Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2025, 14(6), 1268-1275
Published Online November 2025 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2025.146119

Hans汉斯

RCAT：面向拥挤人群导航的机器人中心注意力时空建模方法

吕龙凤[1]，潘为刚[1]，薛秉鑫[2]

[1]山东交通学院轨道交通学院，山东 济南
[2]山东交通学院信息科学与电气工程学院，山东 济南

收稿日期：2025年9月29日；录用日期：2025年10月27日；发布日期：2025年11月4日

摘　要

在拥挤人群环境中实现安全、高效的自主导航是智能机器人面临的核心挑战。现有基于强化学习和自注意力机制的方法在时空建模方面已有一定成效，但往往缺乏对“机器人中心视角”的关注，导致机器人在复杂场景下难以准确聚焦于关键个体。为此，本文提出了一种机器人中心交叉注意力串行时空Transformer (RCAT)方法。该方法首先通过空间模块和时间模块串行建模行人的全局交互与时序动态，随后引入跨注意力机制，以机器人状态为查询向量，从时空特征中筛选与任务最相关的个体信息，从而实现更加任务导向的人群建模。在二维仿真环境中的实验结果表明，RCAT相比SARL方法，在不同人群密度下均表现出更高的成功率、更低的碰撞率以及更短的平均到达时间，并在累积奖励上取得显著优势。研究结果验证了RCAT在复杂人群导航任务中的安全性、效率和鲁棒性。

关键词

机器人导航，人群交互建模，深度强化学习，跨注意力机制，时空Transformer

RCAT: A Spatial-Temporal Modeling Method for Robot Central Attention in Crowd Navigation

Longfeng Lyu[1], Weigang Pan[1], Bingxin Xue[2]

[1]School of Rail Transit, Shandong Jiaotong University, Jinan Shandong
[2]School of Information Science and Electrical Engineering, Shandong Jiaotong University, Jinan Shandong

Received: September 29, 2025; accepted: October 27, 2025; published: November 4, 2025

文章引用: 吕龙凤, 潘为刚, 薛秉鑫. RCAT: 面向拥挤人群导航的机器人中心注意力时空建模方法[J]. 人工智能与机器人研究, 2025, 14(6): 1268-1275. DOI: 10.12677/airr.2025.146119

Abstract

Realizing safe and efficient autonomous navigation in crowded environments is the core challenge faced by intelligent robots. The existing methods based on reinforcement learning and self attention mechanisms have achieved certain results in spatiotemporal modeling, but often lack attention to the "robot central perspective", which makes it difficult for robots to accurately focus on key individuals in complex scenes. Therefore, this article proposes a robot center cross attention serial spatiotemporal Transformer (RCAT) method. This method first models the global interaction and temporal dynamics of pedestrians in series through spatial and temporal modules. Then, a cross attention mechanism is introduced, using the robot state as the query vector, to filter individual information most relevant to the task from spatiotemporal features, thereby achieving more task oriented crowd modeling. The experimental results in a two-dimensional simulation environment show that RCAT exhibits higher success rates, lower collision rates, and shorter average arrival times compared to SARL methods under different population densities, and achieves significant advantages in cumulative rewards. The research results validated the safety, efficiency, and robustness of RCAT in complex crowd navigation tasks.

Keywords

Robot Navigation, Crowd Interaction Modeling, Deep Reinforcement Learning, Cross Attention Mechanism, Spatiotemporal Transformer

1. 引言

具身人工智能(Artificial Intelligence, AI)正逐渐成为智能机器人研究的核心方向，其目标是赋予智能体感知、学习、推理与行动等能力[1]。在实际应用中，一个关键挑战是如何让机器人在复杂且动态的环境中实现安全与高效的自主导航，特别是在拥挤人群场景下，这一能力是诸多任务的前提条件。

随着深度学习与强化学习的快速发展，深度强化学习(Deep Reinforcement Learning, DRL) [2]已成为智能决策的重要手段。大量研究尝试将导航问题建模为马尔可夫决策过程(MDP)，通过最大化长期累积奖励学习最优策略，从而驱动机器人在动态环境中完成目标[3]-[8]。例如，Mirowski 等人[4] [5]基于异步优势 Actor-Critic (A3C) [9]实现了复杂场景下的导航；Josef 等人[7]提出了基于深度 Q-Learning (DQN) [10]的局部路径规划方法。

