Sim2Real Transfer for Traffic Signal Control

"Mitigate the performance impairment for RL agents!"

Longchao Da[1,2], Hao Mei[1,2], Romir Sharma[3] and Hua Wei[1,2*]

[1]Arizona State University

[2]New Jersey Institute of Technology

[3]West Windsor-Plainsboro High School South

IEEE CASE AUCKLAND 2023

19TH INTERNATIONAL CONFERENCE ON AUTOMATION SCIENCE AND ENGINEERING

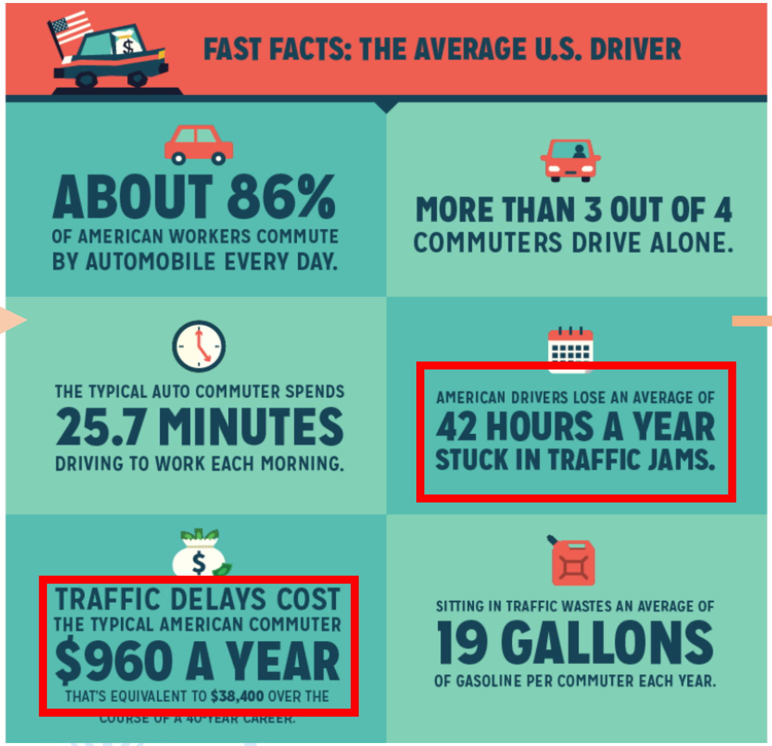
