

A APPENDIX

A.1 IMPLEMENTATION DETAILS

Intelligent Driver Model (IDM). IDM is given by Equation 9 and Equation 10. The model describes the acceleration \dot{v}_{back} of the back agent, as a function of the agent’s velocity v_{back} , the reference velocity v_0 , the difference between the agent velocity and the velocity of the agent in front $\Delta v = v_{back} - v_{front}$, and the following distance $\varphi = s_{front} + L_{length,front} - s_{back}$. Here, s_{front} is the position of the front agent, s_{back} denotes the position of the back agent, and $L_{length,front}$ denotes the length of the front agent. The physical interpretation of the parameters are the minimum following time, T , the minimum following gap, s_0 , the maximum acceleration, a , the minimum following gap, s_0 , the maximum acceleration, a , and the comfortable braking deceleration, b .

$$\dot{v}_{back} = a \left[1 - \left(\frac{v_{back}}{v_0} \right)^\delta - \left(\frac{\phi(v_{back}, \Delta v)}{\varphi} \right)^2 \right] \quad (9)$$

$$\phi(v_{back}, \Delta v) = s_0 + v_{back}T + \frac{v_{back}\Delta v}{2\sqrt{ab}} \quad (10)$$

Table 2: **Parameters of IDM.**

Parameter	Value
Desired speed v_0	6 m/s
Time gap T	1.0 s
Minimum gap s_0	2 m
Acceleration exponent δ	4
Acceleration a	5.0 m/s ²
Comfortable deceleration b	5.0 m/s ²

Policy learning parameters. For Independent Policy Learning (IPL) and single-agent reinforcement learning algorithms, we utilize Soft-Actor-Critic (SAC) (Haarnoja et al., 2018) and Adam optimizer (Kingma & Ba, 2015). Detailed parameters are shown in Table 3.

Table 3: **Hyperparameters of SAC.**

Parameter	Value
optimizer	Adam
actor learning rate	$1 \cdot 10^{-4}$
critic learning rate	$5 \cdot 10^{-4}$
tune learning rate	$1 \cdot 10^{-4}$
discount (γ)	0.9
batch size	128
replay buffer size	10^6
nonlinearity	ReLU
target smoothing coefficient (τ)	0.005
target update interval	200

A.2 RESULTS ON OUR COORDINATED TRAFFIC FLOW

More results are shown in Table 4, Figure 7, Figure 8, Figure 9, Figure 10, Figure 11, Figure 12, Figure 13, and Figure 14.

A.3 RESULTS ON ZERO-SHOT TRANSFER.

More results are shown in Table 5, Table 6, Table 7, Table 8.

Table 4: **Quantitative performance of traffic flows.** The table reports the percentage of different metrics in intersection, bottleneck, merge, and roundabout. A “†” indicates our proposed traffic flow.

Methods	Intersection					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
IDM	70.4 ± 0.0	29.6 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	42.1 ± 0.0
FLOW	79.6 ± 0.5	15.9 ± 0.6	2.0 ± 0.2	1.2 ± 0.1	0.0 ± 0.0	47.4 ± 0.2
CoPO	79.6 ± 0.3	17.6 ± 0.4	1.3 ± 0.1	1.4 ± 0.1	0.1 ± 0.1	46.8 ± 0.1
FailMaker	45.3 ± 0.3	52.5 ± 0.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	28.7 ± 0.1
Ours†	86.9 ± 0.5	9.0 ± 0.4	2.8 ± 0.1	1.2 ± 0.1	0.2 ± 0.1	51.0 ± 0.2
Methods	Bottleneck					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
IDM	67.0 ± 0.0	33.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	49.1 ± 0.0
FLOW	76.2 ± 0.5	9.8 ± 0.4	14.3 ± 0.6	0.0 ± 0.0	0.0 ± 0.0	75.1 ± 0.3
CoPO	80.3 ± 0.6	9.3 ± 0.7	11.4 ± 0.4	0.0 ± 0.0	0.0 ± 0.0	74.5 ± 0.3
FailMaker	21.3 ± 0.2	78.6 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	29.0 ± 0.2
Ours†	83.4 ± 0.4	9.4 ± 0.3	7.3 ± 0.3	0.0 ± 0.0	0.0 ± 0.0	76.4 ± 0.1
Methods	Merge					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
IDM	60.0 ± 0.0	40.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	46.9 ± 0.0
FLOW	66.2 ± 0.4	25.4 ± 0.5	8.5 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	55.1 ± 0.2
CoPO	69.3 ± 0.5	26.9 ± 0.6	3.8 ± 0.3	0.0 ± 0.0	0.0 ± 0.0	54.9 ± 0.2
FailMaker	16.0 ± 0.2	80.9 ± 0.2	3.4 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	16.2 ± 0.1
Ours†	83.1 ± 0.5	16.2 ± 0.5	0.6 ± 0.1	0.2 ± 0.1	0.0 ± 0.0	60.0 ± 0.2
Methods	Roundabout					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
IDM	73.6 ± 0.0	26.4 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	38.1 ± 0.0
FLOW	72.7 ± 0.6	22.4 ± 0.4	4.8 ± 0.2	0.0 ± 0.0	0.1 ± 0.1	39.2 ± 0.1
CoPO	81.2 ± 0.6	14.3 ± 0.5	4.0 ± 0.2	0.0 ± 0.0	0.5 ± 0.1	39.0 ± 0.1
FailMaker	21.3 ± 0.3	77.6 ± 0.3	0.5 ± 0.0	0.0 ± 0.0	0.6 ± 0.1	15.7 ± 0.1
Ours†	84.6 ± 0.5	11.5 ± 0.3	3.6 ± 0.3	0.1 ± 0.0	0.5 ± 0.1	42.2 ± 0.1