在拥挤场景中，研究者进一步探索了基于值函数(Value Function)的 DRL 方法来解决导航与避障问题。Chen 等人[11]提出的 CADRL 使用时间差分(TD)学习结合经验回放来训练值网络，实现了有效的避碰能力。随后，有学者在[12]中通过引入社会规范奖励对[11]的方法进行了扩展，使机器人能够更符合人类行为习惯地完成任务。Everett 等人[13]则采用长短期记忆网络(LSTM)对动态人群进行建模，从而处理不定数量的交互体。Chen 等人[14]进一步引入自注意力机制，能够区分邻近个体的重要性，从而提升决策的精度。然而，这些方法大多偏重空间维度的交互建模，而对时间维度的轨迹信息关注不足，限制了导航的效率与安全性。

为解决时空信息建模的不足，Liu 等人[15]提出了去中心化结构递归神经网络(DSRNN)，用于处理交

互过程中的时间依赖，但 RNN 在捕获全局时空状态方面仍存在局限。近年来，研究者提出了 ST^2 (Spatial-Temporal State Transformer) [16]，该方法利用 Transformer 架构同时建模空间与时间交互关系，从而在动态人群场景下取得了优于 LSTM 等方法的效果。然而，ST^2 的注意力机制是对称的自注意力，即机器人与行人在交互过程中具有相同权重，这可能导致机器人难以区分“与自身任务最相关的个体”，从而在高密度场景下产生效率或安全性下降的问题。

尽管现有的自注意力机制在群体建模中取得了显著成效，但这种对称式的信息交互模式假设机器人与所有行人在信息处理上的地位相同。这种假设在群体行为预测任务中是合理的，但对于自主导航任务而言却存在明显局限。导航的决策主体是机器人，它需要从全局人群中重点关注与自身运动相关的关键行人，而非平均地处理所有目标。因此，导航模型的核心挑战不在于简单地捕捉群体整体关系，而在于如何让机器人主动感知与其路径冲突或潜在风险最高的个体。目前已有部分研究尝试引入“机器人中心”思想，如在社会注意力(Social Attention)中对机器人邻域进行加权，但这些方法仍然属于显式权重分配，难以动态地随环境变化调整交互焦点。

基于此，本文提出了一种机器人中心交叉注意力的串行时空 Transformer (Robot-Centric Cross-Attention Transformer, RCAT)。该方法首先采用空间–时间串行编码捕捉行人间的交互与动态变化，再通过跨注意力机制以机器人状态作为查询向量，从时空特征中选择性聚焦与导航最相关的行人，从而实现“机器人中心”的人群建模。最后，RCAT 融合全局池化特征、上下文向量和机器人自身状态，预测状态价值并指导决策。

与传统方法相比，RCAT 在机制上引入了非对称的信息交互结构，使机器人能够主动筛选关键个体，从而提升模型在高密度人群中的任务适应性与导航安全性。通过将机器人状态作为查询向量，引导注意力机制聚焦于与自身运动最相关的行人，RCAT 在信息聚焦能力与任务导向性方面展现出显著优势。该设计不仅增强了模型对复杂人群动态的响应能力，也为提升导航效率与鲁棒性提供了结构性保障。

2. 问题描述与建模

在动态人群场景中，机器人需要具备在有限感知条件下进行安全高效导航的能力。该问题可以形式化为一个基于部分可观测马尔可夫决策过程(POMDP)的决策学习任务。POMDP 通常由状态集合、动作集合、状态转移概率、奖励函数、折扣因子以及观测机制构成。对于机器人而言，目标是学习一条策略，使其在未知的人群运动模式下能够逐步接近目标点，同时减少潜在碰撞风险并保持行驶的平稳性。

在每个仿真场景中，环境由一个机器人和若干动态行人组成。设机器人编号为 0，其他行人编号为 1 至 n。在任意时刻，个体的观测向量包含 13 个特征，其中既包括自身的几何位置与速度，也包括与机器人相关的信息，如机器人目标位置、首选速度、朝向角及其与目标点的距离。此外，机器人当前的位置信息与运动状态也被附加到行人特征中，以便于显式建模人机之间的交互关系。