Hosted by UNIVERSITY OF AUCKLAND

Supported by IEEE, IEEE Robotics & Automation

26 – 30 August 2023

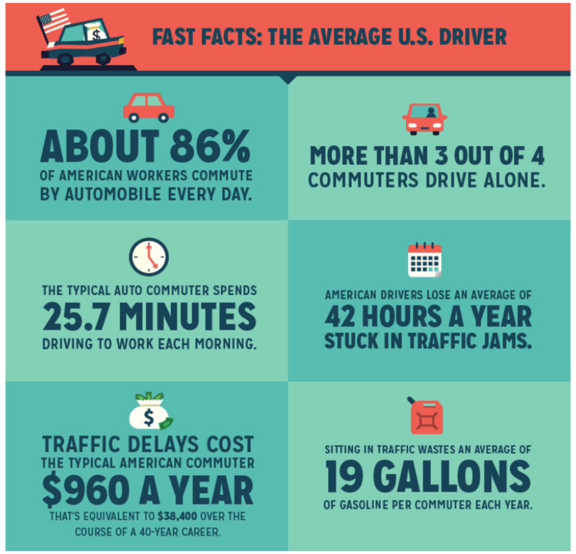
Traffic Problems – Cities' headache

Traffic Signal Control

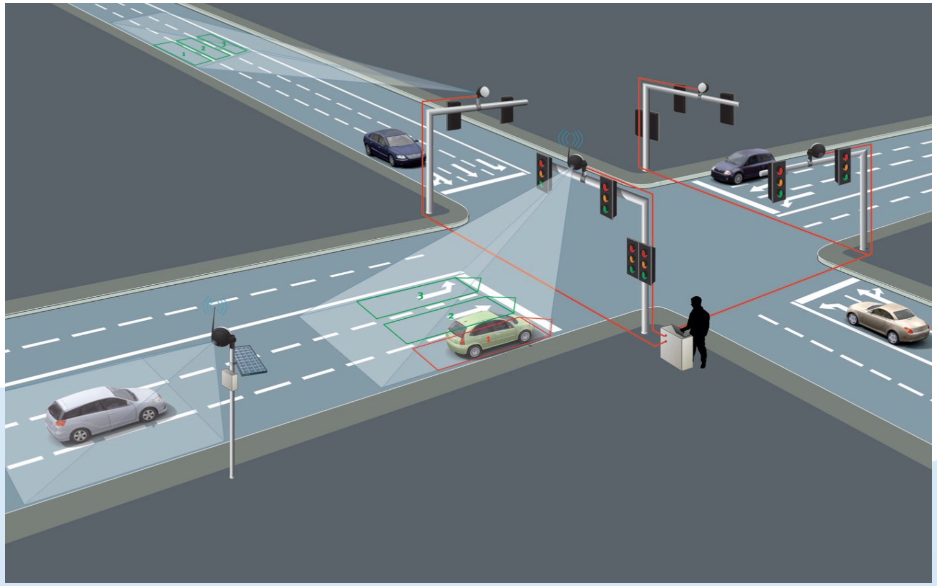


AutoInsurance: cnbc.com 2022

Traffic Signal Control

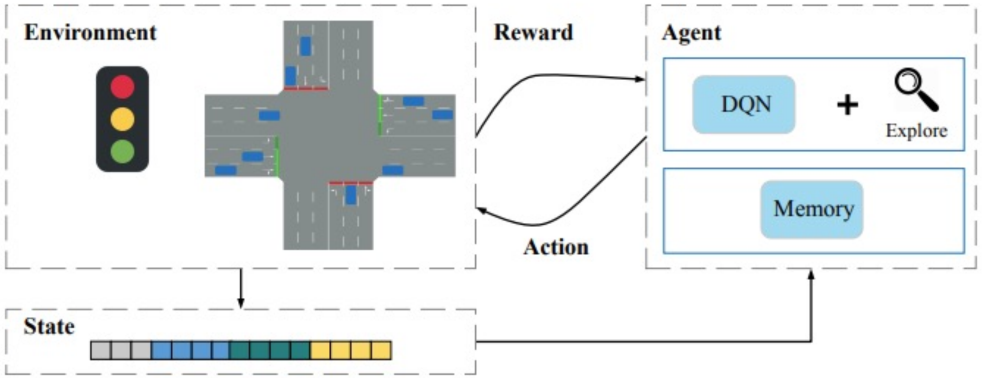


AutoInsurance: cnbc.com

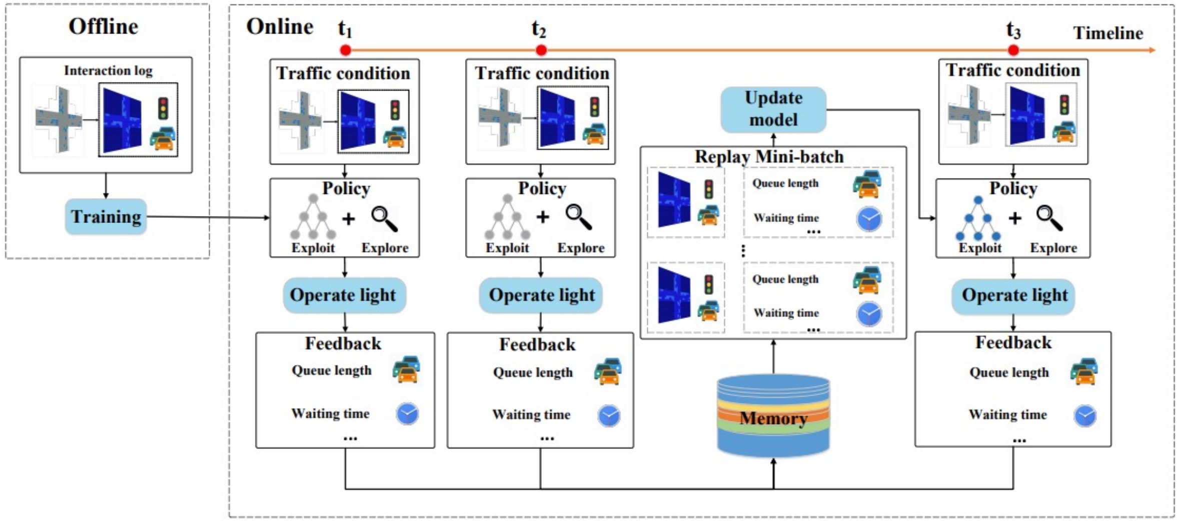


1. Rule-Based Methods
