NEURAL INFORMATION PROCESSING SYSTEMS

URB

Urban Routing Benchmark

for RL-equipped Connected Autonomous Vehicles

Ahmet Onur Akman*[1], Anastasia Psarou[1], Michał Hoffmann[1], Łukasz Gorczyca[1], Łukasz Kowalski[2], Paweł Gora[1], Grzegorz Jamróz[1], Rafał Kucharski[1]

[1] Jagiellonian University, Kraków, Poland

[2] Institute of Urban and Regional Development, Warsaw, Poland

JAGIELLONIAN UNIVERSITY IN KRAKÓW

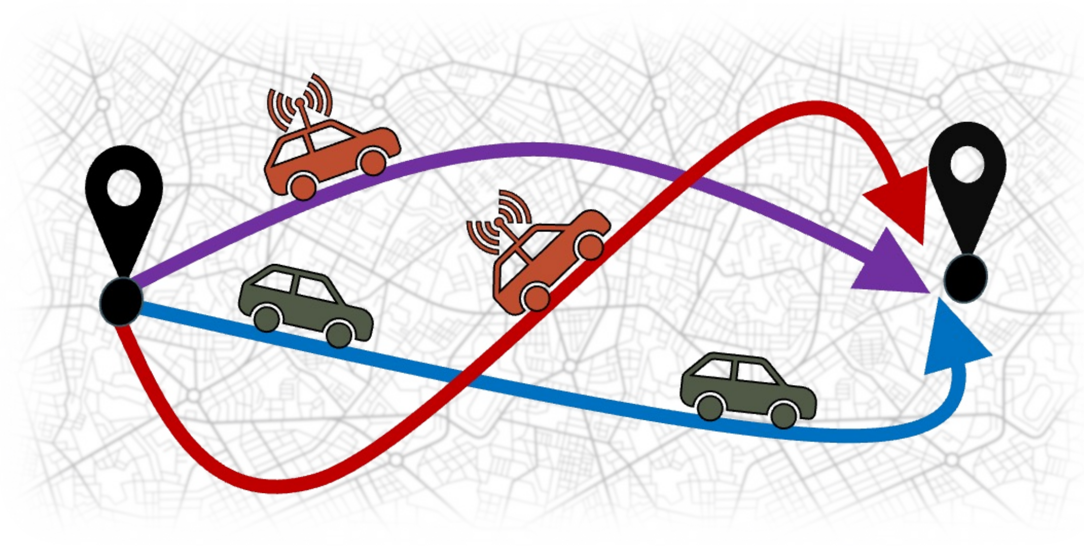
COeXISTENCE

erc Starting Grants

group of machine gmum learning research

Background

- We consider future urban systems where limited resources are shared by **human drivers and CAVs**.
- We study the urban route choice problem: **Which route to take to reach my destination?**



Background

- A possible future scenario:

In city traffic, what happens if ~40% of drivers delegate their routing to a shared policy?

- Which raises the following questions:
  - Which algorithm is most suitable for collective urban fleet routing?
  - How does the problem scale with network complexity, population size, and planning horizon?
  - What is the impact on: humans, congestion, speed, mileage, and equity?

URB

- URB is a comprehensive **benchmarking framework**, aiming to **standardize the assessment** for CAV routing solutions.
- It unifies evaluation across **real-world-inspired tasks**.
- It comes with a catalog of **implementations**, **baselines**, and domain-specific **KPIs**.

URB - Dataset

29 real-world networks

A diverse set of scenarios to test solution robustness under different complexities.

