurations. Since BEV observations cannot directly map to junction-specific action indices, BEVLight grounds visual features onto lanes and uses traffic topology to construct movement and phase representations with consistent semantics. As illustrated in Figure 3, the framework comprises BEV-to-lane grounding, topology-grounded composition and scoring, and human-supervised policy learning. The policy is trained from human traffic-control demonstrations together with traffic-state supervision encouraging visual representations to preserve control-relevant information.

3.1 TOPOLOGY-GROUNDED BEV-TO-PHASE REPRESENTATION

A consistent mapping from BEV observations to signal actions must account for junction-specific geometries and signal plans, under which the same action index can correspond to different movements (Figure 2). BEVLight resolves this ambiguity by grounding visual features onto individual

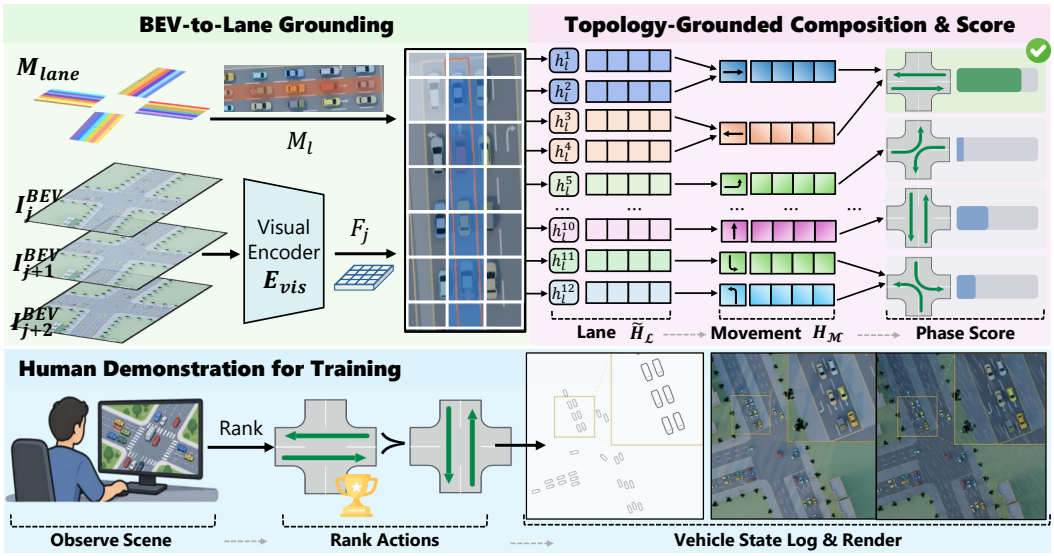


Figure 3: Overview of BEVLight. Visual patch features are grounded onto individual lanes using lane masks with content-adaptive pooling (BEV-to-Lane Grounding). Lane representations are composed along the junction topology into movement and phase representations, and a shared scorer ranks candidate phases (Topology-Grounded Composition & Score). Simulator-provided traffic states supervise lane and movement representations, while human phase selections and rankings supervise the scorer; recorded trajectories are re-rendered under diverse appearances with unchanged supervision (Human Demonstration for Training).

lanes and following the junction topology to progressively compose lane representations into movement and phase representations.

BEV-to-Lane Grounding. For each I_j^{BEV} in $\mathbf{I}_t^{\mathrm{BEV}}$, the visual encoder produces U patch features,

$$\mathbf{F}_j = E_{\mathrm{vis}}(I_j^{\mathrm{BEV}}) = [\mathbf{f}_1^j, \ldots, \mathbf{f}_U^j], \tag{3}$$

where $\mathbf{f}_u^j$ denotes the feature of patch u. To associate these patches with lane ℓ, we derive two geometric attributes from its lane mask $\mathbf{M}_\ell$: the coverage coefficient $c_{\ell,u} \in [0,1]$, which measures the fraction of patch u covered by the lane, and the normalized longitudinal offset $o_{\ell,u} = d_{\ell,u}/L_\ell$, where $d_{\ell,u}$ is the distance from patch u to the stop line along the lane and L_ℓ is the lane length.

Rather than uniformly pooling the patches overlapping a lane, BEVLight assigns content-adaptive weights by combining a mask-derived geometric prior with a learned correction:

$$a_{\ell,u}^j = \underbrace{\log(c_{\ell,u} + \epsilon)}_{\text{geometric mask prior}} + \underbrace{\mathrm{MLP}_{\mathrm{pool}}\left(\mathrm{concat}[\mathbf{f}_u^j, c_{\ell,u}, o_{\ell,u}]\right)}_{\text{content-dependent correction}}, \tag{4}$$

where $\epsilon > 0$ ensures numerical stability. A softmax over the patches overlapping $\mathbf{M}_\ell$ gives $\alpha_{\ell,u}^j = \mathrm{Softmax}_{u:c_{\ell,u}>0}(a_{\ell,u}^j)$. The frame-level lane token is then

$$\mathbf{v}_\ell^j = \sum_u \alpha_{\ell,u}^j \left(\mathbf{f}_u^j + \mathrm{Fourier}(o_{\ell,u})\right). \tag{5}$$

The longitudinal offset, therefore, influences both the adaptive pooling weights and the positional information retained in the resulting lane representation.

Given the lane tokens from consecutive BEV frames, we further model their short-term temporal evolution and interactions across lanes, producing contextualized lane representations $\widetilde{\mathbf{H}}_{\mathcal{L}} = [\widetilde{\mathbf{h}}_\ell^{\mathrm{L}}]_{\ell \in \mathcal{L}}$. The complete temporal and inter-lane modeling procedure is detailed in Appendix B.1.

Topology-Grounded Composition. We next propagate the lane representations through the junction topology. Three fixed binary incidence matrices derived from the junction configuration $\mathcal{C}$, $\mathbf{B}^{\text{in}}_{\mathcal{L}\to\mathcal{M}}$, $\mathbf{B}^{\text{out}}_{\mathcal{L}\to\mathcal{M}}$, and $\mathbf{B}_{\mathcal{M}\to\mathcal{P}}$, encode incoming lane-to-movement, outgoing lane-to-movement, and movement-to-phase relations, respectively. For each movement, we first aggregate its incoming and outgoing lane context:

$$\mathbf{H}^{\text{in}}_{\mathcal{M}} = \mathbf{B}^{\text{in}}_{\mathcal{L}\to\mathcal{M}}\widetilde{\mathbf{H}}_{\mathcal{L}}, \qquad \mathbf{H}^{\text{out}}_{\mathcal{M}} = \mathbf{B}^{\text{out}}_{\mathcal{L}\to\mathcal{M}}\widetilde{\mathbf{H}}_{\mathcal{L}}. \tag{6}$$

For movement m, a shared encoder combines the incoming and outgoing summaries together with their difference,

$$\bar{\mathbf{h}}^{\text{M}}_m = \text{MLP}_M\left(\text{concat}[\mathbf{h}^{\text{M,in}}_m, \mathbf{h}^{\text{M,out}}_m, \mathbf{h}^{\text{M,in}}_m - \mathbf{h}^{\text{M,out}}_m]\right), \tag{7}$$

where the incoming–outgoing difference provides a pressure-like cue for traffic imbalance. We further contextualize the movement representations via self-attention across movements to obtain $\mathbf{h}^{\text{M}}_m$ (Appendix A.1).

We then compose the movements associated with each feasible phase according to $\mathbf{B}_{\mathcal{M}\to\mathcal{P}}$. For phase p, let $\mathbf{H}_p = [\mathbf{h}^{\text{M}}_m]_{m\in p}$. A shared learnable phase query summarizes the relative composition of its constituent movements,

$$\mathbf{c}_p = \text{CrossAttn}\left(\mathbf{q}_P, \mathbf{H}_p, \mathbf{H}_p\right). \tag{8}$$

Since normalized attention primarily captures relative composition, we additionally preserve aggregate traffic magnitude through an additive summary,

$$\mathbf{r}_p = \log\left(\mathbf{1} + \sum_{m\in p} \text{softplus}\left(\text{MLP}_{\text{add}}(\mathbf{h}^{\text{M}}_m)\right)\right). \tag{9}$$

The final phase representation combines these complementary summaries:

$$\mathbf{h}^{\text{P}}_p = \text{MLP}_P\left(\text{concat}[\mathbf{c}_p, \mathbf{r}_p]\right). \tag{10}$$

$\mathbf{h}^{\text{P}}_p$ thus encodes both the movement composition of phase p and its aggregate traffic magnitude.

Candidate Phase Scoring. To accommodate different numbers of feasible phases across junctions and signal plans, BEVLight applies the same scorer to every candidate. Let p_{k_t} denote the current phase and define $\Delta\mathbf{h}^{\text{P}}_p = \mathbf{h}^{\text{P}}_p - \mathbf{h}^{\text{P}}_{p_{k_t}}$. The score of candidate phase p is

$$s_p = \text{MLP}_D\left(\text{concat}[\mathbf{h}^{\text{P}}_p, \mathbf{h}^{\text{P}}_{p_{k_t}}, \Delta\mathbf{h}^{\text{P}}_p, \text{Fourier}\left(\tau_t/\tau_{\max}\right)]\right), \tag{11}$$

where τ_t is the elapsed duration of the current phase and $\tau_{\max}$ is the maximum phase duration. The policy selects the highest-scoring feasible phase, $a_t = \arg\max_{p\in\mathcal{P}} s_p$. Because both phase construction and candidate scoring are shared across junctions, junction-specific signal configurations are handled through the topology rather than junction-specific output heads.

3.2 Human Demonstration Collection

To supervise the shared phase scorer, we collect human traffic-control demonstrations through a web-based virtual teleoperation platform. At each decision step, participants observe the evolving junction through a continuous BEV view, select the preferred phase from the feasible set $\mathcal{P}$, and rank all candidate phases. Each demonstration therefore provides an observation o_t, a selected phase a^*_t, and a preference ranking $r_t(p)$ over $p \in \mathcal{P}$. The ranking distinguishes among non-selected phases that a single action label would treat as equally undesirable.

The simulator additionally records the position of every vehicle, allowing the same traffic evolution to be re-rendered under different visual appearances while preserving the original human preferences. These recorded states also provide traffic supervision for learning control-relevant lane representations. Details of the annotation interface, demonstration protocol, data collection, and processing are provided in Appendix C.

3.3 Policy Training

BEVLight is trained in two stages. The first stage learns traffic-aware lane and movement representations from simulator-provided traffic supervision, while the second stage learns the signal-control policy from human demonstrations.

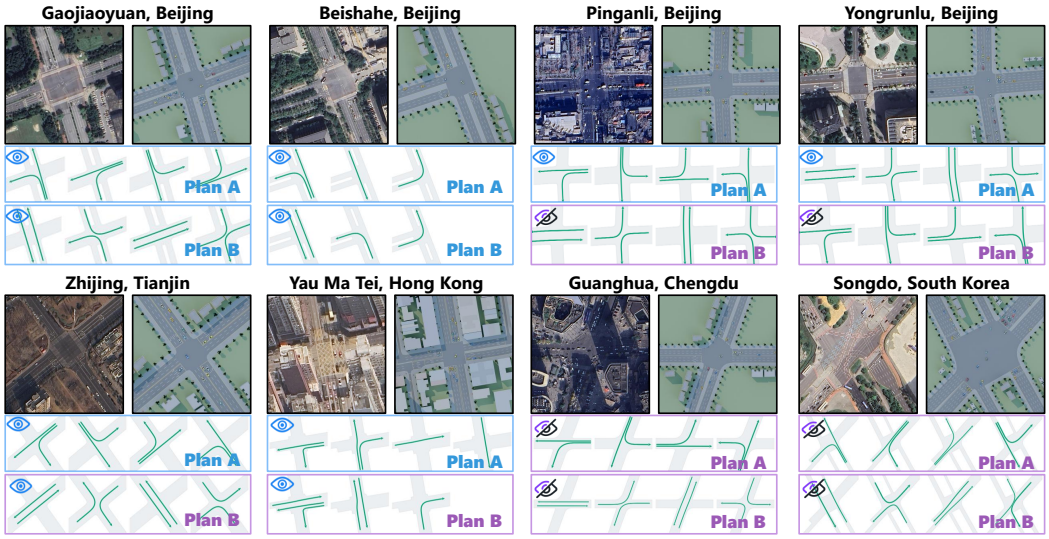


Figure 4: Eight junctions, each with two signal plans. **Blue** borders mark plans used for training, while **purple** borders mark held-out plans.