2. Optimization-Based Methods
3. **Reinforcement Learning Based Methods**

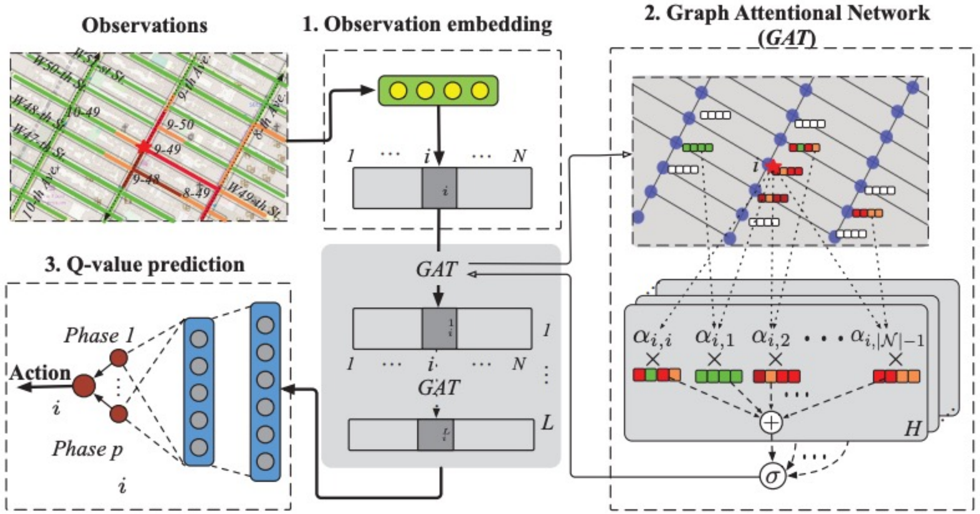
Traffic Signal Control by RL



Phases



Single intersection



Multi intersection

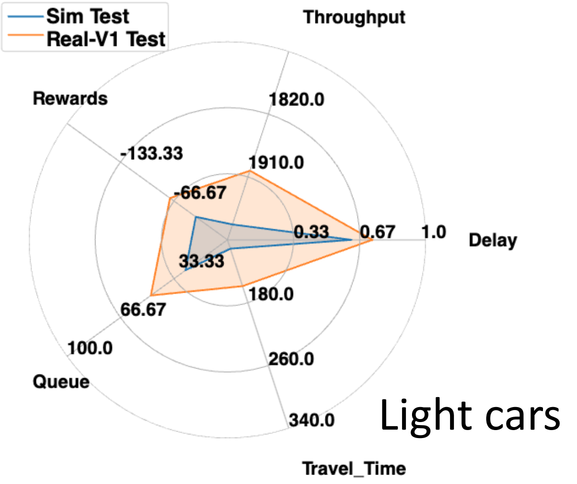
Traffic Signal Control by RL

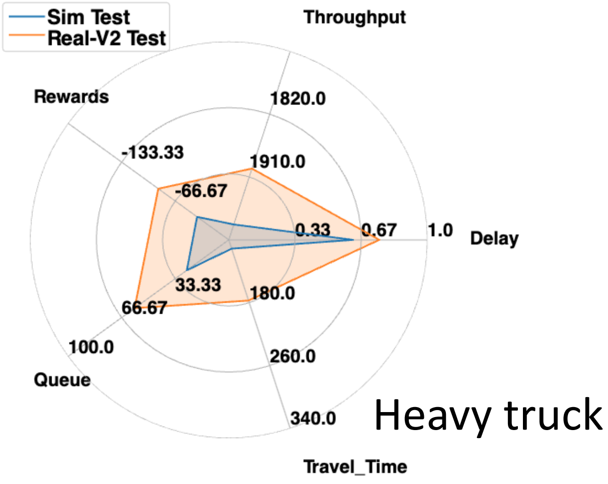
Very **Promising** in Simulation Results!

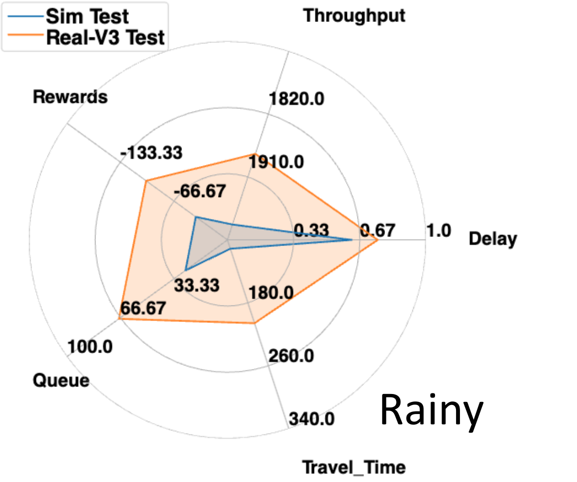
Deploy To Real World?