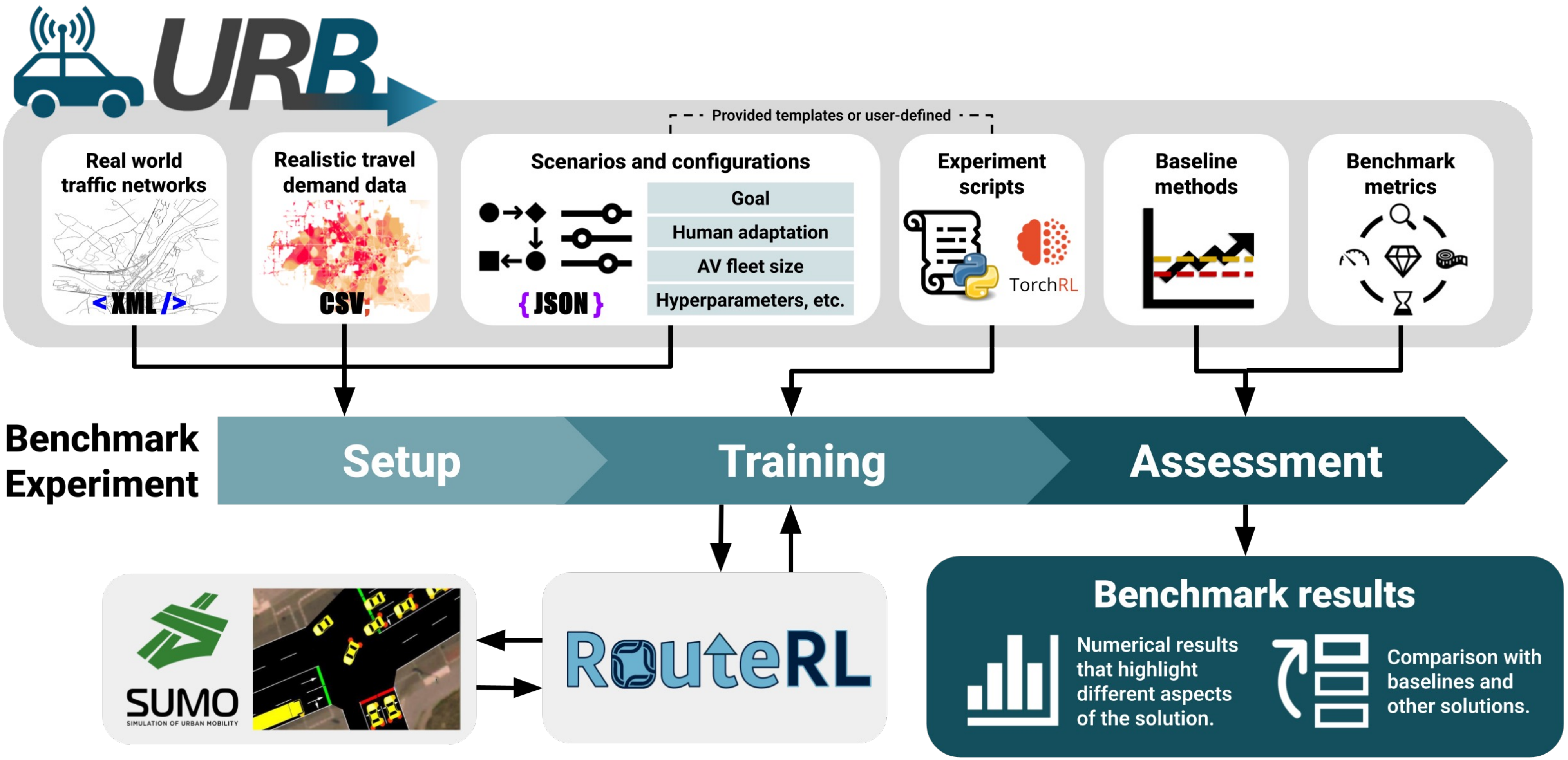
Realistic demand data

Each network is bundled with realistic demand data for assessment in realistic scenarios.

Open access

The dataset and its generation code are documented and made publicly available.