Stage I: Traffic-Aware Representation Pretraining. We attach auxiliary prediction heads to the lane and movement representations to supervise control-relevant traffic information. At the lane level, the model predicts queue length, vehicle count, and vehicle positions along the lane; at the movement level, it predicts queue length and vehicle count. The first-stage objective is

$$\mathcal{L}_{\text{stage 1}} = \left(\mathcal{L}_q^{\mathcal{L}} + \mathcal{L}_n^{\mathcal{L}} + \mathcal{L}_{\text{Ch}}^{\mathcal{L}}\right) + \left(\mathcal{L}_q^{\mathcal{M}} + \mathcal{L}_n^{\mathcal{M}}\right). \tag{12}$$

This supervision encourages the learned representations to retain traffic information relevant to downstream signal decisions. Detailed definitions of Stage I are provided in Appendix A.2.

Stage II: Human-Supervised Policy Learning. Starting from the pretrained representations, we train the phase scorer using both the human-selected phase and the ranking over all feasible candidates. Let $\boldsymbol{\pi}_t$ denote the policy distribution obtained by softmax-normalizing the candidate scores in Equation 11. We construct a soft ranking target $\mathbf{y}_t$ from the human ranking $r_t(p)$ following the top-one listwise formulation of ListNet (Cao et al., 2007). With $\mathbf{e}_{a_t^*}$ denoting the one-hot target for the selected phase, the second-stage objective is

$$\mathcal{L}_{\text{stage 2}} = \text{CE}(\mathbf{e}_{a_t^*}, \boldsymbol{\pi}_t) + \lambda_{\text{soft}}\, \text{CE}(\mathbf{y}_t, \boldsymbol{\pi}_t). \tag{13}$$

The first term learns the human-selected phase, while the second preserves the relative preferences among all feasible candidates. The construction of the ranking target and the individual training objectives are detailed in Appendix A.2.

4 Experiments

4.1 Experimental Setup

Scenarios. We conduct simulation experiments in TranSimHub (Wang et al., 2025a), which uses SUMO (Lopez et al., 2018) to simulate vehicle movements and provides an interface for rendering BEV observations (Appendix B.1). Figure 4 shows the eight junctions used in our experiments. Each junction has two signal plans, A and B, which share the same layout but differ in their phase configurations (Appendix B.2). We consider five traffic demand patterns (Appendix B.3) and evaluate three cumulative generalization settings by progressively introducing unseen demand patterns, signal plans, and junctions. For offline simulation-to-real evaluation, we use real UAV observations from the Songdo dataset of Fonod et al. (2025).

Table 1: Closed-loop average vehicle waiting time (s) across three generalization settings. Values are mean ± sample standard deviation over the two held-out demand patterns. Lower is better. *, †, and ‡ denote the first-, second-, and third-best results in each row, respectively.

	Conventional		Structured-State RL				Vision-Based		
Junction / Plan	FixedTime	MaxPressure	UniTSA	IntelliLight	AttendLight	MPLight	YOLO-C	BEV-S	**BEVLight**
Unseen Demand									
Beishahe A	18.64 ± 0.39	8.02 ± 1.87	11.67 ± 0.62	7.45† ± 0.06	8.08 ± 0.68	8.26 ± 1.24	8.01 ± 2.07	7.73‡ ± 0.68	**7.40*** ± 0.23
Beishahe B	18.30 ± 1.24	8.01 ± 1.81	10.21 ± 1.38	6.68‡ ± 1.32	6.83 ± 1.22	6.67† ± 0.93	7.91 ± 1.76	7.57 ± 2.07	**6.62*** ± 0.60
Gaojiaoyuan A	26.09 ± 3.99	10.20 ± 2.78	13.04 ± 4.15	10.55 ± 3.25	10.28 ± 4.90	10.71 ± 4.38	9.55† ± 4.06	10.08‡ ± 4.45	**9.12*** ± 2.31
Gaojiaoyuan B	25.12 ± 3.01	10.37 ± 1.38	11.69 ± 4.13	9.34 ± 0.43	**9.18*** ± 2.06	10.18 ± 1.81	9.24† ± 0.25	10.07 ± 0.85	9.24† ± 1.21
Zhijingdao A	25.65 ± 0.66	10.21 ± 0.90	11.89 ± 2.05	9.51‡ ± 1.09	9.42† ± 0.56	9.61 ± 0.71	9.79 ± 0.07	9.95 ± 1.00	**8.89*** ± 0.50
Yau Ma Tei A	21.00 ± 0.11	7.45 ± 1.03	8.66 ± 1.20	6.60‡ ± 0.28	6.39† ± 0.58	6.95 ± 0.68	6.75 ± 1.06	7.04 ± 0.37	**6.09*** ± 0.74
Pinganli A	25.21 ± 0.05	10.23† ± 0.14	13.80 ± 0.98	11.14 ± 2.33	10.44‡ ± 1.01	11.08 ± 1.27	10.52 ± 0.81	10.75 ± 0.80	**10.10*** ± 0.84
Yongrunlu A	27.27 ± 2.06	10.85 ± 0.06	12.14 ± 0.10	10.29‡ ± 0.08	10.24† ± 0.88	10.75 ± 0.67	10.45 ± 0.47	10.41 ± 0.84	**9.54*** ± 0.46
Unseen Signal Plan + Demand									
Zhijingdao B	26.22 ± 5.35	10.13‡ ± 1.11	12.62 ± 1.76	**9.96*** ± 1.13	10.31 ± 0.60	11.00 ± 0.37	10.44 ± 0.28	10.60 ± 0.19	10.09† ± 1.14
Yau Ma Tei B	36.86 ± 1.51	21.21 ± 6.29	24.68 ± 7.49	21.40 ± 11.05	20.02‡ ± 7.85	19.44† ± 3.71	20.84 ± 5.00	21.30 ± 9.20	**16.52*** ± 4.67
Pinganli B	25.62 ± 0.63	9.86 ± 0.78	11.29 ± 0.69	10.40 ± 1.14	**9.65*** ± 0.64	9.81 ± 0.13	9.74‡ ± 0.69	10.24 ± 1.39	9.73† ± 0.56
Yongrunlu B	25.37 ± 3.07	11.14 ± 1.13	11.53 ± 0.08	10.59 ± 0.22	10.24 ± 0.95	10.51 ± 0.18	10.23‡ ± 0.23	10.22† ± 0.75	**10.08*** ± 0.15
Unseen Junction + Plan + Demand									
Guanghua A	22.49 ± 1.03	9.19 ± 1.01	13.39 ± 5.87	8.91 ± 0.04	8.48† ± 0.27	8.85 ± 1.15	8.84‡ ± 0.98	9.59 ± 1.14	**7.95*** ± 0.18
Guanghua B	23.69 ± 0.23	9.12 ± 1.15	12.38 ± 0.51	**8.51*** ± 0.45	9.02† ± 0.94	9.31 ± 0.22	9.38 ± 0.22	10.04 ± 0.90	9.04‡ ± 0.62
Songdo A	30.16 ± 0.58	13.38† ± 0.21	17.23 ± 3.93	13.47‡ ± 0.13	13.87 ± 1.31	14.93 ± 1.80	14.46 ± 1.09	14.31 ± 0.75	**13.08*** ± 1.14
Songdo B	30.28 ± 0.90	19.47 ± 5.32	18.52† ± 2.64	19.45 ± 7.67	19.16 ± 7.93	23.43 ± 9.87	19.43 ± 7.16	19.15‡ ± 4.83	**18.10*** ± 7.13

Compared Methods and Metrics. We compare BEVLight with two conventional controllers, FixedTime and MaxPressure Varaiya (2013), and four structured-state RL controllers: UniTSA Wang et al. (2024b), IntelliLight Wei et al. (2018), AttendLight Oroojlooy et al. (2020), and MPLight Chen et al. (2020). We also include two visual perception-to-control baselines. YOLO-based Count (YOLO-C) assigns detected vehicles to lanes and uses the resulting lane-level counts as inputs to AttendLight, while BEVLight-State (BEV-S) feeds BEVLight-estimated lane states into AttendLight. BEVLight directly predicts a signal phase from BEV observations and junction topology. We report average vehicle waiting time, where lower values indicate better performance. Further details on the compared methods are provided in Appendix B.5.

4.2 Closed-Loop Generalization

Table 1 reports closed-loop average vehicle waiting time under the three generalization settings. BEVLight achieves the lowest waiting time in 12 of 16 evaluation scenarios, demonstrating that human-supervised control from visual observations can match or even outperform RL-based TSC methods supplied with ground-truth traffic states. Human demonstrations provide more than a single preferred action. They rank feasible phases, expressing relative preferences among alternatives with different traffic conditions, as illustrated in Appendix C.3. This richer supervision guides the policy in comparing candidate phases rather than only reproducing the selected action.

Unseen demand. BEVLight achieves the lowest waiting time in seven of eight scenarios and closely matches AttendLight on Gaojiaoyuan B (9.24 s versus 9.18 s). Despite receiving accurate simulator-provided quantities such as queue lengths, the structured-state RL methods do not consistently outperform the human-supervised visual policy. This suggests that the information available to the controller and how it is used to prioritize phases both matter for control performance. Visual observations retain vehicle positions and spatial queue patterns, while human rankings supervise how the competing phases should be prioritized. Averaged across this setting, BEVLight reduces waiting time by 7.2% relative to YOLO-C and 9.0% relative to BEV-S.

Unseen signal plans and demand. Across signal plans, the same action index can correspond to different movement combinations, making fixed preferences over action indices insufficient. BEVLight compares phases through the traffic served by their constituent movements, using the topology-grounded representation. On Yau Ma Tei B, BEVLight reduces waiting time from 19.44 s

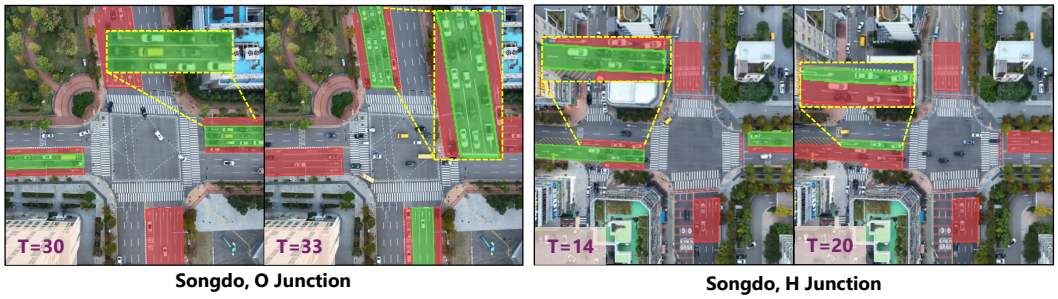


Figure 5: Phase predictions on real UAV imagery from Songdo intersections.

for MPLight, the strongest baseline in this scenario, to 16.52 s, a 15.0% improvement. On Zhijingdao B and Pinganli B, it remains within 0.13 s and 0.08 s of the best baseline, respectively.

Unseen junctions, signal plans, and demand. Finally, we evaluate BEVLight under jointly unseen junction layouts, signal plans, and demand patterns. BEVLight achieves the lowest waiting time on Guanghua A, Songdo A, and Songdo B, while remaining close to AttendLight on Guanghua B (9.04 s versus 9.02 s). It also outperforms both two-stage visual baselines in all four scenarios, with average reductions of 7.6% over YOLO-C and 9.3% over BEV-S. This advantage suggests that directly learning decision-relevant visual cues may alleviate the propagation of intermediate state-estimation errors into control decisions. Together, these results demonstrate that BEVLight can transfer phase preferences learned from human rankings across changes in demand, signal plans, and junction layouts, achieving competitive closed-loop control from visual observations even against RL-based methods using ground-truth traffic states.

4.3 Ablation Study

Table 2 examines the contributions of adaptive pooling, visual pretraining, and soft ranking at Guanghua and Songdo. Combining all three components consistently improves control performance across these unseen settings.

Adaptive pooling alone yields mixed results, suggesting that its additional parameters are difficult to learn from action supervision without explicit perception guidance. Adding visual pretraining improves performance across all four settings, supporting the importance of perception supervision for learning useful lane-level features. Soft ranking further reduces waiting time in all scenarios. These modest gains are consistent with its role in capturing relative phase preferences when multiple actions are plausible, providing richer guidance than a single preferred action.

Table 2: Ablation results in average waiting time (s).