RCAT 方法通过与环境的持续交互，采样状态转移并优化导航策略，以最大化未来累积回报。其核心思想在于对人群特征的建模与融合：首先采用时空编码模块获取行人的动态表示，随后引入交叉注意力机制，由机器人状态生成查询向量，针对所有行人特征执行定向聚合。这种设计能够使机器人更加关注与其路径和安全性最相关的个体，从而得到更具任务导向的上下文表示。最后，融合得到的特征与机器人自身状态共同输入价值网络，并通过时间差分方法进行训练，实现复杂人群环境下的自主导航。

3. 方法学

3.1. 值网络

在 RCAT (Robot-Centric Cross Attention Transformer)中，机器人自主导航任务被建模为基于值的深度

强化学习问题。为了在复杂人群环境中准确建模人机交互，本方法提出了一种时空串行编码架构，并在融合阶段引入跨注意力机制，如图 1 所示。整个框架由空间模块、时间模块和融合模块三部分组成。

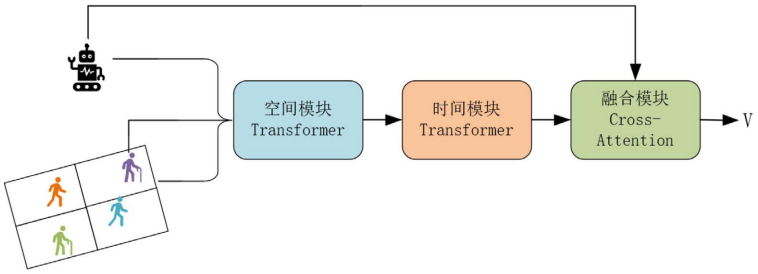

Figure 1. Value network sequence diagram
图 1. 值网络顺序图

3.1.1. 空间模块

在时刻 t，联合状态可表示为：

$$S_t = \left\{S_r^t, S_1^t, S_2^t, \cdots, S_n^t\right\} \tag{1}$$

其中，S_r^t 表示机器人状态，S_i^t 表示第 i 个行人的状态。每个行人的 13 维特征向量包括自身几何与速度信息，以及与机器人相关的目标和当前位置特征。

空间模块通过空间 Transformer 提取人群全局交互信息：

$$H_s = MSA_s\left(\phi\left(S_t\right)\right) \tag{2}$$

其中，$\phi(\bullet)$ 为线性嵌入函数，MSA_S 表示空间多头自注意力机制，得到的不同行人对机器人的全局影响编码为 $H_S \in \mathbb{R}^{N\times d}$。

3.1.2. 时间模块

由于人群运动具有动态特性，机器人需要结合历史帧进行决策。时间模块将连续 T 帧的空间特征输入时间 Transformer，建模轨迹依赖关系：

$$H_t = MSA_t\left(H_s^{0:T}\right) \tag{3}$$

其中，MSA_t 表示时间维度的多头自注意力机制，输出 $H_t \in \mathbb{R}^{T\times N\times d}$，包含行人时序动态。

3.1.3. 融合模块

不同于 ST^2 中的自注意力机制融合，RCAT 在融合阶段引入交叉注意力机制：以机器人状态作为查询向量(Query)，行人的时空特征作为键值(Key/Value)。

$$Q_r = s_r W^Q, K_h = H_t W^K, V_h = H_t W^v \tag{4}$$

$$C = \text{softmax}\left(\frac{Q_r K_h^\top}{\sqrt{d}}\right) V_h \tag{5}$$

其中，W^Q, W^K, W^V 为可学习投影矩阵。这样，机器人能够主动挑选与其决策最相关的行人，得到上下文表示 C。

最终，机器人价值函数通过以下方式估计：

$$V\left(s_t\right) = f_\theta\left(\left[s_r, C, \bar{H}_t\right]\right) \tag{6}$$

其中，$\bar{H}_t$ 为时空特征池化结果，f_θ 为多层感知机(MLP)。价值网络采用时间差分方法进行更新：

$$V(s_t) \leftarrow V(s_t) + \alpha\left(r_t + \gamma V(s_{t+1}) - V(s_t)\right) \tag{7}$$