URB - Framework

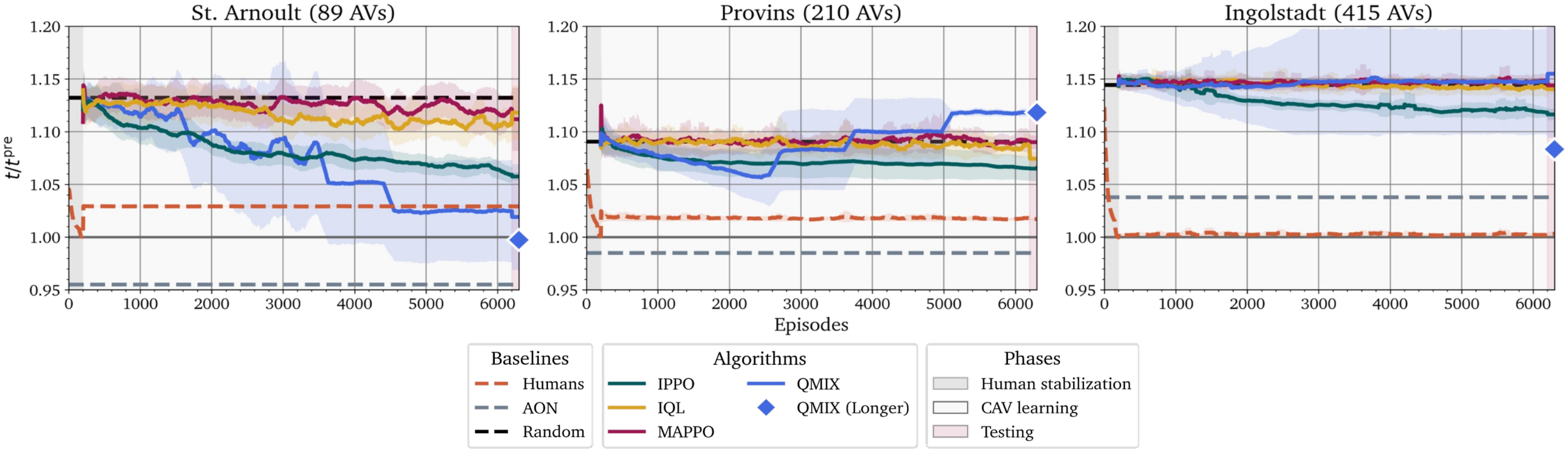


Scenario 1: Mixed Autonomy

*In a given network with a **fixed demand** pattern, **human drivers** have stabilized their route-choice strategies to minimize travel times.*

*After a **40% share of drivers switch to CAVs** and delegate routing decisions, the **CAVs use MARL** to develop new strategies to reduce delay.*

Scenario 1: Results



200, 6000, and 100 episodes for **human stabilization, CAV training,** and **policy testing**, respectively.

We test **IPPO, IQL, QMIX, and MAPPO** against the URB baselines in **3 networks**.

Many algorithms hardly beat the random baseline.

Only QMIX in St. Arnoult managed to outperform humans, though not consistently.