Components			Guanghua		Songdo	
Adaptive Pooling	Visual Pretraining	Soft Loss	A	B	A	B
×	×	×	9.20	10.73	13.96	23.79
✓	×	×	9.38	10.91	13.08	29.03
✓	✓	×	8.78	9.12	12.86	18.20
✓	✓	✓	7.95	9.04	12.77	18.10

4.4 Offline Simulation-to-Real Case Study

We evaluate BEVLight on recorded UAV imagery from four real intersections in Songdo. Figure 5 illustrates BEVLight's zero-shot transfer to real-world traffic scenes through plausible phase decisions without real-world fine-tuning.

At Junction O, BEVLight selects the east-west through phase at $T = 30$ s. By $T = 33$ s, after east-west traffic has discharged, the model favors the north-south through phase, shifting priority to the remaining demand. Similarly, at Junction H, the model initially favors east-west through movements and then switches to the west-approach left-turn movement as through traffic subsides.

These predictions qualitatively align with human control intuition, suggesting that the learned phase preferences remain meaningful under real-world visual conditions.

4.5 Inference Time

We evaluate inference efficiency on an NVIDIA RTX 4090 with batch size 1. As shown in Table 3, BEVLight contains only 1.26M parameters, approximately 53% fewer than YOLO11n. Its full inference latency is 8.59 ms, well below the 10 s control interval, indicating sufficient inference speed on the evaluated hardware.

Table 3: Model size and full inference latency.

Method	Params (M)	GFLOPs	Latency (ms) ↓
YOLO11n	2.662	26.74	18.44
BEV-S	1.260	37.61	8.60
BEVLight	1.260	37.61	8.59

BEVLight's compact design reflects its focus on control-relevant representations: it scores feasible phases from latent traffic features without reconstructing individual vehicle locations during inference. Although its FLOPs exceed those of YOLO11n, its latency is approximately 53% lower, highlighting that operation counts alone do not determine runtime.

5 Related Work

Learning-based TSC has advanced from RL policies (Wei et al., 2019a;b; Chen et al., 2020; Zhang et al., 2022; Wang et al., 2024a) to LLM agents (Lai et al., 2025; Yuan et al., 2026; Zou et al., 2026; Pang et al., 2026). Despite their different architectures, most methods rely on structured traffic states, such as vehicle counts, queue lengths, speeds, and pressure (Wei et al., 2019c). These states are directly available in simulation but require estimation from sensor data in real deployments, introducing errors that complicate transfer. Within this setting, generalizable policies transfer across traffic demands, road topologies, and signal configurations (Oroojlooy et al., 2020; Devailly et al., 2024; Wang et al., 2024b; Lai et al., 2025; Zou et al., 2026), while demonstration-based methods reduce trial-and-error training using data from conventional controllers, virtual experts, or traffic engineers (Xiong et al., 2019; Wang et al., 2022; Ye et al., 2023; Li et al., 2020; 2022). These advances, however, remain grounded in structured states rather than learning control from images.

Vision-based TSC maps images to signal decisions without predefined traffic states. Existing studies explore end-to-end RGB control, synthetic visual data, and vision-language reasoning (Garg et al., 2022; Chen et al., 2024; Wang et al., 2025b), demonstrating the potential of visual control but generally considering a fixed or limited set of junction and signal configurations. UAVs offer a flexible platform for wide-area traffic observation (Menouar et al., 2017), yet existing studies primarily focus on traffic monitoring and trajectory extraction rather than direct signal control (Fonod et al., 2025). BEVLight learns a shared visual TSC policy from human demonstrations on simulated UAV observations, generalizes across heterogeneous junction geometries and signal configurations, and exhibits plausible zero-shot phase decisions on real UAV imagery.

6 Conclusion

We presented BEVLight, a visual traffic signal control framework that maps UAV bird's-eye-view observations directly to signal actions. BEVLight grounds visual features onto lanes and composes them along the junction topology into phase representations, so that a single shared scorer ranks candidate phases across intersections with different layouts and signal plans. The scorer is trained from human demonstrations that select and rank feasible phases, on top of representations pretrained with simulator-provided traffic states and re-rendered under diverse appearances. In closed-loop experiments, BEVLight generalizes zero-shot to unseen demands, signal plans, and junctions, matching or outperforming structured-state RL methods that receive ground-truth traffic states, and transfers to real UAV observations without target-domain adaptation. Future work includes robustness to UAV altitude and viewpoint changes and coordinated control over multiple junctions.

AI USE STATEMENT

We used large language models (LLMs) to assist with language polishing, manuscript organization, exploratory research discussions, and coding. All suggestions were critically reviewed by the authors, who take full responsibility for the scientific claims, analyses, and results.

ETHICS STATEMENT

This study follows the ethical guidelines and submission requirements of ICLR. All datasets, pretrained models, and open-source resources are used in accordance with their respective licenses and intended research purposes. All human annotators were informed about the study and received training before participating. The authors declare no conflicts of interest and take full responsibility for the methodology, experimental results, and conclusions.

REPRODUCIBILITY STATEMENT

Appendix A details the BEVLight architecture, while Appendix B describes the experimental setup and comparison methods. Appendix C explains the human demonstration collection procedure. All code is provided in the supplementary materials to support reproducibility.

REFERENCES

Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the 24th international conference on Machine learning*, pp. 129–136, 2007.

Chacha Chen, Hua Wei, Nan Xu, Guanjie Zheng, Ming Yang, Yuanhao Xiong, Kai Xu, and Zhenhui Li. Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 3414–3421, 2020.

Tiejin Chen, Prithvi Shirke, Bharatesh Chakravarthi, Arpitsinh Vaghela, Longchao Da, Duo Lu, Yezhou Yang, and Hua Wei. SynTrac: a synthetic dataset for traffic signal control from traffic monitoring cameras. In *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2386–2391. IEEE, 2024.

François-Xavier Devailly, Denis Larocque, and Laurent Charlin. Model-based graph reinforcement learning for inductive traffic signal control. *IEEE Open Journal of Intelligent Transportation Systems*, 5:238–250, 2024.

Robert Fonod, Haechan Cho, Hwasoo Yeo, and Nikolas Geroliminis. Advanced computer vision for extracting georeferenced vehicle trajectories from drone imagery. *Transportation Research Part C: Emerging Technologies*, 178:105205, 2025.

Deepeka Garg, Maria Chli, and George Vogiatzis. Fully-autonomous, vision-based traffic signal control: From simulation to reality. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '22, pp. 454–462, Richland, SC, 2022. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9781450392136.

Rahima Khanam and Muhammad Hussain. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*, 2024.

Siqi Lai, Zhao Xu, Weijia Zhang, Hao Liu, and Hui Xiong. LLMlight: Large language models as traffic signal control agents. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 2335–2346, 2025.

Xiaoshuang Li, Fenghua Zhu, and Fei-Yue Wang. Deep behavioral cloning for traffic control with virtual expert demonstration under a parallel learning framework. *IFAC-PapersOnLine*, 53(5): 176–181, 2020.

Xiaoshuang Li, Peijun Ye, Junchen Jin, Fenghua Zhu, and Fei-Yue Wang. Data augmented deep behavioral cloning for urban traffic control operations under a parallel learning framework. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):5128–5137, 2022. doi: 10.1109/TITS.2020.3048151.

Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Fl"otter"od, Robert Hilbrich, Leonhard L"ucken, Johannes Rummel, Peter Wagner, and Evamarie Wiessner. Microscopic traffic simulation using SUMO. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018.

Hamid Menouar, Ismail Guvenc, Kemal Akkaya, A Selcuk Uluagac, Abdullah Kadri, and Adem Tuncer. Uav-enabled intelligent transportation systems for the smart city: Applications and challenges. *IEEE Communications Magazine*, 55(3):22–28, 2017.

Afshin Oroojlooy, Mohammadreza Nazari, Davood Hajinezhad, and Jorge Silva. Attendlight: Universal attention-based reinforcement learning model for traffic signal control. *Advances in Neural Information Processing Systems*, 33:4079–4090, 2020.

Aoyu Pang, Maonan Wang, Man-On Pun, Chung Shue Chen, and Xi Xiong. iLLM-TSC: Integration reinforcement learning and large language model for traffic signal control policy improvement. *IEEE Transactions on Vehicular Technology*, 75(8):15762–15776, 2026. doi: 10.1109/TVT.2026.3674284.

Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of machine learning research*, 22(268):1–8, 2021.

Pravin Varaiya. Max pressure control of a network of signalized intersections. *Transportation Research Part C: Emerging Technologies*, 36:177–195, 2013. doi: 10.1016/j.trc.2013.08.014.

Maonan Wang, Yirong Chen, Yuheng Kan, Chengcheng Xu, Michael Lepech, Man-On Pun, and Xi Xiong. Traffic signal cycle control with centralized critic and decentralized actors under varying intervention frequencies. *IEEE Transactions on Intelligent Transportation Systems*, 25(12):20085–20104, 2024a. doi: 10.1109/TITS.2024.3462153.

Maonan Wang, Xi Xiong, Yuheng Kan, Chengcheng Xu, and Man-On Pun. UniTSA: A universal reinforcement learning framework for v2x traffic signal control. *IEEE Transactions on Vehicular Technology*, 73(10):14354–14369, 2024b. doi: 10.1109/TVT.2024.3403879.

Maonan Wang, Yirong Chen, Yuxin Cai, Aoyu Pang, Yuejiao Xie, Zian Ma, Chengcheng Xu, Kemou Jiang, Ding Wang, Laurent Roullet, et al. TranSimHub: A unified air-ground simulation platform for multi-modal perception and decision-making. *arXiv preprint arXiv:2510.15365*, 2025a.

Maonan Wang, Yirong Chen, Aoyu Pang, Yuxin Cai, Chung Shue Chen, Yuheng Kan, and Man On Pun. VLMLight: Safety-critical traffic signal control via vision-language meta-control and dual-branch reasoning architecture. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 39590–39621. Curran Associates, Inc., 2025b.

Min Wang, Libing Wu, Jianxin Li, Dan Wu, and Chao Ma. Urban traffic signal control with reinforcement learning from demonstration data. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2022. doi: 10.1109/IJCNN55064.2022.9892538.

Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. Intellilight: A reinforcement learning approach for intelligent traffic light control. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2496–2505, 2018.

Hua Wei, Chacha Chen, Guanjie Zheng, Kan Wu, Vikash Gayah, Kai Xu, and Zhenhui Li. Presslight: Learning max pressure control to coordinate traffic signals in arterial network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 1290–1298, 2019a.

Hua Wei, Nan Xu, Huichu Zhang, Guanjie Zheng, Xinshi Zang, Chacha Chen, Weinan Zhang, Yanmin Zhu, Kai Xu, and Zhenhui Li. Colight: Learning network-level cooperation for traffic signal control. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pp. 1913–1922, 2019b.

Hua Wei, Guanjie Zheng, Vikash Gayah, and Zhenhui Li. A survey on traffic signal control methods. *arXiv preprint arXiv:1904.08117*, 2019c.

Yuanhao Xiong, Guanjie Zheng, Kai Xu, and Zhenhui Li. Learning traffic signal control from demonstrations. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pp. 2289–2292, 2019.

Yutong Ye, Yingbo Zhou, Jiepin Ding, Ting Wang, Mingsong Chen, and Xiang Lian. Initlight: Initial model generation for traffic signal control using adversarial inverse reinforcement learning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 4949–4958, 2023. doi: 10.24963/ijcai.2023/550.

Zirui Yuan, Siqi Lai, and Hao Liu. CoLLMLight: Cooperative large language model agents for network-wide traffic signal control. In *International Conference on Learning Representations*, volume 2026, pp. 66687–66715, 2026.

Liang Zhang, Qiang Wu, Jun Shen, Linyuan Lü, Bo Du, and Jianqing Wu. Expression might be enough: Representing pressure and demand for reinforcement learning based traffic signal control. In *International Conference on Machine Learning*, pp. 26645–26654. PMLR, 2022.

Xingchen Zou, Yuhao Yang, Zheng Chen, Xixuan Hao, Yiqi Chen, Chao Huang, and Yuxuan Liang. Traffic-R1: Reinforced llms bring human-like reasoning to traffic signal control systems. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 21823–21838, 2026.

In this appendix, we provide supplementary material for BEVLight, covering implementation and training details, experimental settings, human demonstration collection, additional experimental results, and broader discussion. The appendix is organized as follows:

- Appendix A provides implementation and training details of BEVLight.
- Appendix B describes the experimental setup and comparison methods.
- Appendix C describes the collection and processing of human traffic-control demonstrations.
- Appendix D presents additional real-world case studies.
- Appendix E discusses practical considerations, broader impact, and limitations of BEVLight.