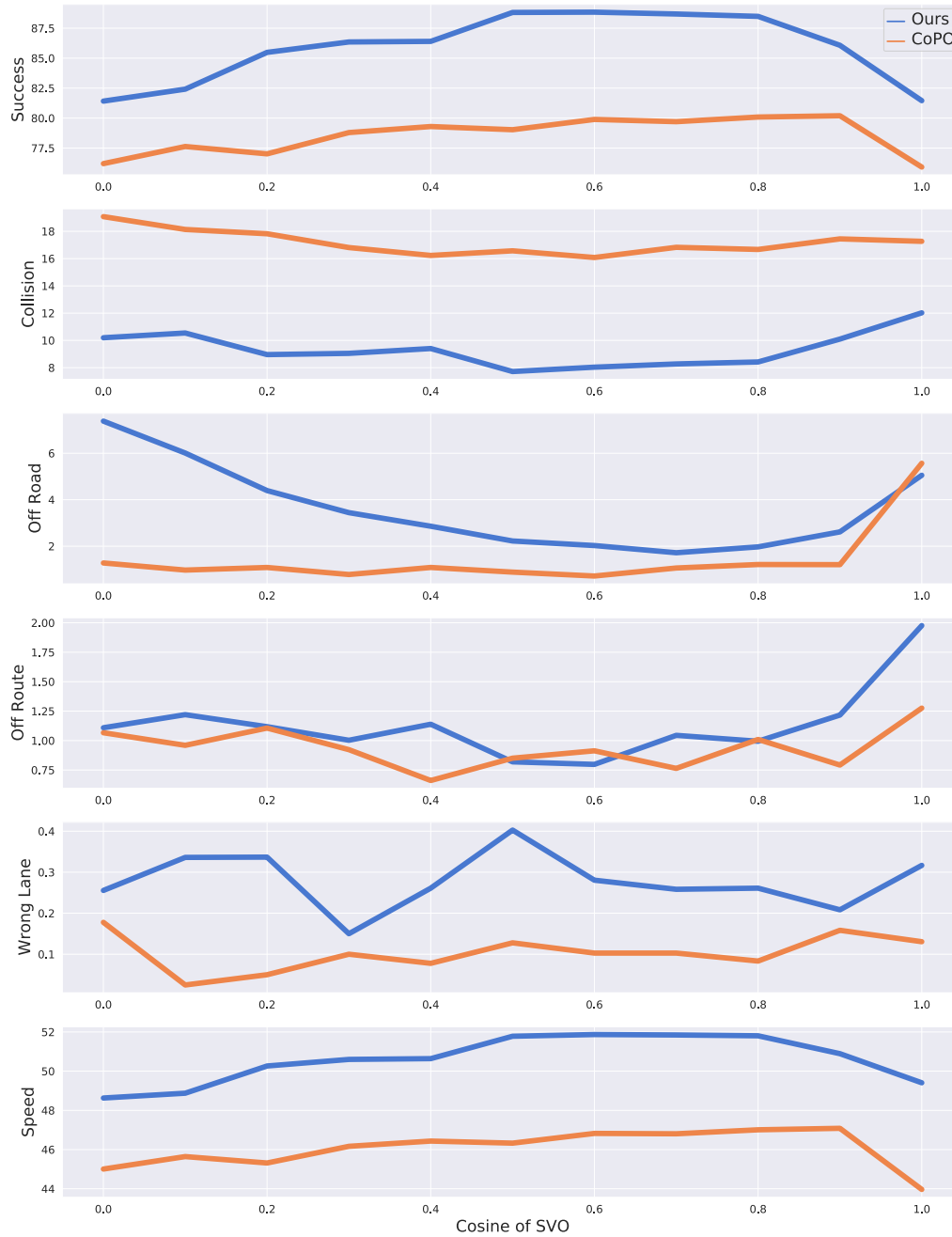


Figure 7: The performance of CoPO and our traffic flow in intersection.