Scenario 1: Results

		t^{TEST}	t_{CAV}	t_{HDV}	c_{ALL}	c_{HDV}	c_{CAV}	Δ_{V}	Δ_{L}	WR
St. Arnoult	IPPO	3.28 (0.004)	3.33 (0.013)	3.25 (0.008)	0.63 (0.015)	0.13 (0.004)	1.38 (0.034)	-0.24 (0.067)	0.06 (0.004)	0%
	IQL	3.36 (0.040)	3.53 (0.104)	3.24 (0.005)	0.66 (0.000)	0.14 (0.000)	1.44 (0.004)	-0.37 (0.115)	0.09 (0.021)	0%
	MAPPO	3.35 (0.049)	3.51 (0.121)	3.25 (0.004)	0.66 (0.000)	0.14 (0.004)	1.45 (0.000)	-0.27 (0.129)	0.09 (0.019)	0%
	QMIX	3.24 (0.080)	3.21 (0.206)	3.25 (0.004)	0.65 (0.004)	0.14 (0.005)	1.43 (0.005)	-0.22 (0.034)	0.03 (0.040)	80%
	Human	**3.15**	N/A	**3.15**	N/A	N/A	N/A	**0.00**	**0.00**	**100%**
	AON	**3.15**	**3.01**	3.25	**0.55**	**0.09**	**1.21**	-0.06	**0.00**	**100%**
	Random	3.38	3.58	3.25	0.60	**0.09**	1.36	-0.33	0.10	0%
Provins	IPPO	2.90 (0.015)	2.98 (0.040)	2.85 (0.004)	0.61 (0.271)	0.31 (0.217)	1.05 (0.356)	-0.52 (0.080)	0.05 (0.009)	0%
	IQL	2.91 (0.011)	3.01 (0.027)	2.85 (0.008)	1.40 (0.104)	0.92 (0.068)	2.12 (0.183)	-0.58 (0.093)	0.05 (0.007)	0%
	MAPPO	2.93 (0.011)	3.05 (0.024)	2.84 (0.005)	1.29 (0.162)	0.83 (0.110)	2.00 (0.247)	-0.69 (0.038)	0.06 (0.004)	0%
	QMIX	2.96 (0.005)	3.14 (0.000)	2.85 (0.000)	0.85 (0.215)	0.52 (0.176)	1.35 (0.278)	-0.82 (0.033)	0.08 (0.000)	0%
	Human	**2.80**	N/A	**2.80**	N/A	N/A	N/A	**0.00**	**0.00**	**100%**
	AON	2.81	**2.76**	2.84	**0.47**	**0.19**	0.99	-0.14	**0.00**	**100%**
	Random	2.93	3.04	2.85	0.51	0.22	**0.95**	-0.62	0.06	0%
Ingolstadt	IPPO	4.41 (0.005)	4.71 (0.030)	**4.21** (0.023)	2.42 (0.497)	1.90 (0.505)	3.19 (0.495)	-0.52 (0.095)	0.06 (0.004)	0%
	IQL	4.46 (0.009)	4.81 (0.024)	4.23 (0.009)	2.54 (0.546)	1.93 (0.533)	3.44 (0.562)	-0.69 (0.067)	0.07 (0.000)	0%
	MAPPO	4.45 (0.011)	4.82 (0.019)	**4.21** (0.008)	2.76 (0.599)	2.16 (0.622)	3.66 (0.562)	-0.72 (0.066)	0.07 (0.004)	0%
	QMIX	4.50 (0.140)	4.87 (0.325)	4.24 (0.015)	1.83 (0.749)	1.27 (0.710)	2.67 (0.810)	-0.97 (0.235)	0.06 (0.045)	0%
	Human	**4.21**	N/A	**4.21**	N/A	N/A	N/A	**0.00**	0.00	**100%**
	AON	4.29	**4.37**	4.23	**0.87**	0.55	**0.24**	-0.45	**-0.01**	0%
	Random	4.45	4.81	4.22	0.99	**0.49**	1.74	-0.68	0.07	0%

Independent learning: Does not scale and yields sub-human performance.

Training is costly in all cases (large c_{CAV}).

Humans are disadvantaged; traffic efficiency is reduced ($t_{test} > t_{pre}$, $\Delta_V < 0$, $\Delta_L > 0$).

