NSF CMMI #1527016 & #1526835

Robot-Assisted Pedestrian Regulation Based on Deep Reinforcement Learning

Yi Guo[1] (PI), Haibo He[2](PI), Chao Jiang[1] , Zhiqiang Wan[2]
[1]Stevens Institute of Technology, [2]University of Rhode Island

Introduction

- Pedestrian crowd regulation is of great importance for avoiding crowd accident in densely-populated areas or during the emergency evacuation;
- We introduce an autonomous mobile robot to dynamically interact with evacuating pedestrians;
- We propose an end-to-end solution which directly inputs the raw image of the environment and outputs real-time robot motion decisions to regulate pedestrian flow.

Problem Formulation

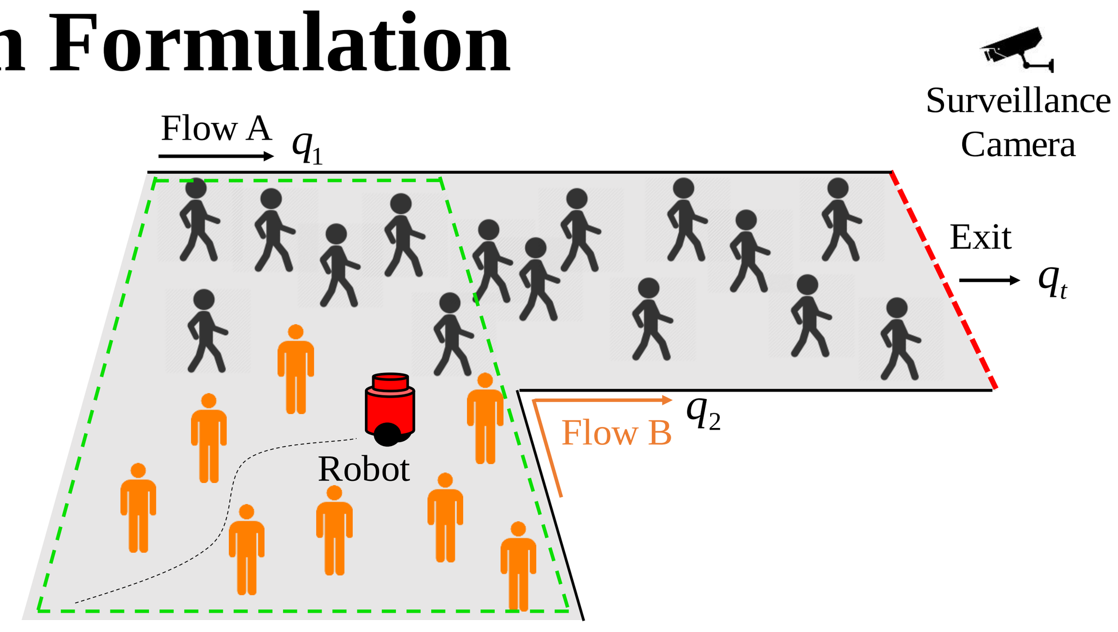


Fig. 1: The schematic diagram of merging pedestrian flows with robot-assisted regulation

- State x_t: the pedestrian position and the robot position.
- Action u_t: the robot motion decision which represents its moving directions, that is, "up", "down", "left", "right".
- State transition $\boldsymbol{x}_{t+1} = f(\boldsymbol{x}_t, \boldsymbol{u}_t)$: determined by human interaction and human-robot interaction (HRI).
- Reward q_t: the instantaneous outflow.
- Action-value function:

$$Q_{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{E}_{\boldsymbol{\pi}}\left[\sum_{k=0}^{K} \gamma^k \cdot q_{t+k} \middle| \boldsymbol{x_t} = \boldsymbol{x}, \boldsymbol{u_t} = \boldsymbol{u}\right]$$

- *Problem statement*: Determine the optimal robot motion decisions $\boldsymbol{u}$ such that the accumulated pedestrian outflow is maximized under the robot-assisted regulation.