A BEVLight Model Architecture and Algorithm

This section provides additional details of the BEVLight architecture and training objectives introduced in Sec. 3. We first describe the complete computation flow of the topology-grounded BEV-to-phase representation, and then elaborate on the traffic-aware and human-supervised objectives used in the two-stage training procedure.

A.1 BEVLight Architecture

This subsection provides the complete architectural formulation of the topology-grounded BEV-to-phase representation introduced in Sec. 3.1. Starting from a sequence of BEV observations, BEVLight first grounds visual patch features onto individual lanes, models their temporal evolution and interactions across lanes, and then follows the junction topology to progressively construct movement and phase representations. The resulting candidate phases are finally evaluated by a shared scoring function.

BEV-to-Lane Grounding. For each frame I_j^{BEV} in the observation sequence $\mathbf{I}_t^{\mathrm{BEV}}$, the visual encoder produces U patch features,

$$\mathbf{F}_j = E_{\mathrm{vis}}(I_j^{\mathrm{BEV}}) = [\mathbf{f}_1^j, \ldots, \mathbf{f}_U^j], \tag{14}$$

where $\mathbf{f}_u^j$ denotes the feature of patch u.

For each lane ℓ, its mask $\mathbf{M}_\ell$ defines the geometric correspondence between the lane and the visual patches. We characterize each lane–patch pair using two quantities. The coverage coefficient $c_{\ell,u} \in [0,1]$ measures the fraction of patch u covered by $\mathbf{M}_\ell$, while the normalized longitudinal offset

$$o_{\ell,u} = \frac{d_{\ell,u}}{L_\ell} \tag{15}$$

specifies its position along the lane, where $d_{\ell,u}$ is the patch-to-stop-line offset and L_ℓ is the lane length.

Rather than assigning fixed weights according to the lane mask, BEVLight predicts a content-dependent correction for each lane–patch pair. Specifically, a two-layer MLP takes the patch feature together with its geometric attributes and combines the resulting correction with a mask-derived geometric prior:

$$a_{\ell,u}^j = \underbrace{\log(c_{\ell,u} + \epsilon)}_{\text{geometric mask prior}} + \underbrace{\mathrm{MLP}_{\mathrm{pool}}\left(\mathrm{concat}[\mathbf{f}_u^j, c_{\ell,u}, o_{\ell,u}]\right)}_{\text{content-dependent correction}}, \tag{16}$$

where $\epsilon > 0$ ensures numerical stability. The scores are normalized only over patches that overlap with lane ℓ,

$$\alpha_{\ell,u}^j = \mathrm{Softmax}_{u:c_{\ell,u}>0}\left(a_{\ell,u}^j\right). \tag{17}$$

We then aggregate the position-aware patch features into a frame-level lane token,

$$\mathbf{v}_\ell^j = \sum_u \alpha_{\ell,u}^j \left(\mathbf{f}_u^j + \mathrm{Fourier}(o_{\ell,u})\right). \tag{18}$$

The longitudinal offset therefore influences the lane representation in two complementary ways: it conditions the adaptive pooling weights in Eq. 16, and its Fourier encoding preserves within-lane positional information in the aggregated feature.

Temporal and Inter-Lane Modeling. Applying the grounding operation to all H frames gives a token history $[\mathbf{v}_\ell^{t-H+1}, \ldots, \mathbf{v}_\ell^t]$ for each lane ℓ. To explicitly expose short-term traffic changes, we compute the difference between adjacent lane tokens,

$$\Delta \mathbf{v}_\ell^j = \mathbf{v}_\ell^j - \mathbf{v}_\ell^{j-1}, \qquad j = t - H + 2, \ldots, t, \tag{19}$$

and set $\Delta \mathbf{v}_\ell^{t-H+1} = \mathbf{0}$.

A temporal Transformer independently summarizes the history of each lane:

$$\mathbf{h}_\ell^{\mathrm{L}} = \mathrm{Transformer}_{\mathrm{temp}}\left([\mathrm{concat}[\mathbf{v}_\ell^j, \Delta \mathbf{v}_\ell^j]]_{j=t-H+1}^{t}\right). \tag{20}$$

Collecting the temporal representations of all lanes gives

$$\mathbf{H}_{\mathcal{L}} = [\mathbf{h}_\ell^{\mathrm{L}}]_{\ell \in \mathcal{L}}. \tag{21}$$

We subsequently apply multi-head self-attention across lanes,

$$\widetilde{\mathbf{H}}_{\mathcal{L}} = \mathrm{SelfAttn}_L\left(\mathbf{H}_{\mathcal{L}}\right), \tag{22}$$

allowing each lane to incorporate information from other traffic streams at the junction. The row $\widetilde{\mathbf{h}}_\ell^{\mathrm{L}}$ of $\widetilde{\mathbf{H}}_{\mathcal{L}}$ denotes the resulting contextualized representation of lane ℓ.

Lane-to-Movement Composition. We next propagate the contextualized lane representations through the junction topology. The incoming and outgoing lane-to-movement relations are encoded by two fixed binary incidence matrices, $\mathbf{B}_{\mathcal{L}\to\mathcal{M}}^{\mathrm{in}}$ and $\mathbf{B}_{\mathcal{L}\to\mathcal{M}}^{\mathrm{out}}$, respectively. Each entry indicates whether a lane belongs to the incoming or outgoing side of a movement.

The corresponding movement-level summaries are obtained through

$$\mathbf{H}_{\mathcal{M}}^{\mathrm{in}} = \mathbf{B}_{\mathcal{L}\to\mathcal{M}}^{\mathrm{in}} \widetilde{\mathbf{H}}_{\mathcal{L}}, \qquad \mathbf{H}_{\mathcal{M}}^{\mathrm{out}} = \mathbf{B}_{\mathcal{L}\to\mathcal{M}}^{\mathrm{out}} \widetilde{\mathbf{H}}_{\mathcal{L}}. \tag{23}$$

For movement m, the rows $\mathbf{h}_m^{\mathrm{M,in}}$ and $\mathbf{h}_m^{\mathrm{M,out}}$ summarize its incoming and outgoing lane context.

A shared MLP combines these two summaries with their difference:

$$\bar{\mathbf{h}}_m^{\mathrm{M}} = \mathrm{MLP}_M\left(\mathrm{concat}[\mathbf{h}_m^{\mathrm{M,in}}, \mathbf{h}_m^{\mathrm{M,out}}, \mathbf{h}_m^{\mathrm{M,in}} - \mathbf{h}_m^{\mathrm{M,out}}]\right). \tag{24}$$

The difference term provides a pressure-like cue that reflects the imbalance between upstream and downstream traffic conditions associated with the movement.

We further allow different movements to exchange information through self-attention:

$$\mathbf{H}_{\mathcal{M}} = \mathrm{SelfAttn}_M\left(\bar{\mathbf{H}}_{\mathcal{M}}\right), \qquad \bar{\mathbf{H}}_{\mathcal{M}} = [\bar{\mathbf{h}}_m^{\mathrm{M}}]_{m \in \mathcal{M}}. \tag{25}$$

The row $\mathbf{h}_m^{\mathrm{M}}$ of $\mathbf{H}_{\mathcal{M}}$ is the resulting contextualized representation of movement m.

Movement-to-Phase Composition. The movement-to-phase membership is encoded by the fixed binary incidence matrix $\mathbf{B}_{\mathcal{M}\to\mathcal{P}}$. For each feasible phase $p \in \mathcal{P}$, let

$$\mathbf{H}_p = [\mathbf{h}_m^{\mathrm{M}}]_{m \in p} \tag{26}$$

collect the representations of the movements served by phase p.

A shared learnable phase query $\mathbf{q}_P$ attends to these constituent movements:

$$\mathbf{c}_p = \mathrm{CrossAttn}\left(\mathbf{q}_P, \mathbf{H}_p, \mathbf{H}_p\right). \tag{27}$$

The attention summary $\mathbf{c}_p$ captures the relative composition of the movements within phase p. However, since the attention weights are normalized to sum to one, it does not directly preserve their aggregate traffic magnitude.

To retain this complementary information, each movement representation is projected to a non-negative vector,

$$\mathbf{r}_m = \mathrm{softplus}\left(\mathrm{MLP}_{\mathrm{add}}(\mathbf{h}_m^{\mathrm{M}})\right), \tag{28}$$

which is accumulated over the movements belonging to phase p:

$$\mathbf{r}_p = \log\left(\mathbf{1} + \sum_{m \in p} \mathbf{r}_m\right). \tag{29}$$

The element-wise logarithm controls the scale of the additive summary.

Finally, the attention-based composition summary and additive traffic summary are combined to form the phase representation:

$$\mathbf{h}_p^{\mathrm{P}} = \mathrm{MLP}_P\left(\mathrm{concat}[\mathbf{c}_p, \mathbf{r}_p]\right). \tag{30}$$

The resulting $\mathbf{h}_p^{\mathrm{P}}$ therefore captures both the relative composition of the movements served by phase p and their aggregate traffic state.

Candidate Phase Scoring. BEVLight applies the same scoring function to every feasible candidate phase, allowing the policy to operate with signal plans containing different numbers of phases.

Let p_{k_t} denote the currently active phase. For candidate phase p, we compute

$$\Delta\mathbf{h}_p^{\mathrm{P}} = \mathbf{h}_p^{\mathrm{P}} - \mathbf{h}_{p_{k_t}}^{\mathrm{P}}, \tag{31}$$

which describes the representation change associated with switching from the current phase to the candidate.

The shared scorer jointly considers the candidate representation, the current phase representation, their difference, and the elapsed duration of the current phase:

$$s_p = \mathrm{MLP}_D\left(\mathrm{concat}[\mathbf{h}_p^{\mathrm{P}}, \mathbf{h}_{p_{k_t}}^{\mathrm{P}}, \Delta\mathbf{h}_p^{\mathrm{P}}, \mathrm{Fourier}\left(\tau_t/\tau_{\max}\right)]\right), \tag{32}$$

where τ_t denotes the elapsed duration of the current phase and $\tau_{\max}$ is the maximum phase duration.

At each decision step, BEVLight selects

$$a_t = \arg\max_{p \in \mathcal{P}} s_p. \tag{33}$$

Because the lane-to-movement and movement-to-phase relations are supplied through junction-specific topology while all learnable transformations are shared, the same architecture can process junctions with different lane layouts, movement sets, and feasible signal phases.

A.2 Training Objectives and Details

This subsection provides additional details of the two-stage training procedure introduced in Sec. 3.3. Stage I uses simulator-provided traffic states to learn traffic-aware lane and movement representations, while Stage II trains the phase-selection policy from human traffic-control demonstrations.

Stage I: Traffic-Aware Representation Pretraining. We apply auxiliary prediction heads to the contextualized lane representations $\widetilde{\mathbf{h}}_\ell^{\mathrm{L}}$ and movement representations $\mathbf{h}_m^{\mathrm{M}}$. These objectives encourage the intermediate representations to preserve traffic information directly relevant to downstream signal-control decisions.

At the lane level, BEVLight predicts queue length and vehicle count. Let q_ℓ and n_ℓ denote their ground-truth values, and $\hat{q}_\ell$ and $\hat{n}_\ell$ the corresponding predictions. We use L2 regression objectives,

$$\mathcal{L}_q^{\mathcal{L}} = \frac{1}{|\mathcal{L}|}\sum_{\ell \in \mathcal{L}} \|\hat{q}_\ell - q_\ell\|_2^2, \qquad \mathcal{L}_n^{\mathcal{L}} = \frac{1}{|\mathcal{L}|}\sum_{\ell \in \mathcal{L}} \|\hat{n}_\ell - n_\ell\|_2^2. \tag{34}$$

The lane head additionally predicts slotwise vehicle presence and longitudinal position. Summing the predicted presence values yields the corresponding vehicle-count estimate. To preserve fine-grained spatial information, we supervise the predicted and observed vehicle positions using a bidirectional Chamfer objective $\mathcal{L}_{\mathrm{Ch}}^{\mathcal{L}}$. Each observed vehicle position is matched to its nearest predicted

position and vice versa, with the contribution of each predicted position weighted by its predicted presence. This objective encourages $\widetilde{\mathbf{h}}_{\ell}^{\mathrm{L}}$ to retain the locations of individual vehicles relative to the stop line rather than only aggregate traffic statistics.