3.2. 奖励函数

在 RCAT 框架中，奖励函数的设计直接影响机器人在动态人群环境中的导航行为。为了实现安全、高效且自然的路径规划，奖励函数考虑了碰撞惩罚、到达奖励、舒适性约束三个方面：

3.2.1. 碰撞惩罚

若机器人与行人发生物理碰撞(欧氏距离小于半径之和)，立即给予一个较大的负奖励：

$$r_t^{\text{collision}} = \begin{cases} R_c, \text{if } d_{r,h} < r_r + r_h, \\ 0, \text{otherwise}, \end{cases} \tag{8}$$

其中，$d_{r,h}$ 表示机器人和行人之间的距离，r_r, r_h 分别为机器人和行人的半径，$R_c \le 0$ 为碰撞惩罚值。

3.2.2. 到达奖励

当机器人到达目标点(当前位置与目标点的欧氏距离小于机器人半径)时，给予正奖励以鼓励任务完成：

$$r_t^{\text{goal}} = \begin{cases} R_g, \text{if } \|(p_x, p_y) - (g_x, g_y)\| < r_r, \\ 0, \text{otherwise}, \end{cases} \tag{9}$$

其中，$R_g \ge 0$ 表示到达奖励。

3.2.3. 舒适性惩罚

为了避免机器人虽然未发生碰撞但过度接近行人，在距离小于安全阈值 d_{safe} 时，给予一定的惩罚：

$$r_t^{\text{discomfort}} = \begin{cases} (d_{r,h} - d_{\text{safe}}) \cdot \lambda, \text{if} d_{r,h} < d_{\text{safe}}, \\ 0, \text{otherwise}, \end{cases} \tag{10}$$

其中，$\lambda \le 0$ 为舒适性惩罚因子。

3.2.4. 总奖励函数

综合上述三部分，时刻 t 的即时奖励为：

$$r_t = r_t^{\text{collision}} + r_t^{\text{goal}} + r_t^{\text{discomfort}} \tag{11}$$

在强化学习中，RCAT 的目标是最大化期望累计折扣回报：

$$J(\pi) = \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t r_t\right] \tag{12}$$

其中 γ 为折扣因子，T 为任务终止时刻。

3.3. 训练环境与参数设置

为了验证 RCAT 在动态人群场景中的导航性能，我们在一个基于仿真的二维环境中进行训练和测试。环境中包含一个机器人和若干($n = 5, 7$)动态行人，行人数量在不同场景中随机变化。机器人和行人均被建模为带有半径的圆盘体，能够在连续二维平面上移动。

3.3.1. 机器人设置

(1) 初始位置和目标位置在环境边界上随机采样。

(2) 运动学模型采用全向运动(holonomic)，动作空间由一组离散的速度和旋转样本组成：

$$\mathcal{A}=\left\{(v,\Delta\theta)\middle|\ v\in V,\Delta\theta\in\Theta\right\} \tag{13}$$

其中，V 表示线速度集合，Θ 表示旋转角度集合。

3.3.2. 行人建模

(1) 行人数量 n 在每个场景中随机设定。

(2) 初始位置和目标位置随机分布，行人以恒定速度朝目标运动。

(3) 行人的速度和半径根据现实人群的统计分布随机采样，以增加场景多样性。

3.3.3. 状态表示

在时刻 t，机器人与每个行人的联合状态用一个 13 维向量表示：

$$o_i^t=\left(px,py,vx,vy,r,g_x,g_y,v_{\text{pref}},\theta,d_g,x_r,y_r,v_r\right) \tag{14}$$

其中包括行人自身的几何和速度特征，以及与机器人目标和机器人当前位置相关的辅助信息。这种状态设计能够显式增强人机交互建模。

3.3.4. 训练方法

(1) 算法采用基于值函数的深度强化学习。

(2) RCAT 使用时间差分(TD)更新价值网络：

$$V\left(s_t\right)\leftarrow V\left(s_t\right)+\alpha\left(r_t+\gamma V\left(s_{t+1}\right)-V\left(s_t\right)\right) \tag{15}$$

其中 α 为学习率，γ 为折扣因子。

(3) 在训练过程中，机器人通过与虚拟人群的交互不断采样状态转移，从而学习最优策略。

3.3.5. 参数设置

训练阶段采用基于值函数的深度强化学习框架，优化器选用 Adam，学习率设置为 0.001。折扣因子 γ 设为 0.9，以平衡短期和长期奖励；时间步长 Δt 为 0.25 s，每个仿真回合的最大时长为 30 s。