$$\underset{\boldsymbol{u}}{\text{maximize}} \quad J = \sum_{t=0}^{T} q_t$$

End-to-End Flow Regulation

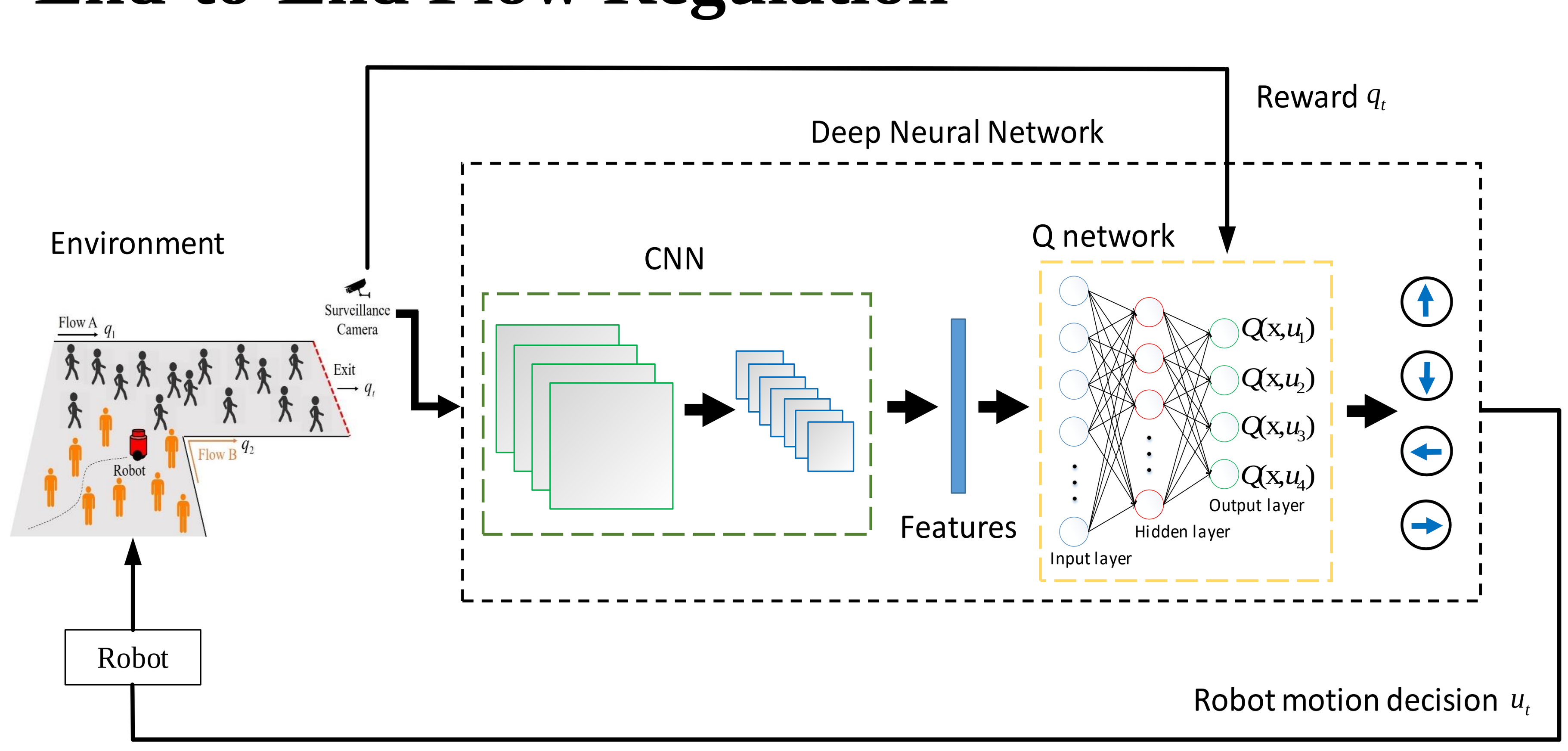


Fig. 2: The overall control diagram. The end-to-end approach uses the raw image obtained by the surveillance camera as input, observes the reward q_t, accordingly, and outputs robot motion decision u_t in real time.