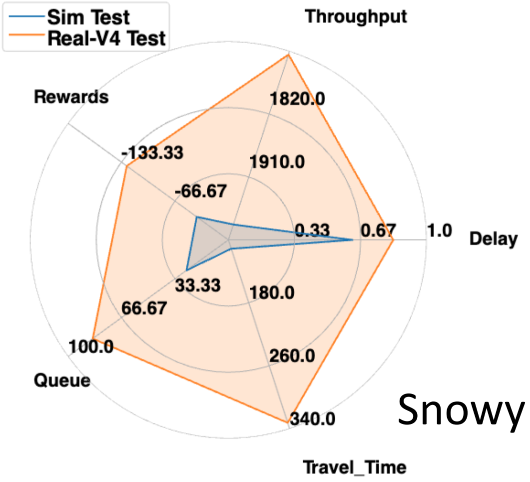
Not really ideal

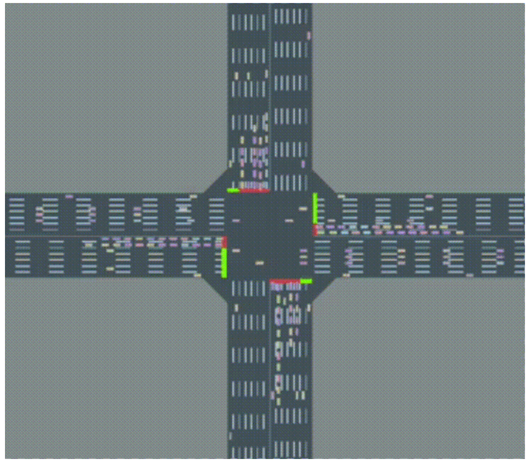
Explore the Sim2Real problem in TSC









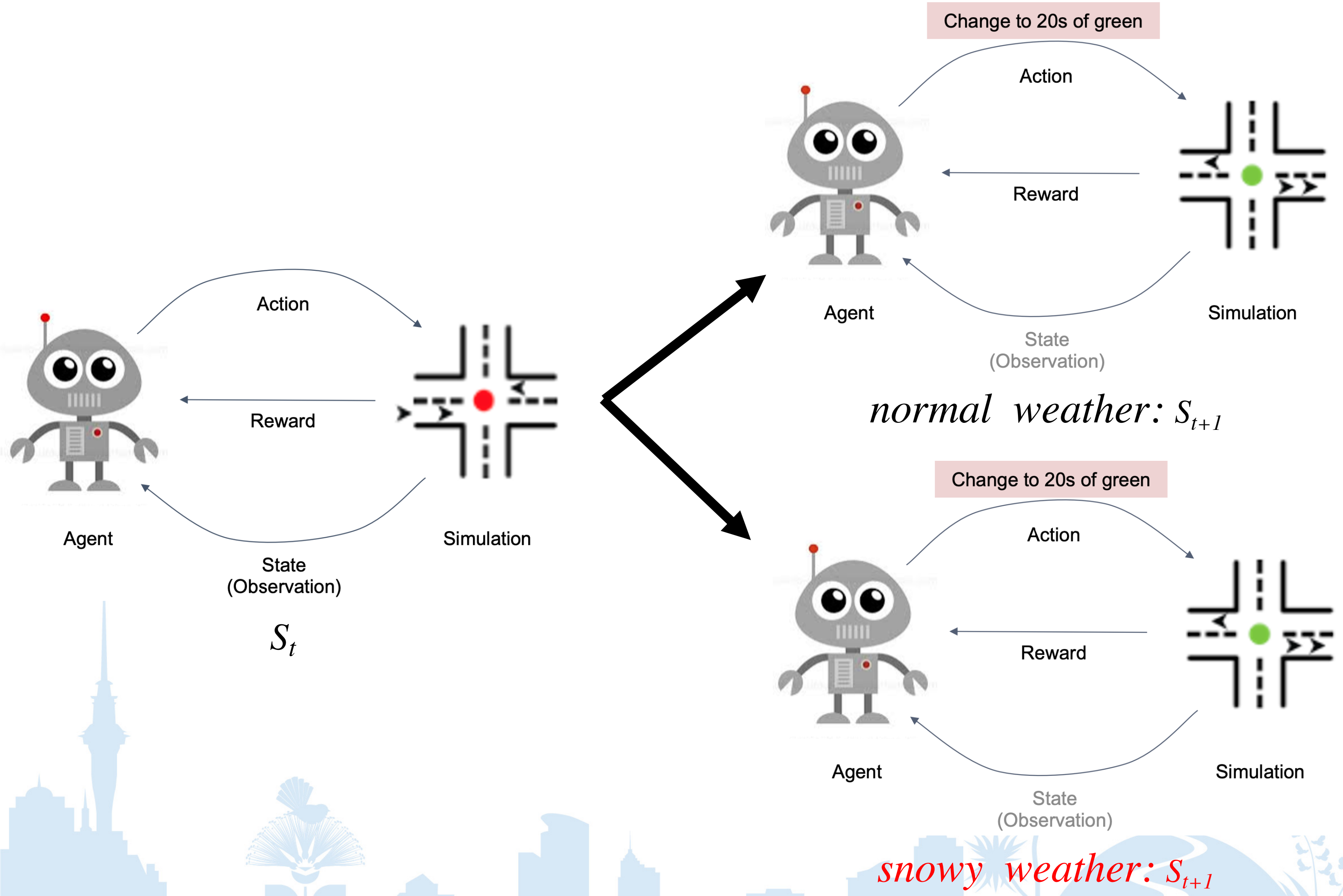


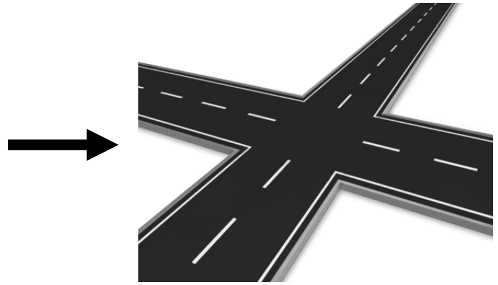
Simulator → Policy → Real Setting

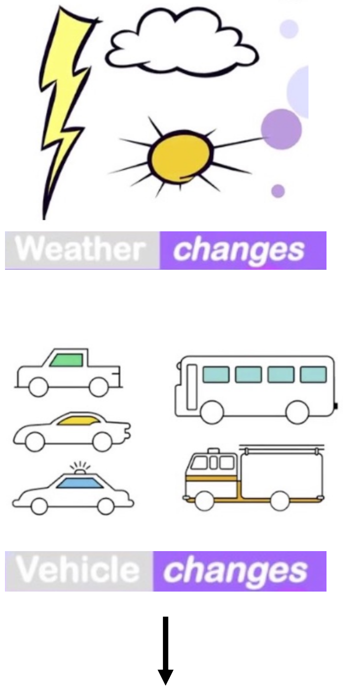
TP | ATT | Queue | Delay | Reward

Why is there a Sim2Real gap?