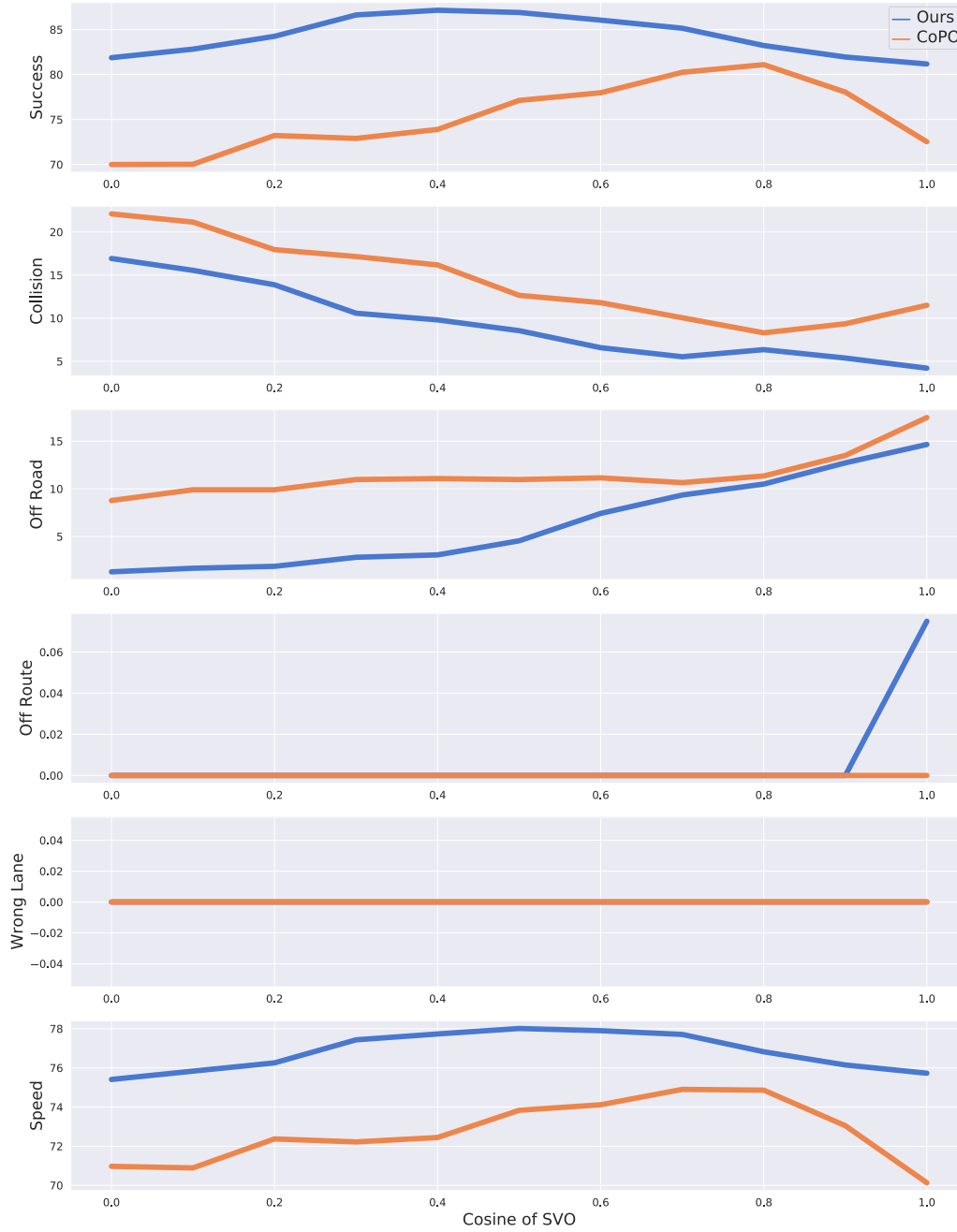


Figure 8: The performance of CoPO and our traffic flow in bottleneck.