Scenario 1: Results

		t^{TEST}	t_{CAV}	t_{HDV}	c_{ALL}	c_{HDV}	c_{CAV}	Δ_{V}	Δ_{L}	WR
St. Arnoult	IPPO	3.28 (0.004)	3.33 (0.013)	3.25 (0.008)	0.63 (0.015)	0.13 (0.004)	1.38 (0.034)	-0.24 (0.067)	0.06 (0.004)	0%
	IQL	3.36 (0.040)	3.53 (0.104)	3.24 (0.005)	0.66 (0.000)	0.14 (0.000)	1.44 (0.004)	-0.37 (0.115)	0.09 (0.021)	0%
	MAPPO	3.35 (0.049)	3.51 (0.121)	3.25 (0.004)	0.66 (0.000)	0.14 (0.004)	1.45 (0.000)	-0.27 (0.129)	0.09 (0.019)	0%
	QMIX	3.24 (0.080)	3.21 (0.206)	3.25 (0.004)	0.65 (0.004)	0.14 (0.005)	1.43 (0.005)	-0.22 (0.034)	0.03 (0.040)	80%
	Human	**3.15**	N/A	**3.15**	N/A	N/A	N/A	**0.00**	**0.00**	**100%**
	AON	**3.15**	**3.01**	3.25	**0.55**	**0.09**	**1.21**	-0.06	**0.00**	**100%**
	Random	3.38	3.58	3.25	0.60	**0.09**	1.36	-0.33	0.10	0%
Provins	IPPO	2.90 (0.015)	2.98 (0.040)	2.85 (0.004)	0.61 (0.271)	0.31 (0.217)	1.05 (0.356)	-0.52 (0.080)	0.05 (0.009)	0%
	IQL	2.91 (0.011)	3.01 (0.027)	2.85 (0.008)	1.40 (0.104)	0.92 (0.068)	2.12 (0.183)	-0.58 (0.093)	0.05 (0.007)	0%
	MAPPO	2.93 (0.011)	3.05 (0.024)	2.84 (0.005)	1.29 (0.162)	0.83 (0.110)	2.00 (0.247)	-0.69 (0.038)	0.06 (0.004)	0%
	QMIX	2.96 (0.005)	3.14 (0.000)	2.85 (0.000)	0.85 (0.215)	0.52 (0.176)	1.35 (0.278)	-0.82 (0.033)	0.08 (0.000)	0%
	Human	**2.80**	N/A	**2.80**	N/A	N/A	N/A	**0.00**	**0.00**	**100%**
	AON	2.81	**2.76**	2.84	**0.47**	**0.19**	0.99	-0.14	**0.00**	**100%**
	Random	2.93	3.04	2.85	0.51	0.22	**0.95**	-0.62	0.06	0%
Ingolstadt	IPPO	4.41 (0.005)	4.71 (0.030)	**4.21** (0.023)	2.42 (0.497)	1.90 (0.505)	3.19 (0.495)	-0.52 (0.095)	0.06 (0.004)	0%
	IQL	4.46 (0.009)	4.81 (0.024)	4.23 (0.009)	2.54 (0.546)	1.93 (0.533)	3.44 (0.562)	-0.69 (0.067)	0.07 (0.000)	0%
	MAPPO	4.45 (0.011)	4.82 (0.019)	**4.21** (0.008)	2.76 (0.599)	2.16 (0.622)	3.66 (0.562)	-0.72 (0.066)	0.07 (0.004)	0%
	QMIX	4.50 (0.140)	4.87 (0.325)	4.24 (0.015)	1.83 (0.749)	1.27 (0.710)	2.67 (0.810)	-0.97 (0.235)	0.06 (0.045)	0%
	Human	**4.21**	N/A	**4.21**	N/A	N/A	N/A	**0.00**	0.00	**100%**
	AON	4.29	**4.37**	4.23	**0.87**	0.55	**0.24**	-0.45	**-0.01**	0%
	Random	4.45	4.81	4.22	0.99	**0.49**	1.74	-0.68	0.07	0%

Independent learning: Does not scale and yields sub-human performance.

Training is costly in all cases (large c_{CAV}).

Humans are disadvantaged; traffic efficiency is reduced ($t_{test} > t_{pre}$, $\Delta_V < 0$, $\Delta_L > 0$).