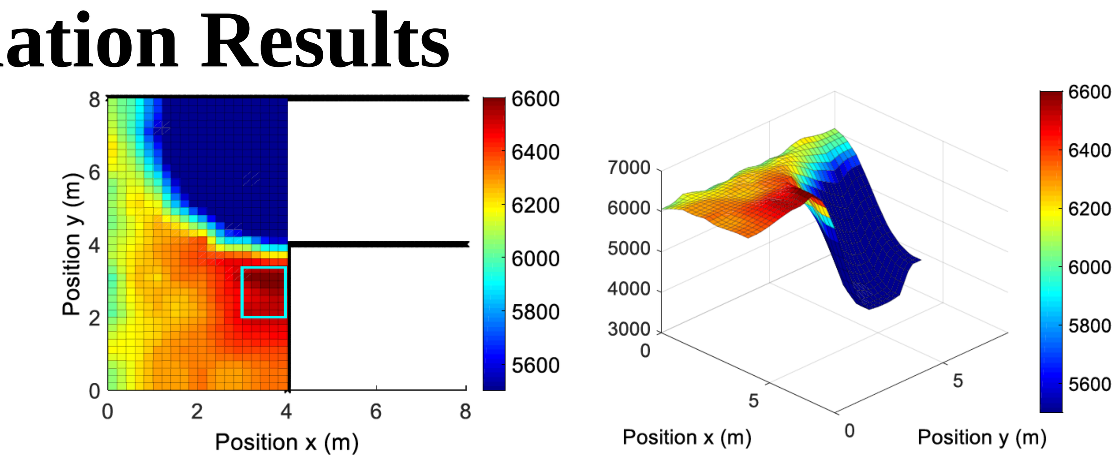
Learning Process

Algorithm 1 Training Process

1: Initialize the deep neural network (DNN) with random parameters θ.
2: **for** Epoch=1:N **do**
3: Initialize the robot position.
4: **for** Time step t = 1:T **do**
5: Input the image captured by the surveillance camera into the DNN.
6: Calculate the action-value $Q(\boldsymbol{x}_t, \boldsymbol{u}; \theta_t)$ with the DNN.
7: Output the robot motion decision $\boldsymbol{u}_t = \text{argmax}_u\, Q(\boldsymbol{x}_t, \boldsymbol{u}; \theta_t)$.
8: Update the robot position and observe the reward q_t.
9: Advance to the next state $\boldsymbol{x}_{t+1}$.
10: Store the experience tuple $(\boldsymbol{x}_t, \boldsymbol{u}_t, q_t, \boldsymbol{x}_{t+1})$ in a buffer $\mathcal{B}$.
11: Sample M tuples $\{(\boldsymbol{x}_j, \boldsymbol{u}_j, q_j, \boldsymbol{x}_{j+1})\}_{j=1}^{M}$ from $\mathcal{B}$.
12: Set the target action-value $y_j = q_j + \gamma \max_{u'} Q\left(\boldsymbol{x}_{j+1}, \boldsymbol{u}'; \theta_t\right)$.
13: Calculate the loss function $L(\theta_t) = \sum_{j=1}^{M} (y_j - Q(\boldsymbol{x}_j, \boldsymbol{u}_j; \theta_t))^2$.
14: Update DNN with mini-batch gradient descent $\theta_{t+1} = \theta_t - \eta \bigtriangledown_{\theta_t} L(\theta_t)$.
15: **end for**
16: **end for**

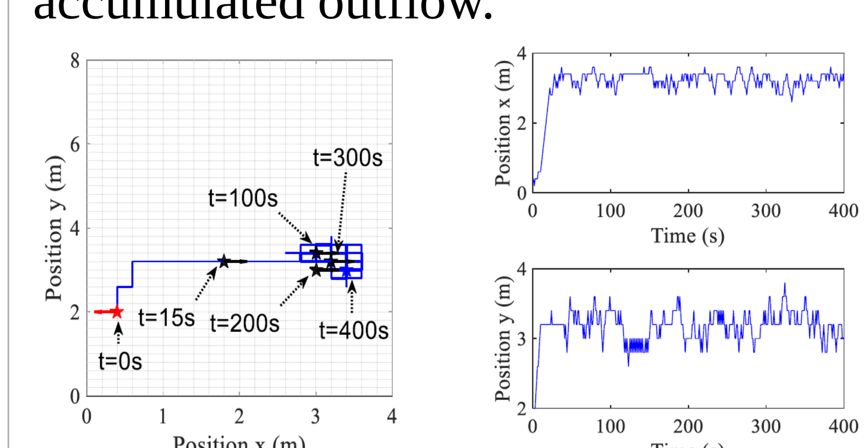
- After the training process, the parameters of the DNN will be saved for the online deployment.