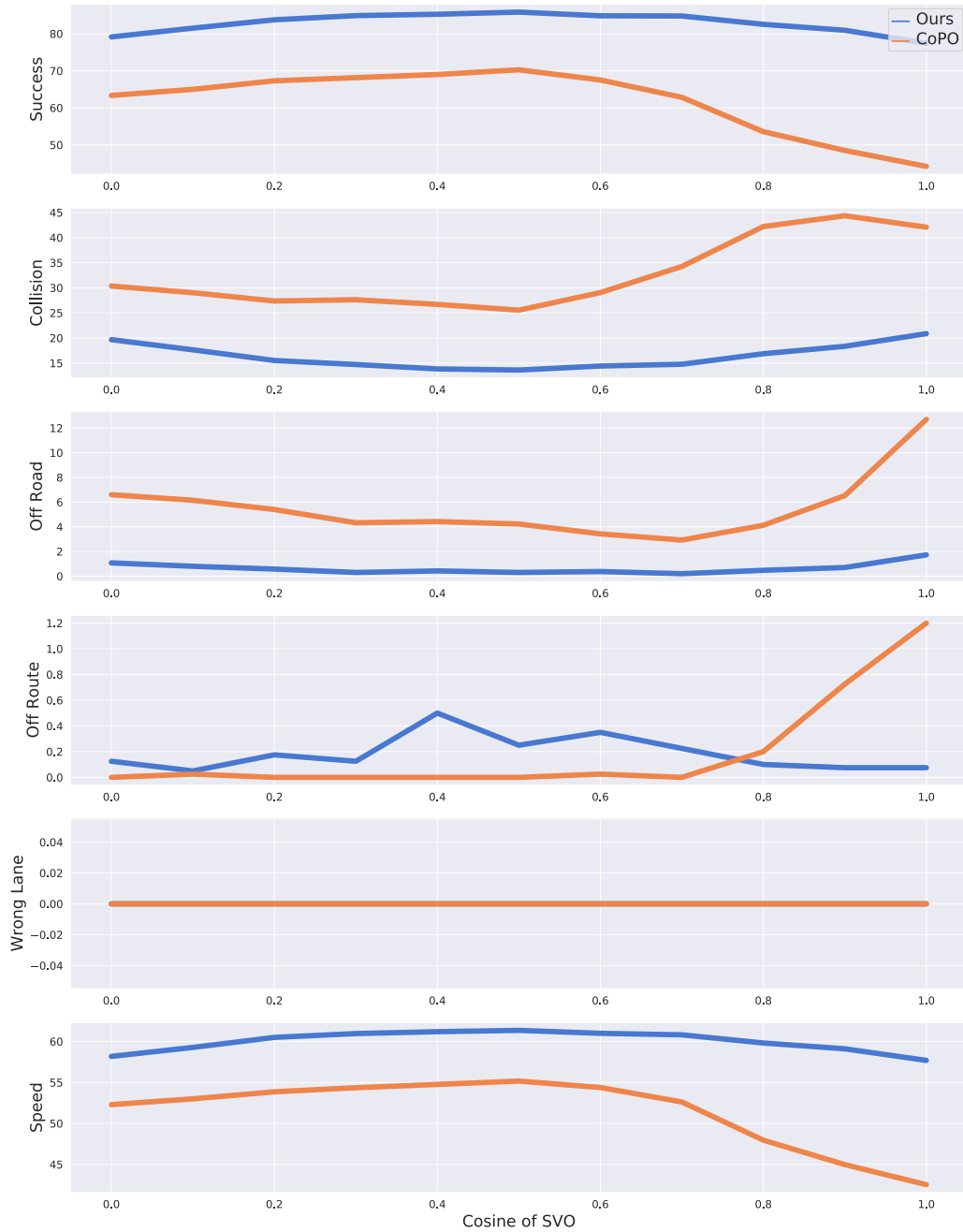


Figure 9: The performance of CoPO and our traffic flow in merge.