Analysis and Reasoning



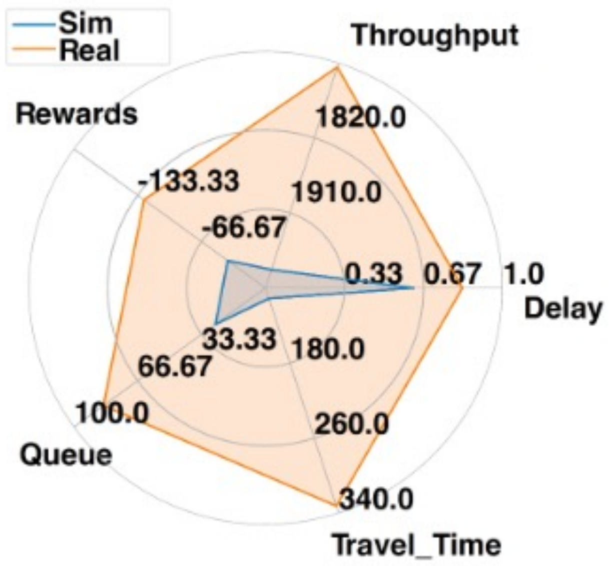


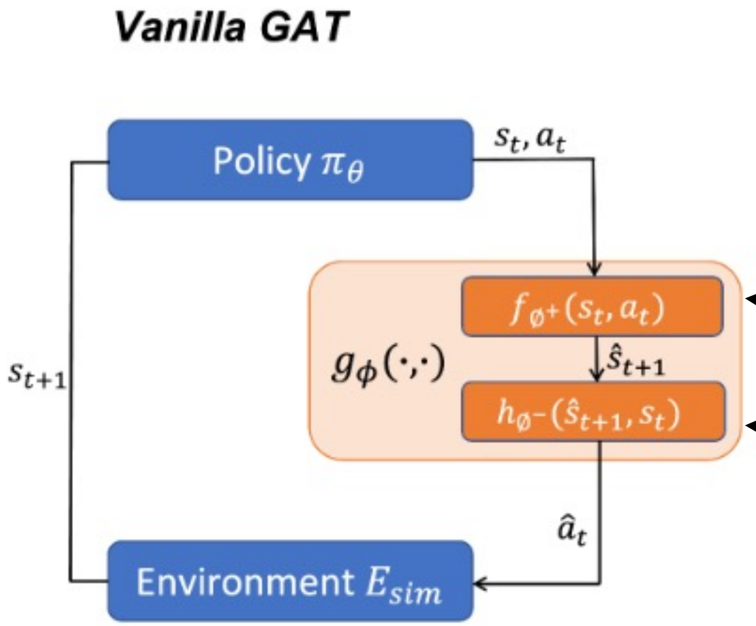


Dynamics change

The same action at a same state in sim and real
does not lead to the same next states !!!

Grounded Action Transformation (GAT)





Key idea

- Learn the policy using the data generated in simulation
- Even though the data is generated in simulation, the dynamic is similar to the real-world

- Forward model
- Inverse model

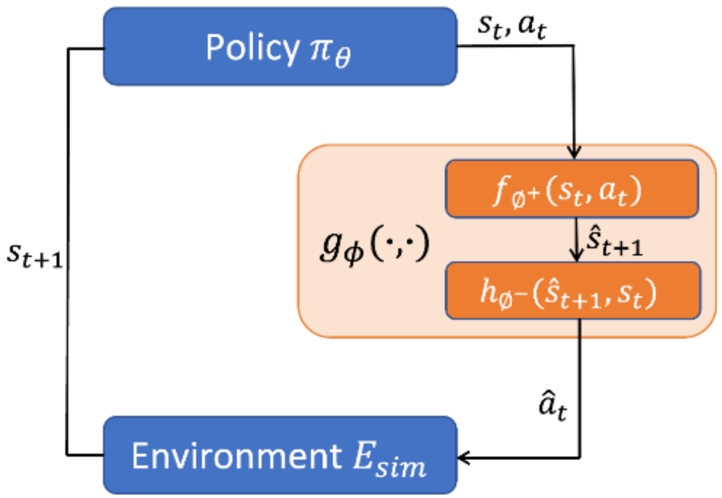
Two key models in GAT

But g_ϕ is likely to have high model uncertainty, may enlarge the performance gap!

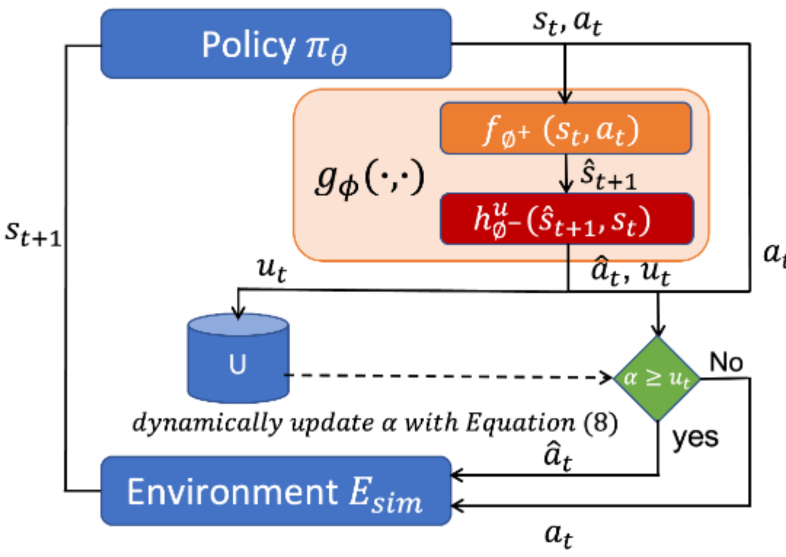
- Early exploration stages in simulation
- Unseen state and action inputs with limited real world data