模型训练在 NVIDIA RTX A5000 GPU (24 GB 显存)上进行，总训练轮数为 10,000，每轮包含一个采样回合与 100 个更新批次。损失函数在约 2000 轮后收敛，验证集成功率趋于稳定。模型参数通过 PyTorch 框架实现，训练时批大小设置为 100。主要参数配置如表 1 所示。

Table 1. Main parameter configuration
表 1. 主要参数配置

参数类别	参数名称	数值
环境参数	时间步长 Δt	0.25 s
	最大仿真时长 T	30 s
	行人数量	5/7
强化学习参数	折扣因子 γ	0.9
	学习率 α	0.001 (Adam 优化器)
	批大小	100
	训练回合数	10,000

4. 实验与结果

在实验结果中，RCAT 在多个指标上均优于传统的 SARL 方法，如表 2 所示。首先，在成功率方面，

当场景中存在 5 个行人时，SARL 的成功率为 0.95，而 RCAT 提升至 0.98；当场景中存在 7 个行人时，SARL 的成功率为 0.93，而 RCAT 提升至 0.96。这表明 RCAT 通过跨注意力机制更好地捕捉了机器人和行人之间的交互关系，使得机器人能够更稳定地找到安全路径。在碰撞率方面，当场景中存在 5 个行人时，RCAT 的碰撞率为 0.01，相比 SARL 的 0.03 明显降低；而在 7 个行人的复杂场景下，RCAT 的碰撞率为 0.03，与 SARL 的 0.02 基本相当。这说明 RCAT 在低到中等密度人群中具有更好的安全性，同时在更复杂的环境中也保持了稳定表现。

在平均到达时间方面，RCAT 在两种场景下的结果分别为 9.43 秒和 10.12 秒，均优于 SARL 的 10.57 秒和 11.09 秒。这表明 RCAT 能够在保证安全性的前提下，使机器人更快到达目标点，从而展现出更高的导航效率。在累积奖励方面，RCAT 在 5 个行人和 7 个行人的场景下分别达到了 0.3315 和 0.3164，均明显高于 SARL 的 0.2906 和 0.2784。奖励值的提升进一步验证了 RCAT 在整体导航表现上的优势。

综上所述，RCAT 相比传统的 SARL 方法，在不同人群规模下均展现出更好的导航性能。它不仅提升了机器人到达目标的成功率，降低了在部分场景下的碰撞率，还有效缩短了到达时间，从而获得了更高的累积奖励。这些改进主要得益于跨注意力机制在建模人机交互时的优势，使机器人能够更加精确地关注与自身运动相关的行人动态信息，从而实现更高效和更安全的导航。

Table 2. Experimental result
表 2. 实验结果

Methods	n_h	Success	Collision	Time (s)	Reward
SARL	5	0.95	0.03	10.57	0.2906
RCAT		0.98	0.01	9.43	0.3315
SARL	7	0.93	0.02	11.09	0.2784
RCAT		0.96	0.03	10.12	0.3164

5. 结论

本文针对拥挤人群环境下的机器人自主导航问题，提出了 RCAT 算法，通过串行时空编码与跨注意力机制实现了机器人中心的人群建模。实验结果显示：与 SARL 方法相比，RCAT 在不同规模的人群场景中均显著提升了导航性能，表现为更高的任务完成率、更低的碰撞风险以及更优的效率。这一成果表明，跨注意力机制能够有效强化机器人对关键个体的关注，从而在复杂环境下实现更安全与高效的路径规划。未来的研究可进一步探索在真实场景中的部署效果，并将 RCAT 扩展到三维动态环境、多机器人协作等更复杂的任务中。

基金项目

国家自然科学基金项目(No. 5247083536)；山东省自然科学基金项目(No. ZR2022MF345)。

参考文献

[1] Zhang, J. and Tao, D. (2021) Empowering Things with Intelligence: A Survey of the Progress, Challenges, and Opportunities in Artificial Intelligence of Things. *IEEE Internet of Things Journal*, **8**, 7789-7817. https://doi.org/10.1109/jiot.2020.3039359

[2] Mnih, V., *et al.* (2013) Playing Atari with Deep Reinforcement Learning. arXiv: 1312.5602.