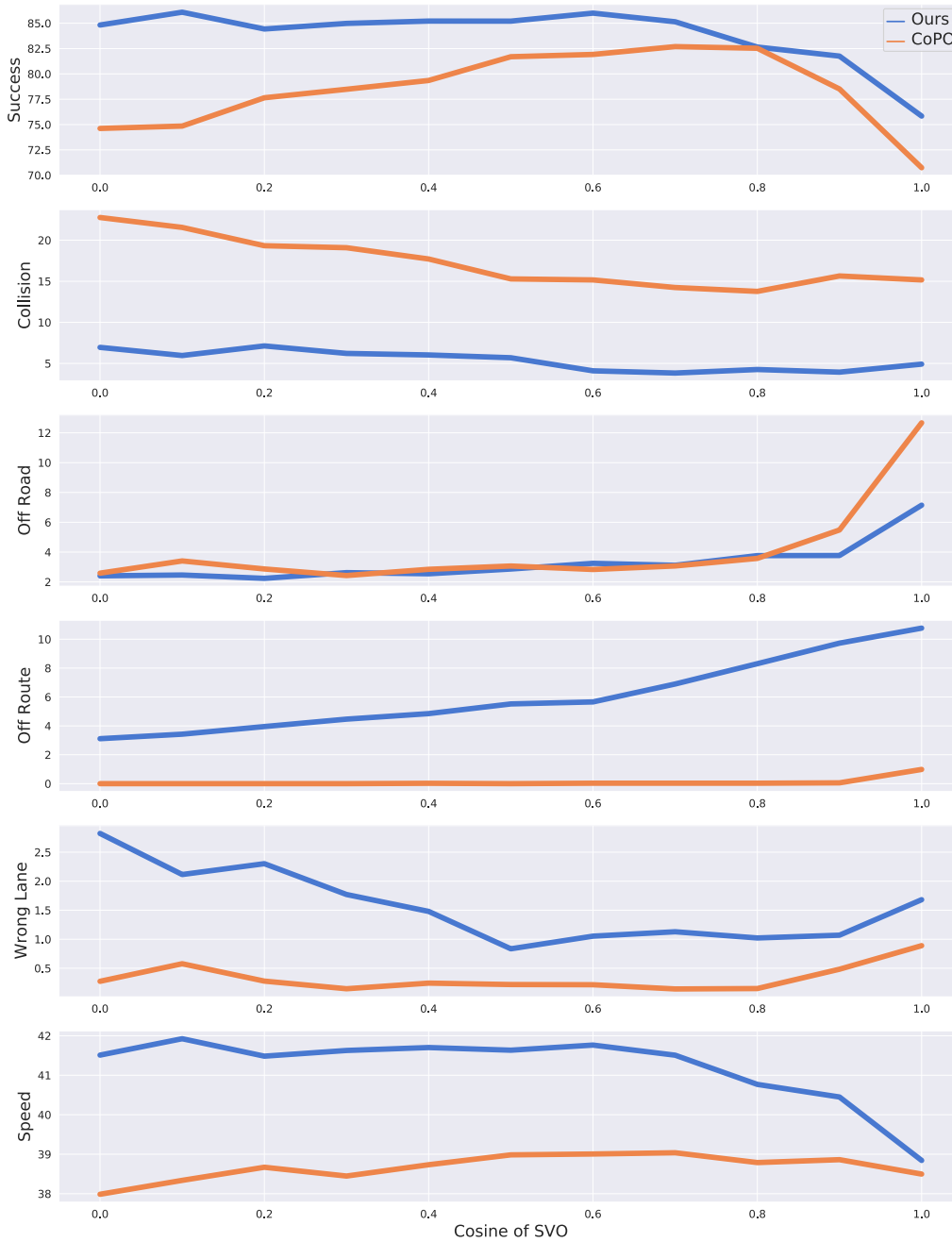


Figure 10: The performance of CoPO and our traffic flow in roundabout.

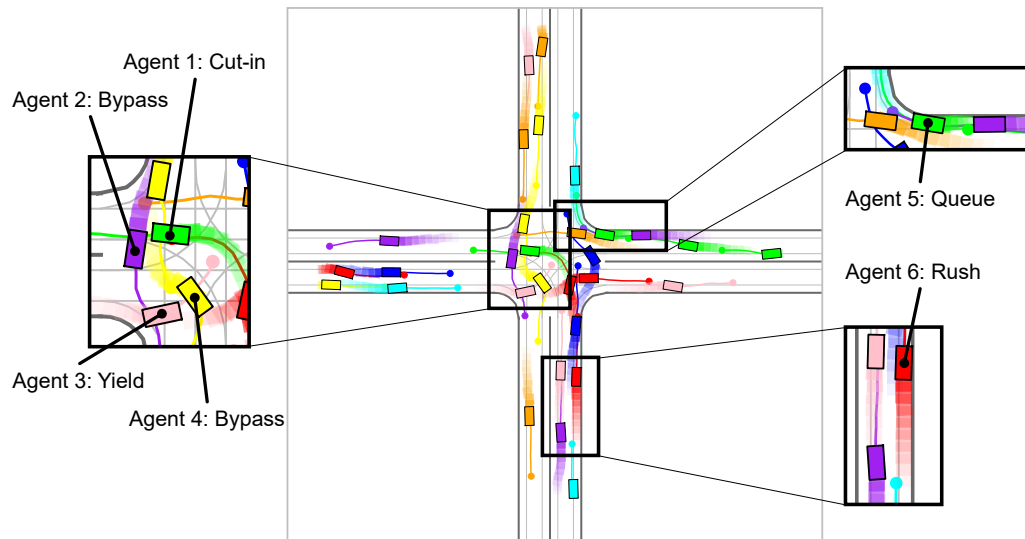


Figure 11: Coordinated behaviors in intersection.

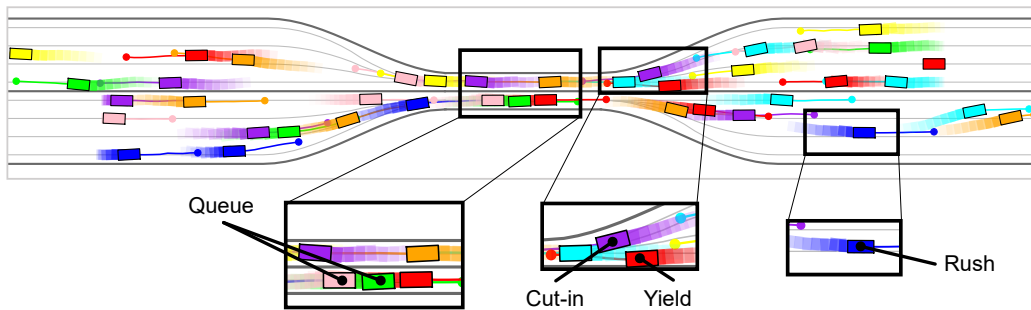


Figure 12: Coordinated behaviors in bottleneck.

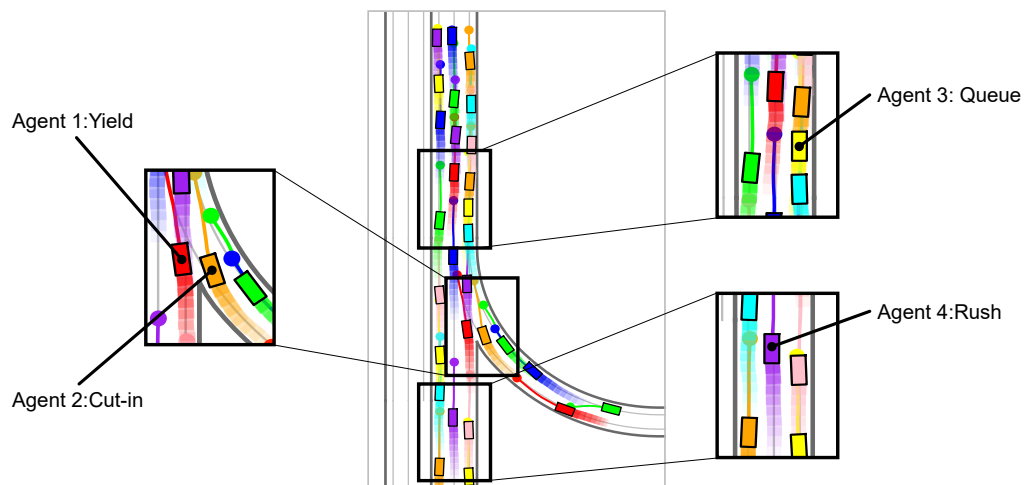


Figure 13: Coordinated behaviors in merge.