Scenario 1: Results

		t^{TEST}	t_{CAV}	t_{HDV}	c_{ALL}	c_{HDV}	c_{CAV}	Δ_{V}	Δ_{L}	WR
St. Arnoult	IPPO	3.28 (0.004)	3.33 (0.013)	3.25 (0.008)	0.63 (0.015)	0.13 (0.004)	1.38 (0.034)	-0.24 (0.067)	0.06 (0.004)	0%
	IQL	3.36 (0.040)	3.53 (0.104)	3.24 (0.005)	0.66 (0.000)	0.14 (0.000)	1.44 (0.004)	-0.37 (0.115)	0.09 (0.021)	0%
	MAPPO	3.35 (0.049)	3.51 (0.121)	3.25 (0.004)	0.66 (0.000)	0.14 (0.004)	1.45 (0.000)	-0.27 (0.129)	0.09 (0.019)	0%
	QMIX	3.24 (0.080)	3.21 (0.206)	3.25 (0.004)	0.65 (0.004)	0.14 (0.005)	1.43 (0.005)	-0.22 (0.034)	0.03 (0.040)	80%
	Human	**3.15**	N/A	**3.15**	N/A	N/A	N/A	**0.00**	**0.00**	**100%**
	AON	**3.15**	**3.01**	3.25	**0.55**	**0.09**	**1.21**	-0.06	**0.00**	**100%**
	Random	3.38	3.58	3.25	0.60	**0.09**	1.36	-0.33	0.10	0%
Provins	IPPO	2.90 (0.015)	2.98 (0.040)	2.85 (0.004)	0.61 (0.271)	0.31 (0.217)	1.05 (0.356)	-0.52 (0.080)	0.05 (0.009)	0%
	IQL	2.91 (0.011)	3.01 (0.027)	2.85 (0.008)	1.40 (0.104)	0.92 (0.068)	2.12 (0.183)	-0.58 (0.093)	0.05 (0.007)	0%
	MAPPO	2.93 (0.011)	3.05 (0.024)	2.84 (0.005)	1.29 (0.162)	0.83 (0.110)	2.00 (0.247)	-0.69 (0.038)	0.06 (0.004)	0%
	QMIX	2.96 (0.005)	3.14 (0.000)	2.85 (0.000)	0.85 (0.215)	0.52 (0.176)	1.35 (0.278)	-0.82 (0.033)	0.08 (0.000)	0%
	Human	**2.80**	N/A	**2.80**	N/A	N/A	N/A	**0.00**	**0.00**	**100%**
	AON	2.81	**2.76**	2.84	**0.47**	**0.19**	0.99	-0.14	**0.00**	**100%**
	Random	2.93	3.04	2.85	0.51	0.22	**0.95**	-0.62	0.06	0%
Ingolstadt	IPPO	4.41 (0.005)	4.71 (0.030)	**4.21** (0.023)	2.42 (0.497)	1.90 (0.505)	3.19 (0.495)	-0.52 (0.095)	0.06 (0.004)	0%
	IQL	4.46 (0.009)	4.81 (0.024)	4.23 (0.009)	2.54 (0.546)	1.93 (0.533)	3.44 (0.562)	-0.69 (0.067)	0.07 (0.000)	0%
	MAPPO	4.45 (0.011)	4.82 (0.019)	**4.21** (0.008)	2.76 (0.599)	2.16 (0.622)	3.66 (0.562)	-0.72 (0.066)	0.07 (0.004)	0%
	QMIX	4.50 (0.140)	4.87 (0.325)	4.24 (0.015)	1.83 (0.749)	1.27 (0.710)	2.67 (0.810)	-0.97 (0.235)	0.06 (0.045)	0%
	Human	**4.21**	N/A	**4.21**	N/A	N/A	N/A	**0.00**	0.00	**100%**
	AON	4.29	**4.37**	4.23	**0.87**	0.55	**0.24**	-0.45	**-0.01**	0%
	Random	4.45	4.81	4.22	0.99	**0.49**	1.74	-0.68	0.07	0%

Independent learning: Does not scale and yields sub-human performance.

Training is costly in all cases (large c_{CAV}).

Humans are disadvantaged; traffic efficiency is reduced ($t_{test} > t_{pre}, \Delta_V < 0, \Delta_L > 0$).