At the movement level, BEVLight predicts queue length and vehicle count from $\mathbf{h}_{m}^{\mathrm{M}}$. Let q_m and n_m denote their ground-truth values, with predictions $\hat{q}_m$ and $\hat{n}_m$. The corresponding objectives are

$$\mathcal{L}_{q}^{\mathcal{M}} = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \|\hat{q}_m - q_m\|_2^2, \qquad \mathcal{L}_{n}^{\mathcal{M}} = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \|\hat{n}_m - n_m\|_2^2. \tag{35}$$

The complete first-stage objective is

$$\mathcal{L}_{\text{stage 1}} = \left(\mathcal{L}_{q}^{\mathcal{L}} + \mathcal{L}_{n}^{\mathcal{L}} + \mathcal{L}_{\mathrm{Ch}}^{\mathcal{L}}\right) + \left(\mathcal{L}_{q}^{\mathcal{M}} + \mathcal{L}_{n}^{\mathcal{M}}\right). \tag{36}$$

The simulator records the position of every vehicle at each time step. These recorded states provide the supervision above and also allow the same traffic evolution to be rendered under different visual appearances while retaining identical traffic-state targets.

Stage II: Human-Supervised Policy Learning. Starting from the representations learned in Stage I, we train the phase-selection policy using the human demonstrations described in Sec. 3.2. Each demonstration provides the preferred phase a_t^* together with a ranking $r_t(p)$ over every feasible phase $p \in \mathcal{P}$, where rank one denotes the most preferred phase.

Given the candidate scores s_p from Eq. 11, we first obtain the policy distribution

$$\pi_t(p) = \frac{\exp(s_p)}{\sum_{p' \in \mathcal{P}} \exp(s_{p'})}. \tag{37}$$

The selected phase provides a hard supervision target $\mathbf{e}_{a_t^*}$. To additionally preserve the preferences among the remaining candidates, we construct a soft target from the complete human ranking. Following the top-one listwise formulation of ListNet (Cao et al., 2007), we define

$$y_t(p) = \frac{\exp\left(-r_t(p)/|\mathcal{P}|\right)}{\sum_{p' \in \mathcal{P}} \exp\left(-r_t(p')/|\mathcal{P}|\right)}. \tag{38}$$

Scaling the ranks by the number of feasible phases keeps the range of the target logits comparable across signal plans containing different numbers of candidates.

The second-stage objective jointly supervises the selected phase and the complete phase ranking:

$$\mathcal{L}_{\text{stage 2}} = \mathrm{CE}\left(\mathbf{e}_{a_t^*}, \boldsymbol{\pi}_t\right) + \lambda_{\text{soft}}\, \mathrm{CE}\left(\mathbf{y}_t, \boldsymbol{\pi}_t\right). \tag{39}$$

The first term directly learns the human-selected phase, while the second transfers the relative preferences among all feasible candidates.

Because the ranking supervision is defined over the feasible phases of each individual signal plan, the same objective can be applied across junctions with different numbers and compositions of signal phases. Moreover, re-rendered observations corresponding to the same recorded traffic evolution retain the original human ranking because the underlying traffic state is unchanged.

B Experimental Setup

B.1 Experimental Settings

We conduct closed-loop experiments in TranSimHub (Wang et al., 2025a), with microscopic vehicle dynamics simulated by SUMO (Lopez et al., 2018). All simulated roads have a speed limit of 13.9 m/s (approximately 50 km/h). Vehicle speeds follow a Gaussian distribution with a mean of 10 m/s and a standard deviation of 3 m/s, representing typical urban traffic conditions.

For the visual encoder, we employ a lightweight ResNet trained from scratch. It consists of a stride-4 convolutional stem followed by two residual blocks and produces 128-channel feature maps at 1/4 input resolution. Each observation contains five consecutive 448×448 RGB BEV frames.

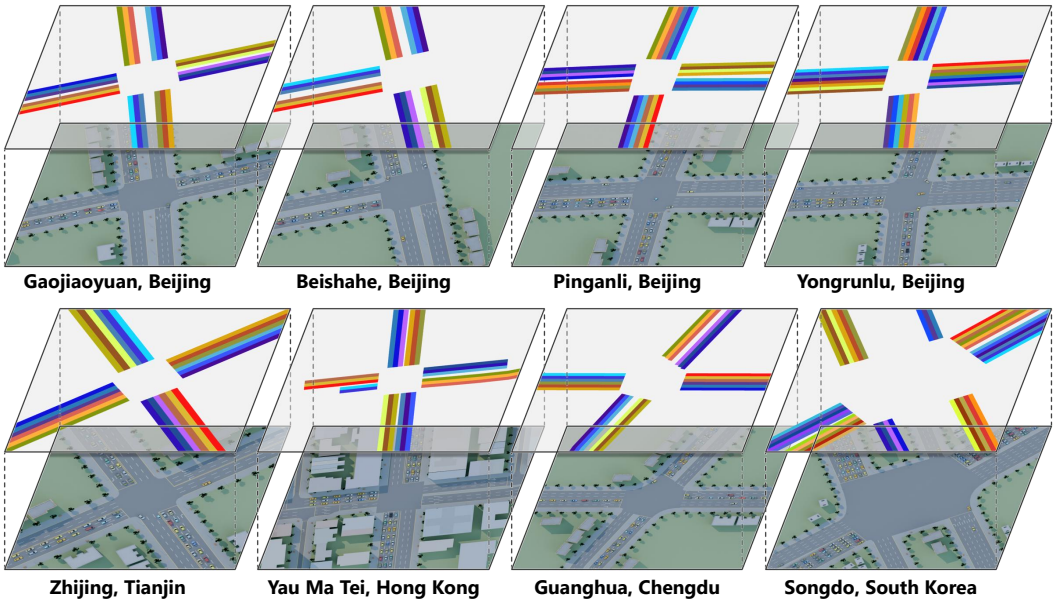


Figure 6: BEV observations and lane masks for the eight simulated intersections. For each scene, the lane masks indicate the visible lane regions used by BEVLight, with different colors denoting individual lanes.

All experiments are conducted on a workstation running Ubuntu 22.04 with an Intel Xeon Gold 6133 CPU, 128 GB of memory, and an NVIDIA GeForce RTX 4090 GPU. Structured-state RL baselines are implemented with Stable-Baselines3 (Raffin et al., 2021) and trained for 10^6 environment steps. UniTSA (Wang et al., 2024b) is trained with PPO, while IntelliLight (Wei et al., 2018), AttendLight (Oroojlooy et al., 2020), and MPLight (Chen et al., 2020) are trained with DQN. Their optimization settings are summarized below.

PPO. For UniTSA (Wang et al., 2024b), we use 16 parallel environments with 128 steps per rollout, resulting in 2,048 transitions per update. Each rollout is optimized for 10 epochs with a minibatch size of 64. The learning rate is 3×10^{-4}, with a discount factor of 0.99 and a GAE parameter of 0.95. We use a policy clipping range of 0.2, disable value-function clipping, and enable advantage normalization. The entropy and value-loss coefficients are set to 0 and 0.5, respectively, with a maximum gradient norm of 0.5.

DQN. For IntelliLight, AttendLight, and MPLight (Wei et al., 2018; Oroojlooy et al., 2020; Chen et al., 2020), we use a learning rate of 10^{-4} and a discount factor of 0.99. The replay buffer stores up to 10^6 transitions, with learning starting after 100 transitions. Each update performs 16 gradient steps with a batch size of 32. We use one-step returns and hard target-network updates ($\tau = 1$) every 10,000 transitions. The exploration rate is linearly annealed from 1.0 to 0.05 over the first 10% of training. Gradients are clipped to a maximum norm of 10.

B.2 Simulation Scenarios

We construct eight simulation scenes from real-world intersection layouts across Beijing, Tianjin, Hong Kong, Chengdu, and Songdo, covering diverse three- and four-arm junction geometries and lane configurations. The scenes represent different urban contexts, ranging from dense inner-city intersections, such as Yau Ma Tei in Hong Kong, to newly developed districts with wider road layouts, such as Songdo.

Figure 6 visualizes the rendered BEV observation and corresponding lane masks for each intersection. The masks define the lane regions used by BEVLight and illustrate the visible extent of both incoming and outgoing lanes under the UAV viewpoint. Table 4 further reports the ground coverage and median visible lane length at a simulated UAV height of 100 m.

Table 4: Spatial coverage of the eight simulated intersections.

	Ground coverage		Visible lane length (m)		
Intersection	Footprint (m × m)	Area (m^2)	Min.	Mean	Max.
Gaojiaoyuan	140.86 × 140.86	19,842	58.13	60.00	62.09
Beishahe	131.48 × 131.48	17,287	50.82	57.90	62.87
Pinganli	147.89 × 147.89	21,871	55.22	58.27	60.67
Yongrunlu	147.50 × 147.50	21,756	56.59	59.09	61.20
Zhijingdao	111.95 × 111.95	12,533	56.01	60.13	64.10
Yau Ma Tei	148.48 × 148.48	22,046	46.29	56.89	61.69
Guanghua	157.07 × 157.07	24,671	57.74	60.41	64.05
Songdo	155.51 × 155.51	24,183	42.76	58.07	69.98

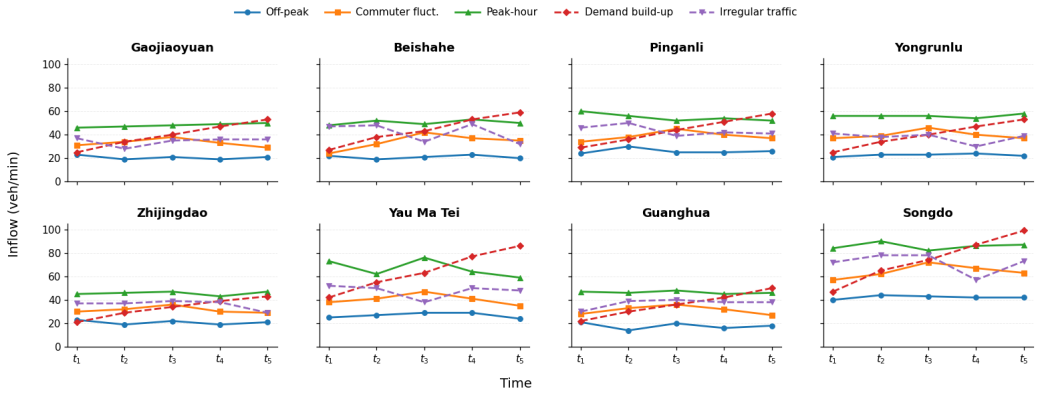


Figure 7: Traffic-demand profiles for the eight simulated scenes.

B.3 Traffic Demand

For each scene, we construct five demand patterns with different traffic patterns. Three patterns are used for training: *off-peak*, characterized by low and relatively stable inflow; *peak-hour*, characterized by sustained high inflow; and *commuter fluctuation*, which introduces structured variations over time. For evaluation, we introduce two held-out demand patterns that are not observed during training. *Demand build-up* progressively increases the inflow over time, while *irregular traffic* introduces stochastic temporal variations.

Figure 7 shows the temporal profiles of the five demand patterns across the eight scenes. Although the absolute inflow varies across scenes, the demand patterns preserve their intended temporal characteristics. Table 5 further summarizes the mean and standard deviation of each demand profile over the five time intervals.

B.4 Real-UAV Data

For real-world evaluation, we use UAV videos from the dataset introduced by Fonod et al. Fonod et al. (2025), selecting four intersections in Songdo, denoted as O, P, T, and H. The videos are captured at approximately 140–150 m above ground and provide bird's-eye views of diverse intersection geometries and lane configurations. These scenes are used to evaluate the transfer of BEVLight from simulation to real UAV observations.

Figure 8 shows representative UAV observations and the corresponding lane masks for the four intersections. The masks indicate the visible lane regions used by BEVLight and illustrate how the lane-level representation is transferred from simulation to real UAV imagery. Table 6 further reports the visible lane lengths in each scene.

Table 5: Traffic-demand statistics across the eight simulated scenes. Each entry reports the mean ± standard deviation of the vehicle inflow over five time intervals, in vehicles per minute.