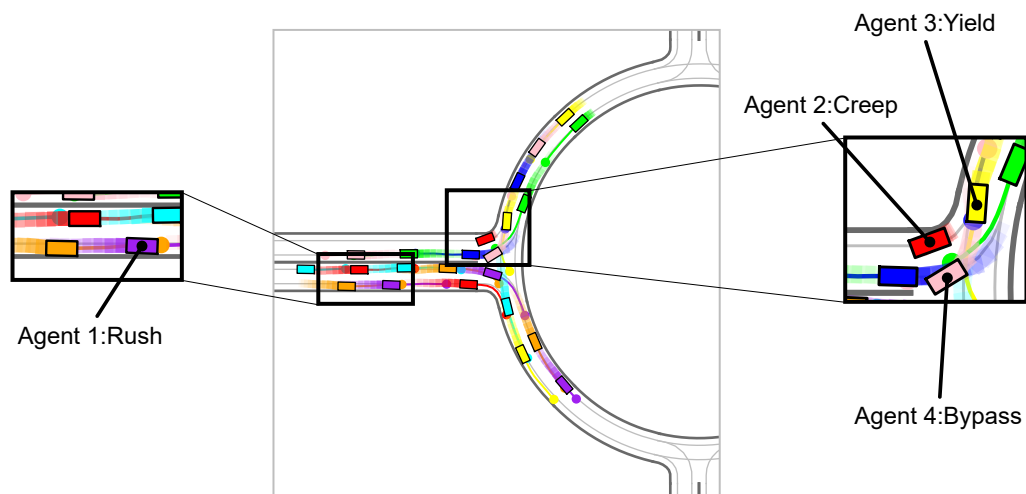


Figure 14: Coordinated behaviors in roundabout.

Table 5: **Zero-shot transfer performance in intersection.** Each subtable stores results of different driving policies in the same traffic flow. A “†” indicates our proposed method.

Evaluate in IDM	Intersection					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	79.5	6.0	2.0	0.0	0.0	47.1
VRL/FLOW	67.0	30.5	0.5	0.5	0.5	39.5
VRL/CoPO	69.0	27.0	0.5	0.0	0.0	43.0
eRARL	22.5	20.5	19.5	11.0	0.0	25.2
VRL/Ours†	69.0	28.5	3.5	2.0	0.0	40.8
iRARL†	71.5	21.5	6.0	1.0	0.0	42.1

Evaluate in FLOW	Intersection					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	65.5	18.5	13.0	1.5	0.5	42.8
VRL/FLOW	87.5	6.5	1.5	0.0	0.0	51.5
VRL/CoPO	76.0	12.0	2.0	3.0	0.0	46.5
eRARL	26.0	13.5	13.0	16.0	0.5	25.7
VRL/Ours†	79.0	11.0	2.0	4.5	0.0	47.9
iRARL†	83.5	8.0	2.0	3.0	0.5	49.7

Evaluate in CoPO	Intersection					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	69.5	20.0	8.0	1.0	0.5	45.0
VRL/FLOW	75.5	20.0	1.5	0.0	0.0	46.1
VRL/CoPO	81.5	16.0	1.0	0.0	0.0	48.7
eRARL	25.5	11.5	15.5	14.5	0.5	26.0
VRL/Ours†	74.5	19.0	1.0	2.0	0.0	45.8
iRARL†	79.5	16.5	0.0	1.0	0.0	47.7

Evaluate in FailMaker	Intersection					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	55.0	22.5	22.0	0.0	0.5	37.3
VRL/FLOW	52.0	37.0	12.0	0.0	0.0	36.0
VRL/CoPO	51.0	43.5	4.0	1.0	0.5	34.7
eRARL	26.5	11.5	17.5	14.5	0.0	25.9
VRL/Ours†	51.5	33.0	16.5	1.5	0.0	36.5
iRARL†	51.0	40.5	8.0	0.0	0.0	35.9

Evaluate in Ours wo Adv	Intersection					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	77.0	10.0	11.0	1.0	0.0	47.1
VRL/FLOW	84.0	13.5	1.0	0.5	0.5	50.1
VRL/CoPO	81.5	14.5	1.0	2.0	0.0	48.8
eRARL	21.5	12.5	21.0	15.0	0.5	24.9
VRL/Ours†	87.0	7.0	1.5	1.5	0.5	51.9
iRARL†	85.5	11.0	1.0	0.5	0.0	50.9

Evaluate in Ours	Intersection					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	74.5	14.5	7.5	2.5	0.0	45.9
VRL/FLOW	82.5	15.0	2.0	0.0	0.0	49.6
VRL/CoPO	78.5	18.5	1.5	1.0	0.0	47.5
eRARL	24.0	12.5	15.5	18.0	0.5	25.3
VRL/Ours†	85.0	9.0	3.0	1.0	0.5	50.8
iRARL†	86.0	10.5	0.5	1.5	0.0	51.4

Table 6: **Zero-shot transfer performance in bottleneck.** Each subtable stores results of different driving policies in the same traffic flow. A “†” indicates our proposed method.