Simulation Results



(a) (b)

Fig. 3: HRI characteristics for inflow rate $q_1/q_2 = 1/3$: (a) top-view; (b) 3D-view. The color indicates the quantity of the accumulated outflow, $\sum_{t=0}^{T} q_t$, at T=400s. The rectangle in (a) highlights the robot positions with the highest accumulated outflow.



(a) (b)

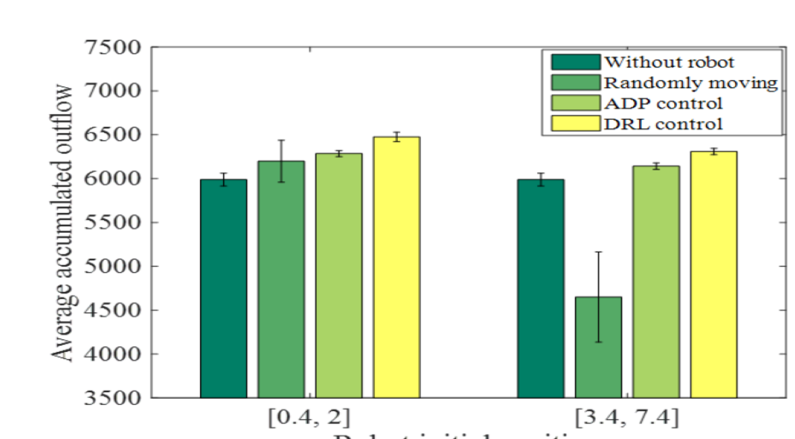
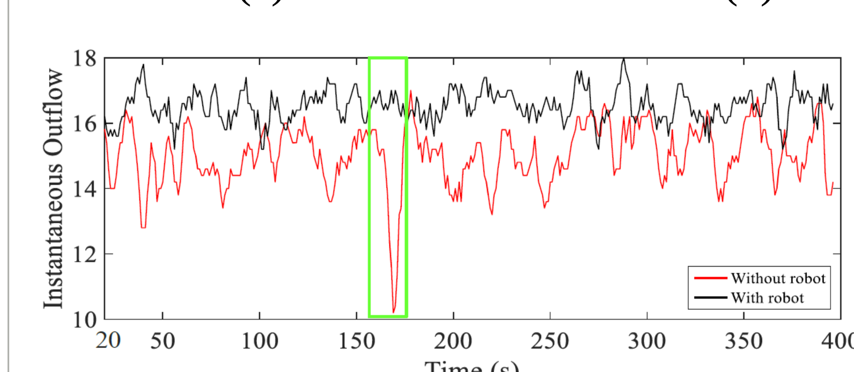
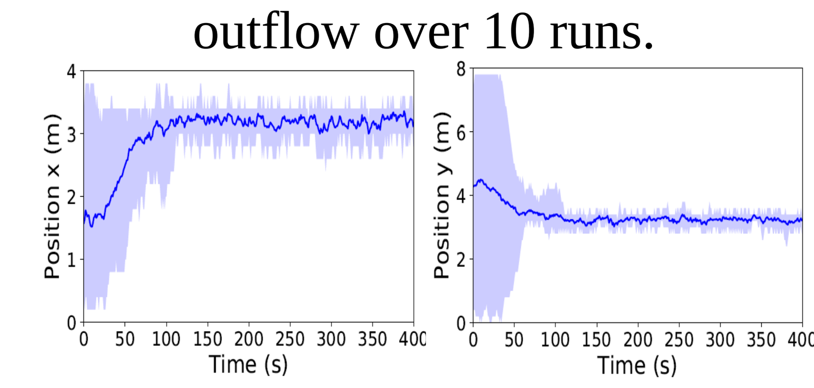


Fig. 5 Average accumulated outflow over 10 runs.



(c)



(a) (b)

Fig. 4: (a) Robot trajectory; (b) Time history of robot position in x and y directions; (c) Instantaneous outflow, q_t.

Fig. 6 Time history of robot position with 10 robot initial positions.

- Robot position converges to the rectangular region in Fig. 3 (a) that maximizes the accumulated outflow.
- The accumulated outflow is improved by about 8.1% in comparison with no-robot case.
- The proposed approach achieves best regulation performance.

Conclusion

- We proposed to regulate merging pedestrian flow in a T-shaped junction through HRI;
- We presented an end-to-end approach based on deep Q-learning to solve this problem with offline training and online deployment;
- Various simulation results demonstrated promising regulation performances.