	Training			Held-out Evaluation	
Scene	Off-peak	Commuter fluctuation	Peak-hour	Demand build-up	Irregular traffic
Gaojiaoyuan	20.6 ± 1.5	33.0 ± 3.0	48.0 ± 1.4	39.8 ± 9.8	34.4 ± 3.3
Beishahe	21.0 ± 1.4	34.0 ± 6.0	50.4 ± 1.9	44.0 ± 11.2	42.0 ± 7.4
Pinganli	26.0 ± 2.1	38.8 ± 3.7	54.8 ± 3.0	43.6 ± 10.3	43.6 ± 3.9
Yongrunlu	22.6 ± 1.0	39.8 ± 3.3	56.0 ± 1.3	39.8 ± 9.8	37.6 ± 3.9
Zhijingdao	20.8 ± 1.6	31.4 ± 2.5	45.6 ± 1.5	33.2 ± 7.7	36.0 ± 3.6
Yau Ma Tei	26.8 ± 2.0	40.4 ± 4.0	66.8 ± 6.6	64.6 ± 15.6	47.6 ± 5.0
Guanghua	17.8 ± 2.6	31.2 ± 3.3	46.4 ± 1.0	36.0 ± 9.6	37.0 ± 3.6
Songdo	42.2 ± 1.3	64.2 ± 5.0	85.8 ± 2.7	74.4 ± 17.9	71.6 ± 7.7

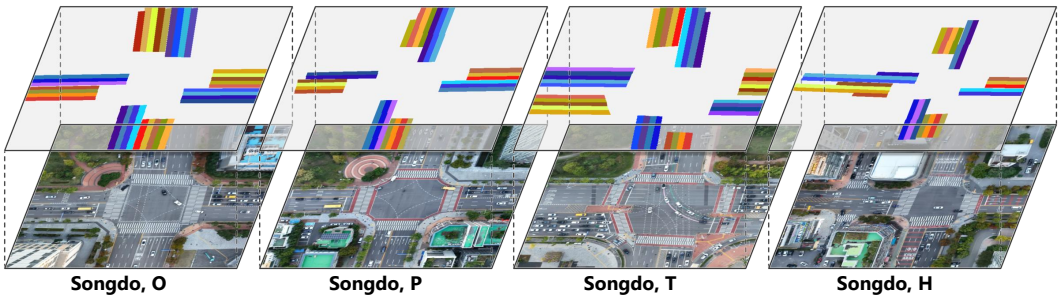


Figure 8: Real UAV observations and corresponding lane masks for intersections O, P, T, and H.

B.5 Compared Methods

We compare BEVLight with three groups of traffic signal methods: two conventional methods, four structured-state RL methods, and two vision-based two-stage methods. The conventional and RL methods directly access structured traffic states provided by the simulator, whereas the vision-based methods first extract lane-level traffic information from BEV observations before applying an RL controller.

Conventional Controllers. We include two commonly used traffic-responsive controllers:

- **FixedTime** follows a predefined signal schedule and holds each phase for 20 s in our experiment, independent of the observed traffic conditions.
- **MaxPressure** (Varaiya, 2013) selects the signal phase with the highest traffic pressure, computed from the imbalance between upstream and downstream traffic states. It provides a strong adaptive baseline without requiring policy learning.

Structured-State RL Methods. These methods receive lane-level traffic states directly from the simulator and learn signal-control policies through RL:

- **UniTSA** (Wang et al., 2024b) introduces a junction-matrix representation to accommodate different intersection layouts and uses traffic-state augmentation to expose the policy to diverse intersection structures and traffic conditions.
- **IntelliLight** (Wei et al., 2018) learns adaptive signal control from structured intersection states using a value-based RL formulation.
- **AttendLight** (Oroojlooy et al., 2020) employs attention mechanisms to aggregate traffic information across lanes and candidate signal phases, allowing a shared policy to operate across different intersection configurations.
- **MPLight** (Chen et al., 2020) incorporates traffic-pressure information into RL-based signal control and models competition among candidate phases.

Table 6: Spatial coverage of the four real-UAV intersections.

	Ground coverage		Visible lane length (m)		
Intersection	Footprint (m × m)	Area (m^2)	Min.	Mean	Max.
Songdo, O	118.5×118.5	14,050	25.61	30.85	37.90
Songdo, P	119.3×119.3	14,239	22.28	27.11	32.90
Songdo, T	120.7×120.7	14,568	14.53	25.53	43.24
Songdo, H	121.2×121.2	14,695	17.54	25.06	51.38

Vision-Based Two-Stage Methods. We further construct two baselines that first recover traffic states from BEV observations and then use AttendLight for signal control:

- **YOLO-Count** uses YOLOv11 Khanam & Hussain (2024) to detect vehicles from BEV images with oriented bounding boxes and assigns the detections to individual lanes to obtain lane-level vehicle counts. The detector is fine-tuned on synthetic images generated from the same training scenes used by BEVLight, ensuring a fair comparison without introducing additional visual training data. The resulting lane-level states are then provided to AttendLight for phase selection.
- **BEVLight-State+AttendLight** uses BEVLight's auxiliary lane-level prediction heads to estimate traffic states directly from BEV observations. These predicted states are subsequently passed to AttendLight for signal control, providing a two-stage alternative that does not rely on explicit object detection.

In contrast to the two vision-based baselines, BEVLight directly maps BEV observations and junction topology to signal actions without an intermediate vehicle-detection or structured-state reconstruction stage.

C Human Demonstrations

C.1 Web-Based Human Demonstration Collection

We develop a virtual teleoperation platform to collect human traffic-control demonstrations. Eight graduate researchers with backgrounds in computer science or transportation, including master's and PhD students, participated in the data collection. As illustrated in Fig. 9, the interface presents a bird's-eye view of the current intersection together with all feasible signal actions. To make the control semantics of each action explicit, hovering over an action highlights the corresponding lanes and permitted traffic movements directly on the intersection view.

Participants make a control decision every 10 s and rank all feasible actions using arrow-based controls. Rather than recording only the selected action, the ranking provides relative preference supervision among feasible signal phases, indicating not only which action is preferred but also how the remaining alternatives are ordered. This provides richer supervision for learning the phase-scoring function and avoids treating all non-selected actions as equally undesirable. For each decision, we record the complete simulator state, including the positions of all vehicles, together with the human-provided action ranking. The recorded vehicle states are subsequently used to reconstruct the same traffic scene under different visual appearances, as described in Appendix C.2.

To ensure the quality of the collected demonstrations, we use MaxPressure as a reference controller and perform filtering at the episode level. Each human-controlled episode is compared with the MaxPressure controller under the same traffic scenario, and only episodes achieving a lower average waiting time are retained. After filtering, the resulting dataset contains more than 5K human decisions, corresponding to approximately 14 hours of traffic-control demonstrations.

C.2 Appearance-Diverse Scene Rendering

For each human decision, we record the underlying simulator state rather than storing only the rendered observation. In particular, the recorded state contains the positions of all vehicles at the decision time, allowing the same traffic configuration to be reconstructed independently of its original visual appearance. The associated human action ranking is preserved throughout this process.

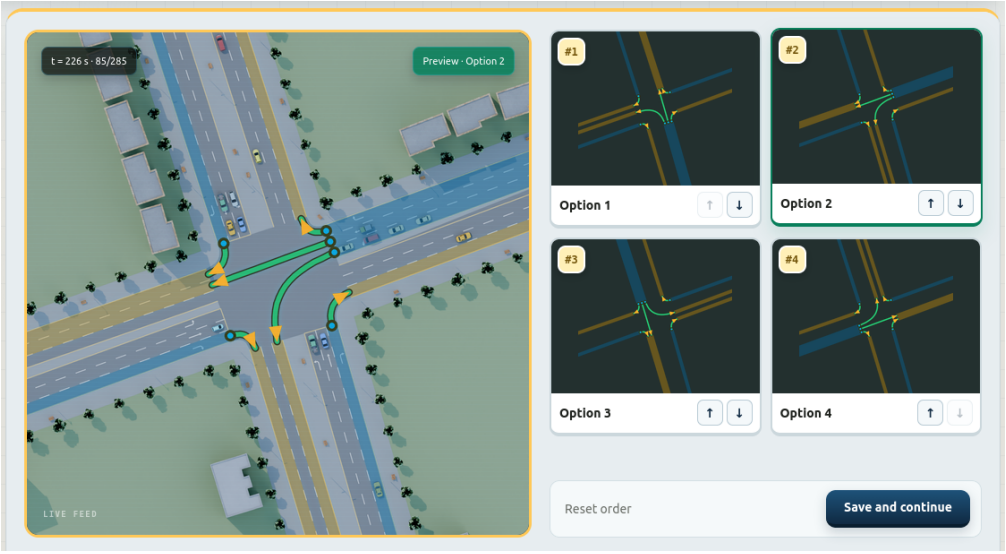


Figure 9: Web-based virtual TSC platform for collecting human demonstrations. The interface displays the current bird's-eye-view traffic scene and all feasible signal actions. Hovering over an action highlights its corresponding lanes and permitted movements on the intersection.

In this way, a single human demonstration can provide supervision for multiple visually different observations while retaining exactly the same traffic state and control preference.

We render each recorded state under five visual conditions: *Morning*, *Noon*, *Dusk*, *Dusk-Rain*, and *Afternoon-Rain*. The rendering process changes illumination, weather, and scene appearance while keeping vehicle positions, junction geometry, and control supervision unchanged. As illustrated in Figs. 10–13, each example additionally includes a visualization of the recorded vehicle configuration before rendering. The figures cover all eight intersections used in our experiments: *Beishahe, Beijing*, *Gaojiaoyuan, Beijing*, *Pinganli, Beijing*, *Yongrunlu, Beijing*, *Guanghua, Chengdu*, *Yau Ma Tai, Hongkong*, *Songdo, SouthKorea*, and *Zhijingdao, Tianjin*.

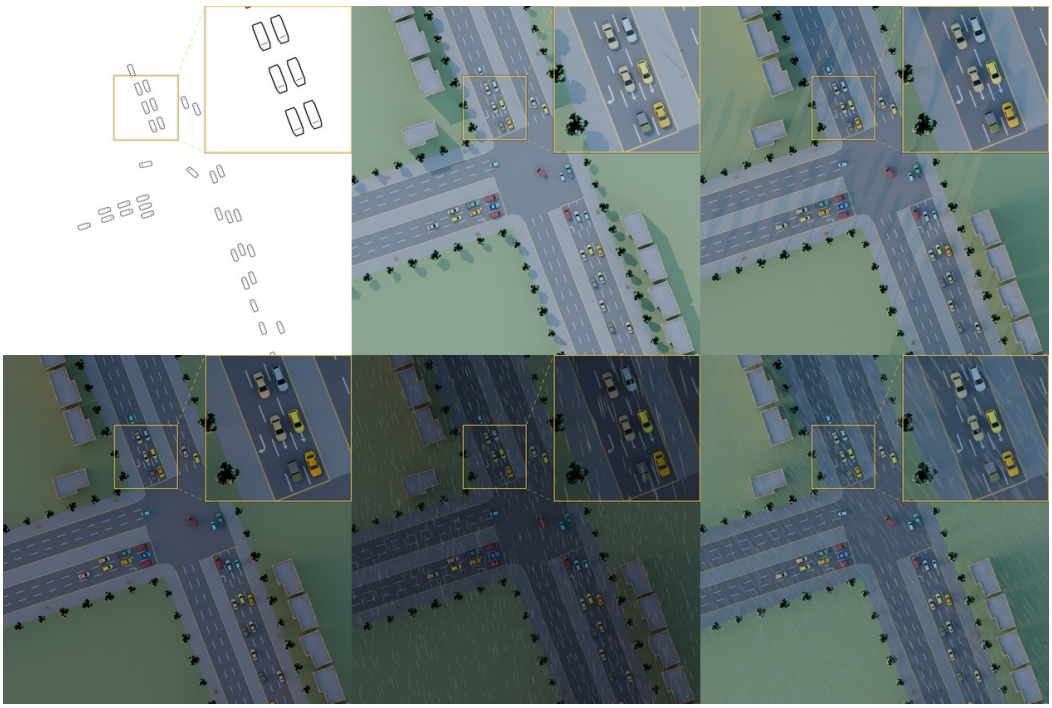
This state-preserving rendering strategy substantially increases the visual diversity associated with the collected human demonstrations. More importantly, because the same control preference is observed across different appearances of the same traffic state, the model is encouraged to rely on traffic-relevant visual information rather than appearance-specific cues. We therefore use these appearance-diverse renderings throughout policy training as an important component for improving robustness to visual-domain changes and supporting transfer from simulated BEV observations to real UAV imagery.