Evaluate in IDM	Bottleneck					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	91.5	6.0	2.5	0.0	0.0	72.8
VRL/FLOW	53.0	25.5	19.5	2.5	0.0	51.7
VRL/CoPO	64.5	24.0	10.5	2.5	0.0	59.6
eRARL	50.0	10.0	37.5	3.0	0.0	33.9
VRL/Ours†	67.0	20.5	12.5	0.0	0.0	55.8
iRARL†	74.5	21.0	5.5	0.0	0.0	57.6

Evaluate in FLOW	Bottleneck					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	51.0	24.0	26.5	0.5	0.0	54.8
VRL/FLOW	79.0	18.5	2.0	0.5	0.0	76.1
VRL/CoPO	60.5	27.5	12.5	0.0	0.0	65.5
eRARL	42.0	18.5	39.5	0.0	0.0	25.6
VRL/Ours†	72.5	24.0	4.5	0.0	0.0	68.7
iRARL†	73.0	25.5	2.0	0.0	0.0	68.4

Evaluate in CoPO	Bottleneck					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	48.5	16.5	33.0	2.0	0.0	57.2
VRL/FLOW	61.0	6.5	33.0	0.0	0.0	64.6
VRL/CoPO	89.0	9.0	1.5	0.5	0.0	81.8
eRARL	51.0	11.0	34.5	3.5	0.0	35.1
VRL/Ours†	65.0	30.5	5.5	0.0	0.0	67.2
iRARL†	74.0	25.5	0.5	0.0	0.0	71.0

Evaluate in FailMaker	Bottleneck					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	20.5	59.0	26.0	0.0	0.0	34.5
VRL/FLOW	4.5	94.0	3.0	0.0	0.0	24.2
VRL/CoPO	6.5	88.5	6.5	0.0	0.0	27.5
eRARL	38.0	23.0	38.5	0.5	0.0	26.1
VRL/Ours†	2.5	87.5	13.5	0.0	0.0	24.3
iRARL†	4.5	94.0	2.0	0.0	0.0	24.1

Evaluate in Ours wo Adv	Bottleneck					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	52.5	26.0	21.0	0.5	0.0	58.3
VRL/FLOW	75.5	19.5	4.5	0.5	0.0	74.6
VRL/CoPO	71.5	13.5	15.5	1.0	0.0	70.9
eRARL	36.0	23.5	42.0	0.5	0.0	26.6
VRL/Ours†	91.0	4.0	5.0	0.0	0.0	81.3
iRARL†	89.5	8.5	2.0	0.0	0.0	79.6

Evaluate in Ours	Bottleneck					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	50.5	27.0	22.0	0.5	0.0	57.0
VRL/FLOW	73.5	24.0	2.0	0.5	0.0	73.8
VRL/CoPO	69.5	11.0	19.0	1.0	0.0	69.9
eRARL	36.5	19.0	43.5	1.0	0.0	27.7
VRL/Ours†	85.0	7.0	8.0	0.0	0.0	78.3
iRARL†	88.5	11.5	0.0	0.0	0.0	79.8

Table 7: **Zero-shot transfer performance in merge.** Each subtable stores results of different driving policies in the same traffic flow. A “†” indicates our proposed method.

Evaluate in IDM	Merge					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	92.0	6.0	2.0	0.0	0.0	62.3
VRL/FLOW	41.5	55.5	7.5	0.5	0.0	29.1
VRL/CoPO	54.5	19.0	22.5	9.0	0.0	38.0
eRARL	66.5	5.0	14.5	14.0	0.0	45.9
VRL/Ours†	67.5	18.5	14.5	0.0	0.0	43.6
iRARL†	85.0	15.0	0.0	0.0	0.0	53.2

Evaluate in FLOW	Merge					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	35.0	38.0	27.5	0.0	0.0	40.5
VRL/FLOW	69.5	25.0	5.5	0.0	0.0	56.0
VRL/CoPO	52.0	35.0	12.0	1.0	0.0	48.2
eRARL	17.0	40.5	33.5	9.5	0.0	24.2
VRL/Ours†	62.5	33.5	4.0	0.0	0.0	54.5
iRARL†	67.5	21.5	11.0	0.0	0.0	56.0