Scenario 1: Results

		t^{TEST}	t_{CAV}	t_{HDV}	c_{ALL}	c_{HDV}	c_{CAV}	Δ_{V}	Δ_{L}	WR
ST. ARNOULT	IPPO	3.28 (0.004)	3.33 (0.013)	3.25 (0.008)	0.63 (0.015)	0.13 (0.004)	1.38 (0.034)	-0.24 (0.067)	0.06 (0.004)	0%
	IQL	3.36 (0.040)	3.53 (0.104)	3.24 (0.005)	0.66 (0.000)	0.14 (0.000)	1.44 (0.004)	-0.37 (0.115)	0.09 (0.021)	0%
	MAPPO	3.35 (0.049)	3.51 (0.121)	3.25 (0.004)	0.66 (0.000)	0.14 (0.004)	1.45 (0.000)	-0.27 (0.129)	0.09 (0.019)	0%
	QMIX	3.24 (0.080)	3.21 (0.206)	3.25 (0.004)	0.65 (0.004)	0.14 (0.005)	1.43 (0.005)	-0.22 (0.034)	0.03 (0.040)	80%
	HUMAN	**3.15**	N/A	**3.15**	N/A	N/A	N/A	**0.00**	**0.00**	**100%**
	AON	**3.15**	**3.01**	3.25	**0.55**	**0.09**	**1.21**	-0.06	**0.00**	**100%**
	RANDOM	3.38	3.58	3.25	0.60	**0.09**	1.36	-0.33	0.10	0%
PROVINS	IPPO	2.90 (0.015)	2.98 (0.040)	2.85 (0.004)	0.61 (0.271)	0.31 (0.217)	1.05 (0.356)	-0.52 (0.080)	0.05 (0.009)	0%
	IQL	2.91 (0.011)	3.01 (0.027)	2.85 (0.008)	1.40 (0.104)	0.92 (0.068)	2.12 (0.183)	-0.58 (0.093)	0.05 (0.007)	0%
	MAPPO	2.93 (0.011)	3.05 (0.024)	2.84 (0.005)	1.29 (0.162)	0.83 (0.110)	2.00 (0.247)	-0.69 (0.038)	0.06 (0.004)	0%
	QMIX	2.96 (0.005)	3.14 (0.000)	2.85 (0.000)	0.85 (0.215)	0.52 (0.176)	1.35 (0.278)	-0.82 (0.033)	0.08 (0.000)	0%
	HUMAN	**2.80**	N/A	**2.80**	N/A	N/A	N/A	**0.00**	**0.00**	**100%**
	AON	2.81	**2.76**	2.84	**0.47**	**0.19**	0.99	-0.14	**0.00**	**100%**
	RANDOM	2.93	3.04	2.85	0.51	0.22	**0.95**	-0.62	0.06	0%
INGOLSTADT	IPPO	4.41 (0.005)	4.71 (0.030)	**4.21** (0.023)	2.42 (0.497)	1.90 (0.505)	3.19 (0.495)	-0.52 (0.095)	0.06 (0.004)	0%
	IQL	4.46 (0.009)	4.81 (0.024)	4.23 (0.009)	2.54 (0.546)	1.93 (0.533)	3.44 (0.562)	-0.69 (0.067)	0.07 (0.000)	0%
	MAPPO	4.45 (0.011)	4.82 (0.019)	**4.21** (0.008)	2.76 (0.599)	2.16 (0.622)	3.66 (0.562)	-0.72 (0.066)	0.07 (0.004)	0%
	QMIX	4.50 (0.140)	4.87 (0.325)	4.24 (0.015)	1.83 (0.749)	1.27 (0.710)	2.67 (0.810)	-0.97 (0.235)	0.06 (0.045)	0%
	HUMAN	**4.21**	N/A	**4.21**	N/A	N/A	N/A	**0.00**	0.00	**100%**
	AON	4.29	**4.37**	4.23	**0.87**	0.55	**0.24**	-0.45	**-0.01**	0%
	RANDOM	4.45	4.81	4.22	0.99	**0.49**	1.74	-0.68	0.07	0%

Independent learning: Does not scale and yields sub-human performance.

Training is costly in all cases (large c_{CAV}).

Humans are disadvantaged; traffic efficiency is reduced ($t_{test} > t_{pre}$, $\Delta_V < 0$, $\Delta_L > 0$).

Algorithmic Performance:
Sensitivity Analysis

- Using **Scenario 1** in **St. Arnoult**, we test different demand levels and information restrictions.
- **URB exposes the realism gap**: large populations, restricted information, and credit assignment **remain open challenges for practical applications**.

Experiment	t^{PRE}	t_{CAV}	$\Delta_{\%} t^{\mathrm{PRE}}$
IQL (Half demand)	3.27	3.45	-5.50%
IQL (Original)	3.15	3.53	-12.06%
IQL (Double demand)	3.24	5.81	-79.32%
IQL (Global observations)	3.15	3.26	-3.49%
MAPPO (Half demand)	3.27	3.44	-5.20%
MAPPO (Original)	3.15	3.45	-9.52%
MAPPO (Double demand)	3.24	5.23	-61.42%

Summary

- We introduce **URB: an open-source, realistic,** and **comprehensive MARL benchmark** for large-scale urban routing.
- URB addresses the **limited social awareness and practicality** often found in RL research benchmarks.
- We contribute an **extensive urban routing dataset** for testing and assessment using URB metrics.
- We initiate the **URB leaderboard** by evaluating widely used algorithms and empirically highlighting their practical shortcomings.

Thank you!

Contact:

Ahmet Onur Akman

onur.akman@uj.edu.pl

Dataset

Code

JAGIELLONIAN UNIVERSITY IN KRAKÓW

COeXISTENCE

erc Starting Grants

group of machine gmum learning research

Research financed by the European Union within the Horizon Europe Framework Programme (ERC Starting Grant COeXISTENCE no. 101075838).