C.3 Human Control Preferences and Action Ranking

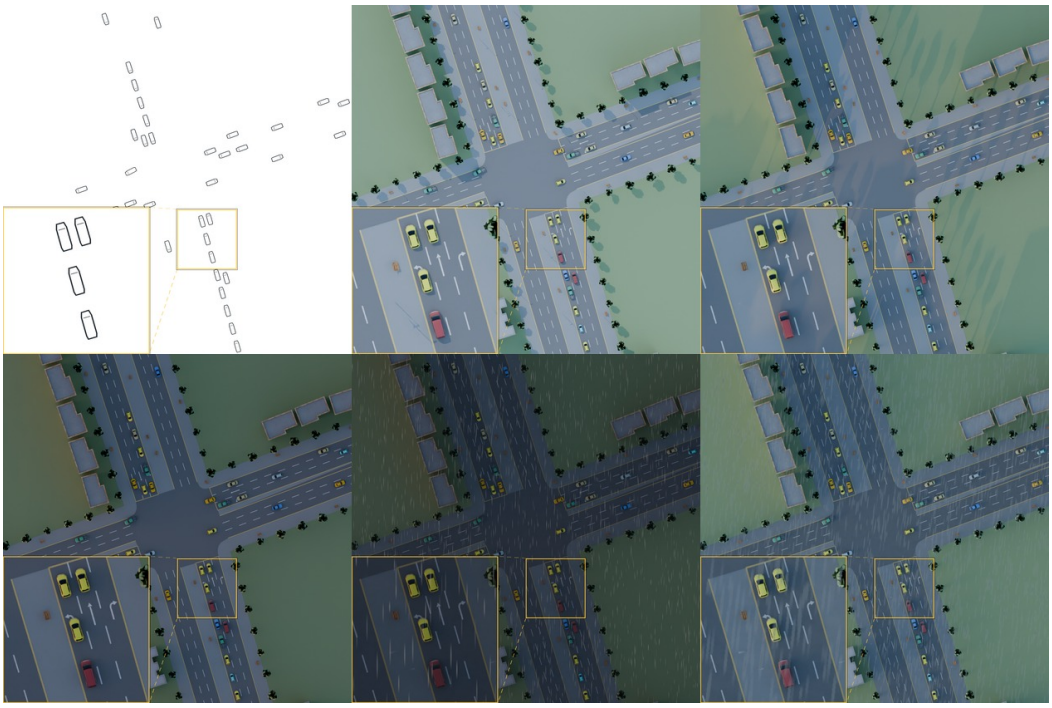
In addition to the selected action, we collect a complete ranking over all feasible signal phases at each decision step. The motivation is that a single selected-action label provides only the top choice, while all non-selected actions would otherwise be treated equivalently. In the TSC task, however, feasible actions that are not selected can still differ substantially in their relative priority. By recording the full ranking, we preserve this relative preference structure and provide richer supervision for learning the phase-scoring function in BEVLight.

Fig. 14 presents representative examples from four intersections: *Beishahe, Beijing*, *Gaojiaoyuan, Beijing*, *Pinganli, Beijing*, and *Yau Ma Tei, Hong Kong*. Each example shows the traffic observation at a decision step together with the human-assigned ranking over all feasible signal phases.

More specifically, the ranking supervision specifies not only the preferred phase, but also the relative ordering of the remaining feasible phases. In Figure 14(a), for example, the action ranked third serves no vehicles under the current traffic state, whereas the second-ranked action serves active traffic, providing a meaningful distinction between two non-selected actions. Meanwhile, the first-

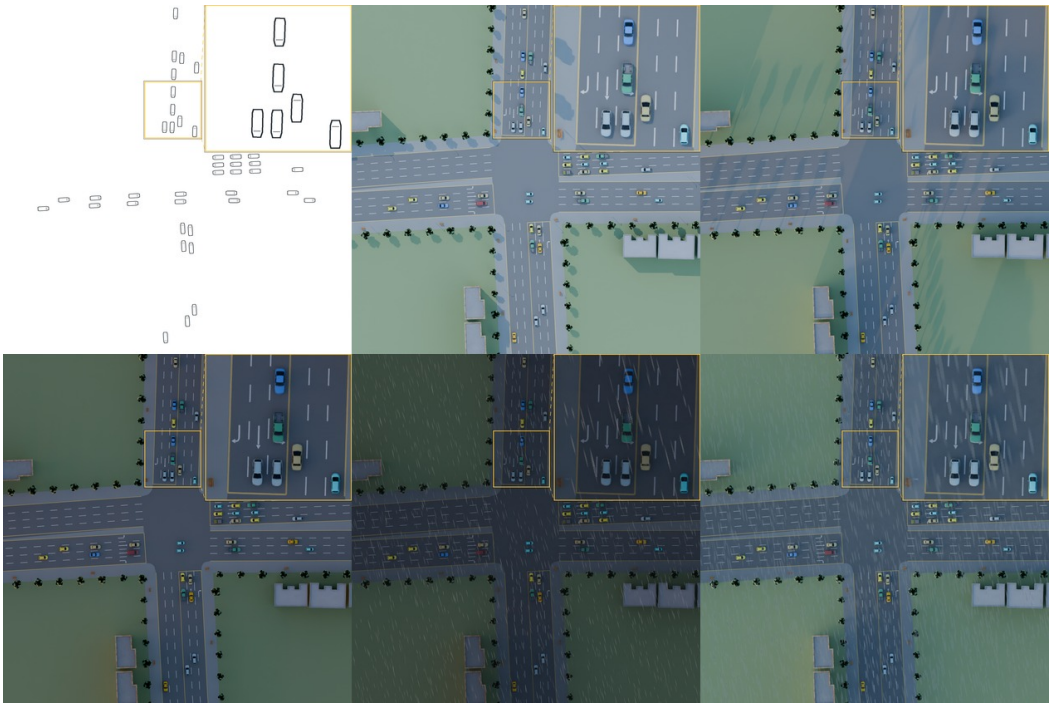


(a) Beishahe, Beijing

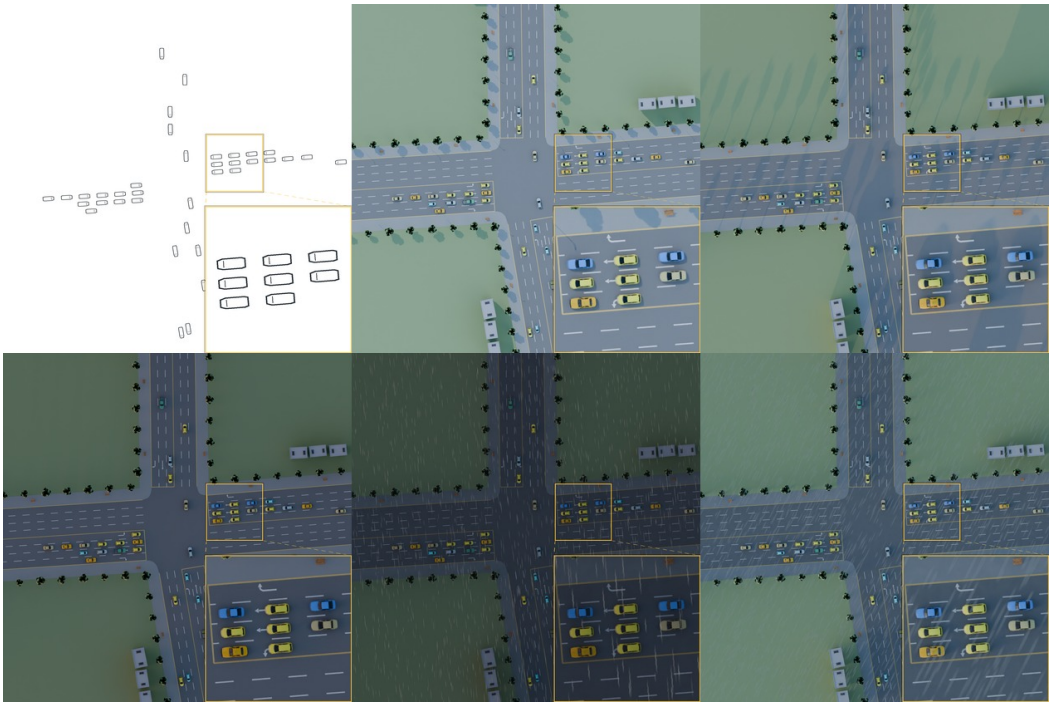


(b) Gaojiaoyuan, Beijing

Figure 10: Appearance-diverse rendering of recorded human demonstration states. For each intersection, the top-left panel visualizes the recorded vehicle configuration, while the remaining panels show the same traffic state under five visual conditions: *Morning*, *Noon*, *Dusk*, *Dusk-Rain*, and *Afternoon-Rain*. The underlying traffic state and human action ranking remain unchanged across all rendered variants. The examples shown here correspond to *Beishahe, Beijing* and *Gaojiaoyuan, Beijing*.



(c) Pinganli, Beijing

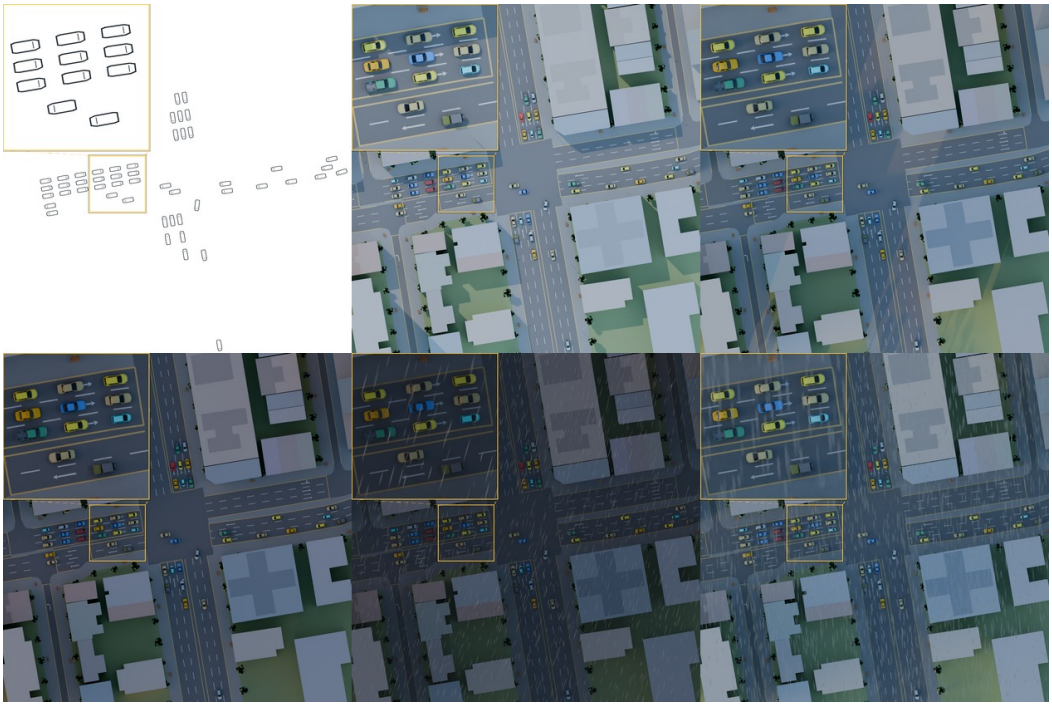


(d) Yongrunlu, Beijing

Figure 11: Appearance-diverse rendering examples for *Pinganli, Beijing* and *Yongrunlu, Beijing*.