Evaluate in CoPO	Merge					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	36.5	44.5	19.0	0.0	0.0	40.0
VRL/FLOW	50.0	44.0	5.5	0.5	0.0	46.6
VRL/CoPO	70.5	24.5	5.0	0.0	0.0	55.4
eRARL	17.0	39.0	34.0	10.0	0.0	23.5
VRL/Ours†	57.0	37.5	5.5	0.0	0.0	49.9
iRARL†	59.5	40.0	0.5	0.0	0.0	52.0

Evaluate in FailMaker	Merge					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	9.0	74.5	18.0	0.0	0.0	15.1
VRL/FLOW	9.0	81.0	11.5	0.0	0.0	17.5
VRL/CoPO	10.0	83.5	10.0	0.0	0.0	19.2
eRARL	19.5	37.5	39.5	6.5	0.0	21.6
VRL/Ours†	8.5	88.0	4.0	0.0	0.0	16.8
iRARL†	12.0	78.5	11.0	0.0	0.0	18.5

Evaluate in Ours wo Adv	Merge					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	65.0	16.0	19.0	0.0	0.0	51.3
VRL/FLOW	71.5	21.0	7.5	0.0	0.0	55.0
VRL/CoPO	70.0	24.5	5.5	0.5	0.0	55.1
eRARL	18.5	29.0	47.5	6.5	0.0	27.6
VRL/Ours†	84.5	14.0	1.5	0.0	0.0	60.7
iRARL†	83.5	12.5	3.5	0.5	0.0	61.2

Evaluate in Ours	Merge					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	64.5	18.0	17.5	0.0	0.0	51.1
VRL/FLOW	69.5	22.5	8.0	0.0	0.0	54.1
VRL/CoPO	66.0	23.5	10.5	0.5	0.0	53.4
eRARL	18.5	25.0	50.5	7.0	0.0	27.2
VRL/Ours†	83.5	15.0	1.5	0.0	0.0	61.0
iRARL†	85.0	13.5	1.5	0.0	0.0	61.3

Table 8: **Zero-shot transfer performance in roundabout.** Each subtable stores results of different driving policies in the same traffic flow. A “†” indicates our proposed method.

Evaluate in IDM	Roundabout					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	83.0	9.0	5.5	0.0	2.5	38.7
VRL/FLOW	44.5	51.5	4.0	1.0	0.0	23.8
VRL/CoPO	59.0	33.0	9.5	0.0	0.0	31.5
eRARL	32.5	16.0	55.5	6.0	0.5	19.7
VRL/Ours†	40.5	30.5	31.0	0.5	1.0	21.1
iRARL†	63.0	18.5	17.0	0.0	1.5	32.3

Evaluate in FLOW	Roundabout					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	54.0	27.0	18.5	0.0	0.5	32.7
VRL/FLOW	84.0	11.5	4.5	0.0	0.0	42.7
VRL/CoPO	61.5	29.5	9.0	0.0	0.0	38.4
eRARL	30.5	24.0	46.0	5.0	0.5	17.7
VRL/Ours†	66.5	25.5	9.5	0.0	0.0	38.6
iRARL†	70.5	26.5	2.5	0.0	0.5	40.1

Evaluate in CoPO	Roundabout					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	55.5	18.5	23.0	0.5	2.5	32.4
VRL/FLOW	62.0	12.0	24.0	0.0	2.0	35.7
VRL/CoPO	86.5	12.0	1.5	0.0	0.0	41.8
eRARL	32.0	25.0	46.5	4.5	0.5	17.2
VRL/Ours†	70.0	26.5	3.5	0.0	0.5	38.8
iRARL†	76.5	17.5	5.5	0.0	0.5	39.9

Evaluate in FailMaker	Roundabout					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	19.5	64.0	20.5	0.0	0.5	20.2
VRL/FLOW	29.5	57.5	14.0	0.0	0.0	25.0
VRL/CoPO	21.0	70.5	10.0	0.0	0.5	22.7
eRARL	29.5	17.0	51.5	5.0	0.0	14.5
VRL/Ours†	17.0	63.0	24.5	0.0	0.0	19.9
iRARL†	18.5	57.5	25.5	0.0	0.0	20.5

Evaluate in Ours wo Adv	Roundabout					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	75.5	9.5	13.0	0.0	2.0	38.0
VRL/FLOW	74.0	12.0	14.0	0.0	0.0	39.7
VRL/CoPO	82.0	13.5	1.5	1.5	2.5	42.4
eRARL	30.0	16.5	54.5	5.0	0.0	17.7
VRL/Ours†	87.5	9.0	3.5	0.0	0.0	43.2
iRARL†	86.0	12.0	2.0	0.0	0.0	42.8

Evaluate in Ours	Roundabout					
	Success (↑)	Collision (↓)	Off Road (↓)	Off Route (↓)	Wrong Lane (↓)	Speed (↑)
VRL/IDM	72.5	13.0	13.0	0.0	1.5	37.8
VRL/FLOW	81.0	4.0	14.5	0.5	0.0	41.2
VRL/CoPO	80.5	17.5	1.5	1.0	0.0	42.2
eRARL	31.0	15.5	56.0	4.5	0.5	18.2
VRL/Ours†	84.5	11.5	5.0	0.0	0.0	42.3
iRARL†	87.0	10.5	2.0	0.0	0.5	42.9