Uncertainty-aware GAT

Vanilla GAT



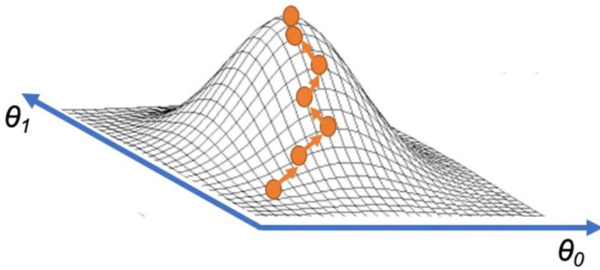
UGAT



$$\phi^* = \arg\min_{\phi} \sum_{\tau^i \in \mathcal{D}_{real}} \sum_{t=0}^{T-1} d(P^*(s^i_{t+1}|s^i_t, a^i_t), P_\phi(s^i_{t+1}|s^i_t, a^i_t))$$

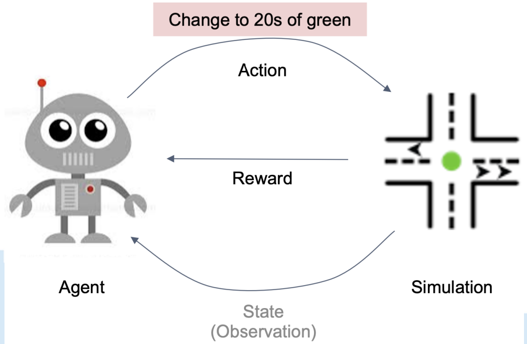
$$\hat{a}_t, u_t = g_\phi(s_t, a_t) = h_{\phi^-}(f_{\phi^+}(s_t, a_t), s_t)$$

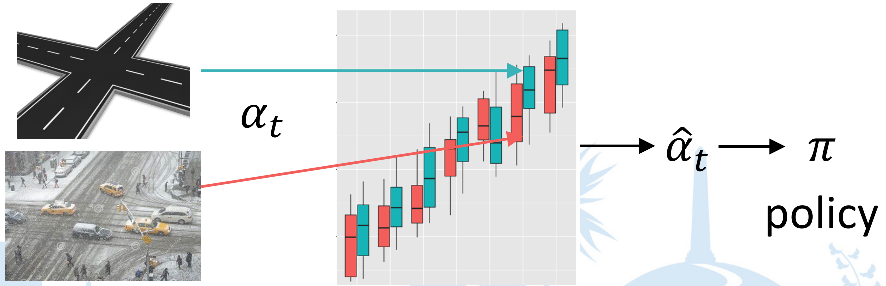
$$\alpha = \frac{\sum_{e=1}^{E} \sum_{t=0}^{T-1} u^e_t}{T \times E}$$



$$\hat{s}_{t+1} = f_{\phi^+}(s_t, a_t)$$

$$\hat{a}_t = h_{\phi^-}(\hat{s}_{t+1}, s_t)$$





- $g_\emptyset$ outputs an additional uncertainty u_t
- dynamic control by adjusting param α

Result analysis

Setting	Direct Transfer					UGAT				
	$ATT(\Delta\downarrow)$	$TP(\Delta\uparrow)$	$Reward(\Delta\uparrow)$	$Queue(\Delta\downarrow)$	$Delay(\Delta\downarrow)$	$ATT(\Delta\downarrow)$	$TP(\Delta\uparrow)$	$Reward(\Delta\uparrow)$	$Queue(\Delta\downarrow)$	$Delay(\Delta\downarrow)$
V1	158.93(47.69)	1901(-77)	-71.55(-32.11)	47.71(21.59)	0.73(0.11)	**144.72(33.49)**$_{\pm 3.61}$	**1925(-52)**$_{\pm 4.58}$	**-59.38(-19.94)**$_{\pm 3.08}$	**39.58(13.47)**$_{\pm 2.04}$	**0.67(0.05)**$_{\pm 0.01}$
V2	177.27(66.03)	1898(-80)	-87.71(-48.27)	58.59(32.47)	0.76(0.14)	**164.65(53.52)**$_{\pm 12.94}$	**1907(-71)**$_{\pm 13.06}$	**-75.18(-35.74)**$_{\pm 8.37}$	**50.25(24.14)**$_{\pm 5.56}$	**0.72(0.10)**$_{\pm 0.01}$
V3	205.86(94.63)	1877(-101)	-101.26(-61.82)	67.62(41.51)	0.76(0.14)	**183.22(71.99)**$_{\pm 13.22}$	**1900(-78)**$_{\pm 13.08}$	**-82.38(-42.94)**$_{\pm 9.11}$	**55.05(28.94)**$_{\pm 6.08}$	**0.72(0.10)**$_{\pm 0.01}$
V4	332.48(221.25)	1735(-252)	-126.71(-87.23)	84.53(58.42)	0.83(0.21)	**284.26(173.03)**$_{\pm 6.67}$	**1794(-184)**$_{\pm 12.05}$	**-111.68(-72.24)**$_{\pm 7.25}$	**74.54(48.43)**$_{\pm 4.82}$	**0.8(0.18)**$_{\pm 0.01}$