(e) Guanghua, Chengdu

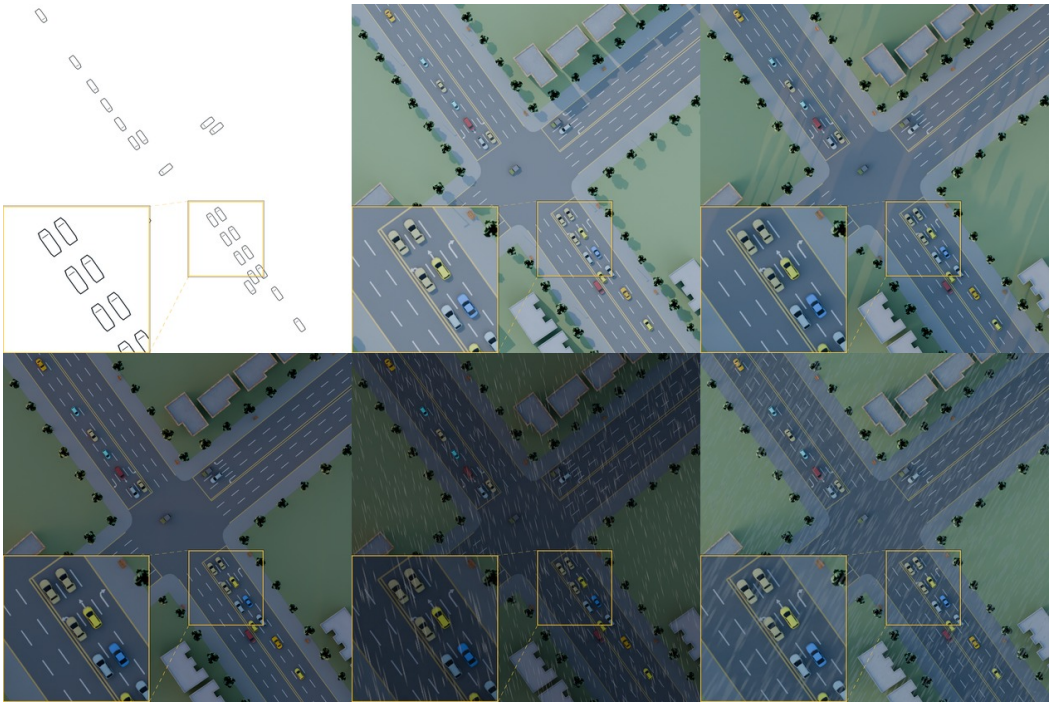


(f) Yau Ma Tei, Hongkong

Figure 12: Appearance-diverse rendering examples for *Guanghua, Chengdu* and *Yau Ma Tei, Hongkong*.

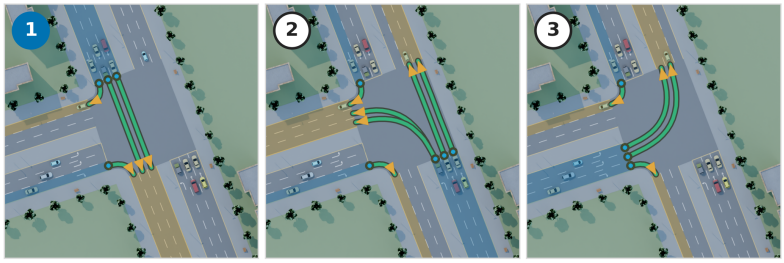


(g) Songdo, SouthKorea

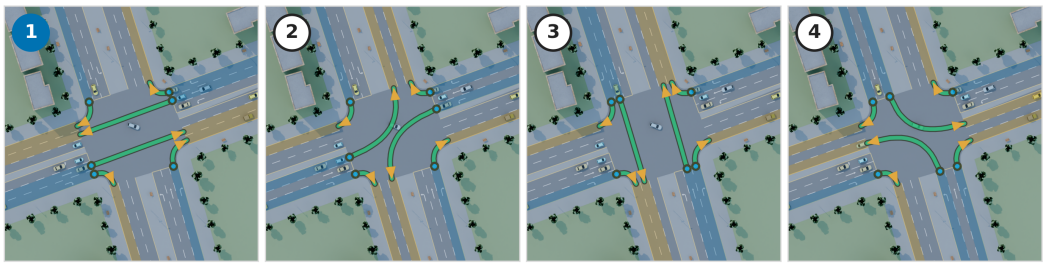


(h) Zhijingdao, Tianjin

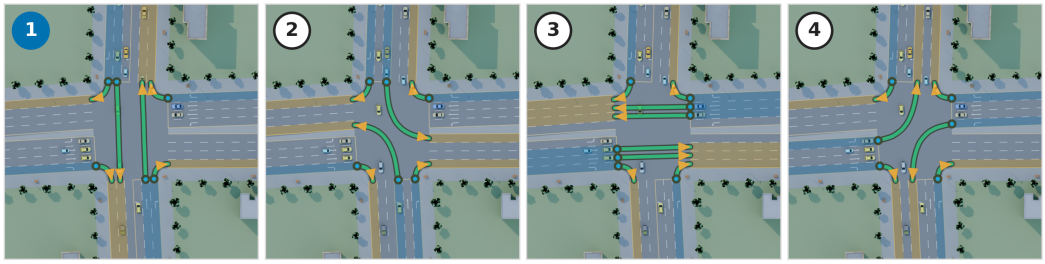
Figure 13: Appearance-diverse rendering examples for *Songdo, SouthKorea* and *zhijingdao, Tianjin*.



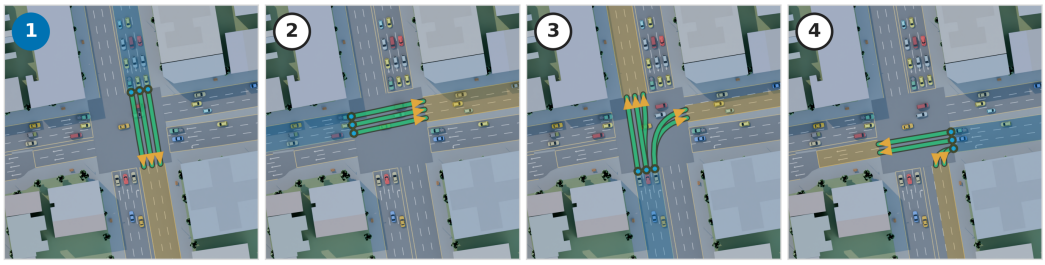
(a) Beishahe, Beijing



(b) Gaojiaoyuan, Beijing



(c) Pinganli, Beijing



(d) Yau Ma Tei, Hong Kong

Figure 14: Representative examples of human control preferences and action rankings. Each example shows a traffic observation together with the ranking assigned by the human operator over all feasible signal phases. The ranking preserves the relative ordering among feasible actions rather than retaining only the top-ranked phase, providing richer supervision for learning the phase-scoring function in BEVLight.

and second-ranked actions correspond to traffic conditions of comparable congestion and are both placed ahead of the remaining alternatives. In Figure 14(c), the ranking further reflects the current signal context: the active phase is prioritized to allow vehicles already being served to clear the intersection before switching to another phase. In Figure 14(d), the highest-ranked action serves visibly heavier traffic than the other alternatives and is accordingly assigned the highest priority. These examples illustrate that the full ranking preserves distinctions among both selected and non-selected actions that would be lost if only the final action were recorded.

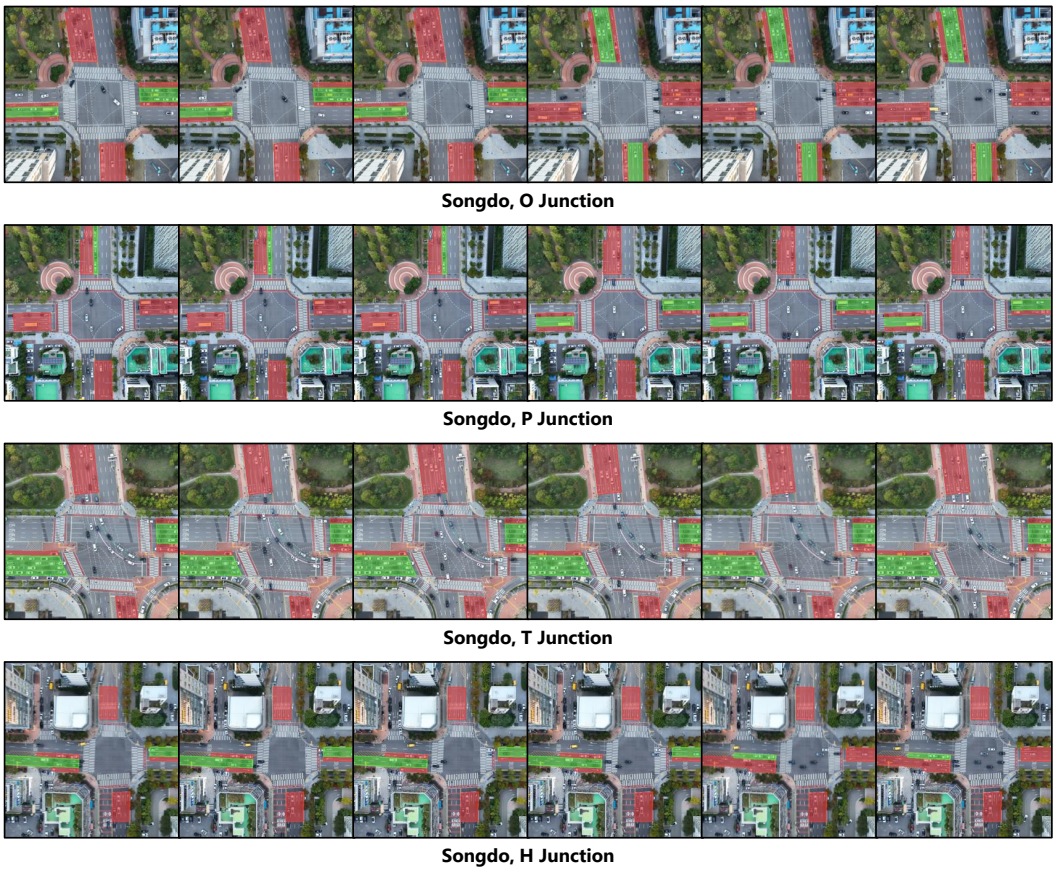


Figure 15: Zero-shot phase predictions over six consecutive UAV frames at Junctions O, P, T, and H in Songdo. Predictions are evaluated offline without controlling actual signals.

D Additional Real-World Case Studies

Figure 15 presents six consecutive UAV frames from each of four real intersections in Songdo: Junctions O, P, T, and H, together with BEVLight's predicted signal phases. The model is applied directly to real UAV imagery without real-world fine-tuning. Across these sequences, its phase predictions respond to changes in vehicle distribution, shifting priority as traffic clears and demand changes across movements.

Since the evaluation uses recorded videos, the predicted phases do not affect actual signals or subsequent traffic. These examples therefore provide qualitative evidence of zero-shot transfer rather than closed-loop control performance: BEVLight produces plausible phase decisions that adapt to evolving traffic conditions under real-world visual inputs.

E Discussion

Practical considerations. Deployment hinges on the two properties evaluated in Sec. 4: a policy that is not tied to a particular junction or signal plan can be moved between sites without retraining, and a policy trained entirely in simulation can be applied to real UAV observations without collecting and labeling real control data, which is impractical at scale. Beyond the UAV video stream, BEVLight requires lane masks in the BEV frame and the lane-to-movement and movement-to-phase relations of the active signal plan. The relations are part of any signal timing plan, and the masks are specified once per junction and viewpoint. Decisions are made every 10 s, leaving ample margin for transmission and inference. The policy only selects among feasible phases, so it cannot release conflicting movements; phase transitions and minimum green times remain under the signal con-

troller, and a human operator can override the policy at any step. UAV observation is unavailable under strong wind, heavy precipitation, or restricted airspace, so BEVLight is best viewed as an on-demand controller that complements installed infrastructure.

Limitations. Our simulation-to-real evaluation is offline: queue estimates and phase decisions transfer to recorded UAV videos, but we do not control real signals, and at the two sites with the sparsest queues the queue-count error remains above an always-zero predictor. BEVLight controls a single junction, and coordination across junctions is left for future work. The policy imitates demonstrations from graduate researchers filtered against MaxPressure, so it inherits their quality and is not optimized for a long-horizon objective. Lane masks are assumed static, so UAV drift would require image registration. Finally, the simulated UAV height of 100 m differs from the 140–150 m of the real recordings, and appearance rendering covers daytime and rain but not night-time.

Broader impact. BEVLight lowers the infrastructure barrier for adaptive signal control and can be deployed on demand where fixed sensing is absent or temporarily disrupted, with reduced waiting time translating into lower fuel consumption and emissions. UAV imagery raises privacy and regulatory concerns: BEVLight operates on lane-level features from images captured above 100 m, at which individuals and license plates are not identifiable, and does not track vehicles across frames, but deployments should follow local airspace regulations and data-retention policies. The demonstrations were collected in simulation from consenting participants and contain no personal data. Since a learned controller reflects the priorities of its demonstrations, operators should verify that its phase preferences do not systematically disadvantage particular approaches before deployment. The virtual teleoperation platform is not specific to BEVLight. It records complete simulator states alongside human decisions, so any recorded demonstration can be re-rendered under new appearances or used to supervise other visual controllers, and it can be extended to other observation modalities such as roadside cameras. We plan to release the platform and the collected demonstrations to support future work on human-supervised visual TSC.