[3] Hu, H., Zhang, K., Tan, A.H., Ruan, M., Agia, C.G. and Nejat, G. (2021) A Sim-To-Real Pipeline for Deep Reinforcement Learning for Autonomous Robot Navigation in Cluttered Rough Terrain. *IEEE Robotics and Automation Letters*, **6**, 6569-6576. https://doi.org/10.1109/lra.2021.3093551

[4] Mirowski, P., *et al.* (2017) Learning to Navigate in Complex Environments. arXiv: 1611.03673.

[5] Mirowski, P., *et al.* (2018) Learning to Navigate in Cities without a Map. *Advances in Neural Information Processing Systems*, **31**, 2419-2430.

[6] Han, R., Chen, S., Wang, S., Zhang, Z., Gao, R., Hao, Q., *et al.* (2022) Reinforcement Learned Distributed Multi-Robot Navigation with Reciprocal Velocity Obstacle Shaped Rewards. *IEEE Robotics and Automation Letters*, **7**, 5896-5903. https://doi.org/10.1109/lra.2022.3161699

[7] Josef, S. and Degani, A. (2020) Deep Reinforcement Learning for Safe Local Planning of a Ground Vehicle in Unknown Rough Terrain. *IEEE Robotics and Automation Letters*, **5**, 6748-6755. https://doi.org/10.1109/lra.2020.3011912

[8] Cimurs, R., Suh, I.H. and Lee, J.H. (2022) Goal-Driven Autonomous Exploration through Deep Reinforcement Learning. *IEEE Robotics and Automation Letters*, **7**, 730-737. https://doi.org/10.1109/lra.2021.3133591

[9] Mnih, V., *et al.* (2016) Asynchronous Methods for Deep Reinforcement Learning. *Proceedings of the International Conference on Machine Learning* (*ICML*), New York, 19-24 June 2016, 1928-1937.

[10] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., *et al.* (2015) Human-Level Control through Deep Reinforcement Learning. *Nature*, **518**, 529-533. https://doi.org/10.1038/nature14236

[11] Chen, Y.F., Liu, M., Everett, M. and How, J.P. (2017) Decentralized Non-Communicating Multiagent Collision Avoidance with Deep Reinforcement Learning. 2017 *IEEE International Conference on Robotics and Automation* (*ICRA*), Singapore, 29 May-3 June 2017, 285-292. https://doi.org/10.1109/icra.2017.7989037

[12] Chen, Y.F., Everett, M., Liu, M. and How, J.P. (2017) Socially Aware Motion Planning with Deep Reinforcement Learning. 2017 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Vancouver, 24-28 September 2017, 1343-1350. https://doi.org/10.1109/iros.2017.8202312

[13] Everett, M., Chen, Y.F. and How, J.P. (2018. Motion Planning among Dynamic, Decision-Making Agents with Deep Reinforcement Learning. 2018 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Madrid, 1-5 October 2018, 3052-3059. https://doi.org/10.1109/iros.2018.8593871

[14] Chen, C., Liu, Y., Kreiss, S. and Alahi, A. (2019) Crowd-Robot Interaction: Crowd-Aware Robot Navigation with Attention-Based Deep Reinforcement Learning. 2019 *International Conference on Robotics and Automation* (*ICRA*), Montreal, 20-24 May 2019, 6015-6022. https://doi.org/10.1109/icra.2019.8794134

[15] Liu, S., Chang, P., Liang, W., Chakraborty, N. and Driggs-Campbell, K. (2021) Decentralized Structural-RNN for Robot Crowd Navigation with Deep Reinforcement Learning. 2021 *IEEE International Conference on Robotics and Automation* (*ICRA*), Xi'an, 30 May-5 June 2021, 3517-3524. https://doi.org/10.1109/icra48506.2021.9561595

[16] Yang, Y., Jiang, J., Zhang, J., Huang, J. and Gao, M. (2023) ST^2: Spatial-Temporal State Transformer for Crowd-Aware Autonomous Navigation. *IEEE Robotics and Automation Letters*, **8**, 912-919. https://doi.org/10.1109/lra.2023.3234815