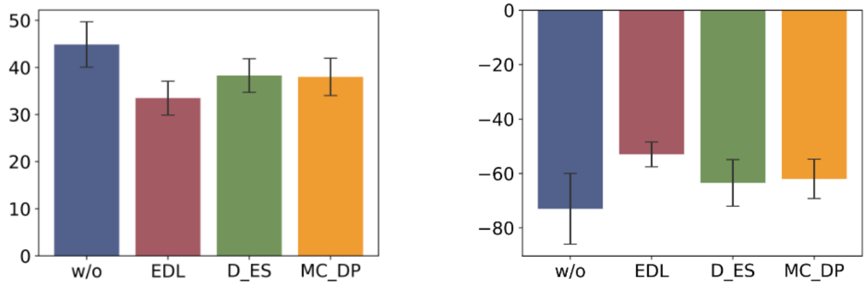
Comparison between using Direct-Transfer and UGAT

α	ATT_Δ $(\Delta\downarrow)$	TP_Δ $(\Delta\uparrow)$	$Reward_\Delta$ $(\Delta\uparrow)$	$Queue_\Delta$ $(\Delta\downarrow)$	$Delay_\Delta$ $(\Delta\downarrow)$
dynamic	**33.49**$_{\pm 3.61}$	**-52**$_{\pm 4.58}$	**-19.94**$_{\pm 3.08}$	**13.47**$_{\pm 2.04}$	**0.05**$_{\pm 0.01}$
0.2	68.59$_{\pm 7.14}$	-117$_{\pm 12.53}$	-40.42$_{\pm 3.92}$	27.11$_{\pm 4.29}$	0.12$_{\pm 0.05}$
0.4	55.87$_{\pm 7.83}$	-73$_{\pm 13.01}$	-30.69$_{\pm 4.54}$	20.30$_{\pm 3.28}$	0.12$_{\pm 0.01}$
0.5	39.12$_{\pm 4.21}$	-72$_{\pm 7.61}$	-25.07$_{\pm 5.71}$	16.88$_{\pm 5.11}$	0.08$_{\pm 0.01}$
0.6	47.09$_{\pm 2.79}$	-77$_{\pm 4.68}$	-34.11$_{\pm 3.99}$	21.31$_{\pm 2.38}$	0.10$_{\pm 0.03}$
0.8	48.53$_{\pm 6.70}$	-85$_{\pm 9.17}$	-37.85$_{\pm 6.23}$	25.60$_{\pm 2.91}$	0.11$_{\pm 0.01}$

Dynamic vs Static Control Grounding Action

Structure	ATT_Δ $(\Delta\downarrow)$	TP_Δ $(\Delta\uparrow)$	$Reward_\Delta$ $(\Delta\uparrow)$	$Queue_\Delta$ $(\Delta\downarrow)$	$Delay_\Delta$ $(\Delta\downarrow)$
UGAT	**33.49**$_{\pm 3.61}$	**-52**$_{\pm 4.58}$	**-19.94**$_{\pm 3.08}$	**13.47**$_{\pm 2.04}$	**0.05**$_{\pm 0.01}$
w/o dynamic α	39.12$_{\pm 4.21}$	-72$_{\pm 7.61}$	-25.07$_{\pm 5.71}$	16.88$_{\pm 5.11}$	0.08$_{\pm 0.01}$
w/o α, uncertainty	44.87$_{\pm 4.81}$	-73$_{\pm 12.99}$	-30.59$_{\pm 3.80}$	20.50$_{\pm 1.97}$	0.09$_{\pm 0.01}$
w/o Grounding	47.71$_{\pm 6.73}$	-77$_{\pm 10.64}$	-32.11$_{\pm 4.24}$	21.60$_{\pm 3.12}$	0.11$_{\pm 0.02}$

Ablation study



(a) ATT_Δ of 4 methods (b) TP_Δ of 4 methods

Evidential Deep Learning Module to capture the model uncertainty

Uncertainty investigation across 4 methods (EDL best)

Take away

- While RL works for decision making, the sim2real gap exists
- Grounded action transformation is a possible solution to Sim2Real
  - But only conduct grounding when model is certain

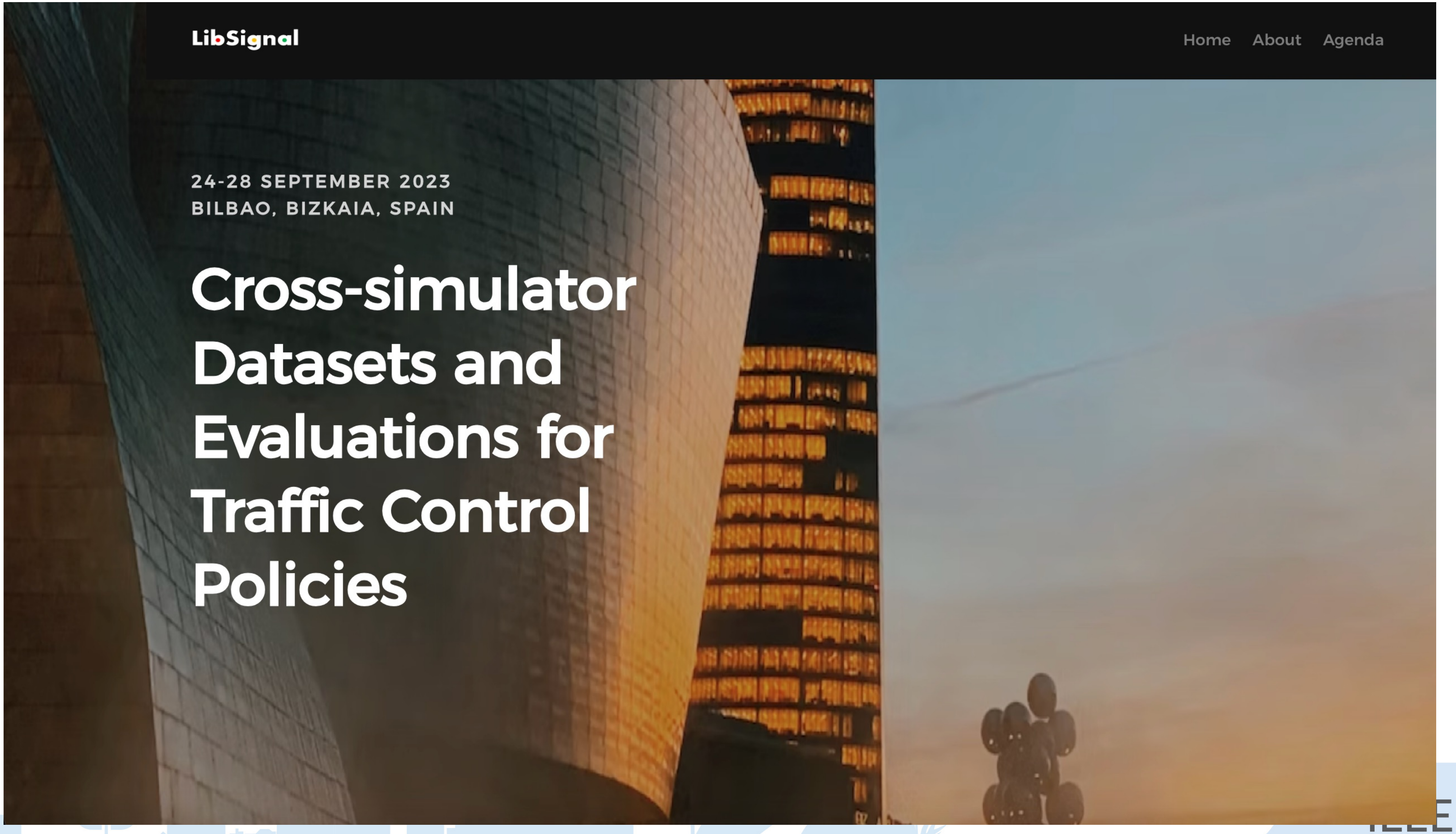
Future work for mitigate Sim2Real problems:

1) A more realistic simulator by learning internal transitions using real world data
2) Same simulator, but how to better rectify policy actions to learn a transferrable policy
3) Deploy to real world for testing the effectiveness

Uncertainty-aware Grounded Action Transformation towards Sim-to-Real Transfer for Traffic Signal Control

IEEE CDC 2023 [Singapore] & arXiv

Tutorial on the corss-simulator analysis and experiment IEEE-ITSC – CBLab & LibSignal [Spain]


LibSignal
Home
About
Agenda
24-28 SEPTEMBER 2023
BILBAO, BIZKAIA, SPAIN
Cross-simulator Datasets and Evaluations for Traffic Control Policies

Reference

[1] IntelliLight: A Reinforcement Learning Approach for Intelligent Traffic Light Control H Wei, G Zheng, et,al.
[2] Colight: Learning network-level cooperation for traffic signal control H Wei, N Xu, et,al.
[3] Grounded action transformation for robot learning in simulation Hanna J, Stone P.
[4] Evidential deep learning to quantify classification uncertainty. M Sensoy, L Kaplan
[5] Intellilight: A reinforcement learning approach for intelligent traffic light control H Wei, G Zheng
[6] A survey on traffic signal control methodsH Wei, G Zheng
[7] Toward A Thousand Lights: Decentralized Deep Reinforcement Learning for Large-Scale Traffic Signal ControlChacha Chen, Hua Wei
[8] Modeling Network-level Traffic Flow Transitions on Sparse DataX Lei, H Mei,
[9] Rebuilding city-wide traffic origin destination from road speed data G Zheng, C Liu
[10] Reinforcement Learning Approaches for Traffic Signal Control under Missing DataH Mei, J Li
[11] AutoInsurance: cnbc.com
[12] NewYork times report on Traffic Congestion and waste on resources