文章编号：1003－1251(2026)05－0041－07

基于改进双重深度Q网络的交通信号控制

史兰泽，文 峰

（沈阳理工大学 信息科学与工程学院，沈阳 110159）

摘 要：深度强化学习算法被广泛用于解决具有空间复杂性、时序随机性的交通信号控制问题。然而，现有深度强化学习算法中的神经网络提取交通特征的能力不足，难以全面解析复杂交通场景下的时空信息。为了更好地缓解交通拥堵压力，提出一种基于改进双重深度Q网络的交通信号控制算法。该算法以基于经验策略的深度强化学习算法为基础，通过在卷积神经网络中构建全局空间建模机制和时序记忆增强机制，深度挖掘路网空间关联信息，准确捕捉车流时序依赖关系，全面提取路网环境的空间特征和时间特征，从而输出高效的交通信号控制决策。在SUMO仿真平台上的实验结果表明，该算法可减少车辆等待数量、缩短车辆平均等待时间，有效提高交通系统运行效率。

关 键 词：深度强化学习；交通信号控制；卷积神经网络；改进双重深度Q网络

中图分类号：TP391.9 **文献标志码**：A **DOI**：10.3969/j.issn.1003－1251.2026.05.006

Traffic Signal Control Based on Enhanced Double DQN

SHI Lanze，WEN Feng

（Shenyang Ligong University，Shenyang 110159，China）

Abstract：Deep reinforcement learning algorithms are widely used to address traffic signal control problems with spatial complexity and temporal stochasticity. However，neural networks in existing deep reinforcement learning algorithms lack sufficient capability to extract traffic features，failing to fully analyze spatiotemporal information in complex traffic scenarios. To better alleviate congestion pressure，an enhanced Double Deep Q Network-based traffic signal control algorithm is proposed. This algorithm is built upon an empirical strategy based deep reinforcement learning algorithm，integrating a global spatial modeling mechanism and a temporal memory enhancement mechanism into the convolutional neural network，which deeply mines spatial correlations of road networks，accurately captures temporal dependencies of traffic flow，comprehensively extracts the spatial and temporal features of the road network environment，and therefore generates efficient signal control decisions. Experimental results on the SUMO simulation platform demonstrate that this algorithm reduces vehicle waiting counts and shortens the average waiting times，efficiently improving the operational efficiency of the traffic system.

Key words：deep reinforcement learning；traffic signal control；convolutional neural network；enhanced double deep Q Network

收稿日期：2025－08－21

基金项目：国家重点研发计划“社会治理与智慧社会科技支撑”重点专项(2022YFC3302502)

作者简介：史兰泽(1999—)，女，硕士研究生；文峰(1977—)，通信作者，男，教授，博士。

随着城市化的推进,驾车出行日益普及,但有限的道路资源难以满足车辆增长需求,导致城市交通拥堵[1]问题严重。交通拥堵现象不仅降低了城市交通系统运行效率,还带来了空气污染、能源浪费等一系列社会问题,已成为社会关注焦点。

相较于道路扩建、立交桥建设等基础设施改造方案,交通信号控制[2]的部署灵活、实现成本低、见效周期短,是提升交通系统运行效率的关键措施。

近年来,人工智能算法[3]在交通信号控制中的研究日益深入,深度强化学习算法同时具备深度学习的感知能力和强化学习的决策能力[4],为解决城市交通治理难题提供了技术支撑。Li 等[5]针对交叉口协同控制问题,提出了自适应多智能体深度混合强化学习算法;Gu 等[6]针对四相位信号灯的控制问题,提出了基于约束网络分区和自适应深度强化学习的大规模交通信号控制算法;Xu 等[7]为了解决智能体状态空间中数据缺失的问题,提出了基于图结构的深度强化学习交通信号控制算法;刘智敏等[8]通过采集实时车辆数建立奖励函数,提出了基于改进深度强化学习的交通信号控制算法。

然而,现有深度强化学习交通信号控制算法中的神经网络利用远距离空间位置的关联信息进行建模的能力有限,处理长时序依赖的能力也有待提升。为了解决上述问题,进一步缓解交通拥堵压力,本文提出一种基于改进双重深度 Q 网络的交通信号控制算法。算法以基于经验策略的深度强化学习算法[9]为基础,通过在卷积神经网络中构建全局空间建模机制和时序记忆增强机制,增强深度强化学习算法在复杂交通环境下的鲁棒性,从而实现高效的交通信号控制。本文算法的主要内容如下。

1)构建全局空间建模机制。在卷积神经网络中引入自适应稀疏自注意力机制(adaptive sparse self-attention,ASSA)[10],利用该注意力的双分支结构,在过滤低匹配度特征的同时保留关键信息,从而提升卷积神经网络利用远距离空间位置的关联信息进行建模的能力。

2)设计时序记忆增强机制。在卷积神经网络中集成门控循环单元(gated recurrent unit,GRU)单元[11],利用单元的门控机制动态调节时序信息流,弥补传统卷积神经网络处理长时序依赖能力弱的缺陷。

1 基于改进双重深度 Q 网络的交通信号控制算法

1.1 状态定义

交通路网的空间结构与图像数据不同,缺乏规则化排列特性。当使用卷积神经网络处理交通道路数据时,需要先将道路状态信息转化为有序的空间形式,把路网划分成多个存在交叠区域的感受野。这种交叠式划分既能确保各感受野之间的空间关联性,又能维持算法对路网全局状态的感知能力。

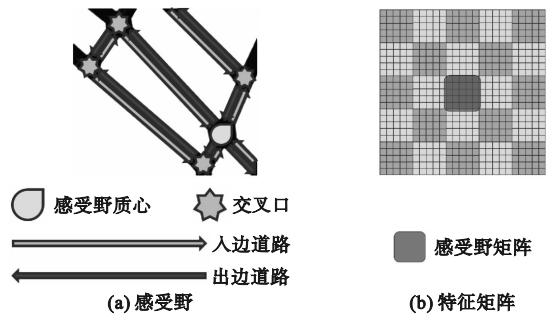


图 1 感受野及特征矩阵

Fig. 1 Receptive field and feature matrix

本文通过多个步骤对路网进行预处理以获得可直接用于输入卷积神经网络的状态,具体步骤如下。

1)确立感受野质心。质心为感受野内的中心交叉口,为实现路网中感受野的规则化排布,采用 K-means 聚类算法[12]完成质心的布局。

2)划分感受野。基于已确定的质心集合,利用广度优先搜索算法[13]遍历路网中所有交叉口的入边道路与出边道路,构建 25 个具有空间交叠特性的感受野,感受野的结构如图 1(a)所示。每个感受野由 25 对双向道路组成,每对道路都包含入边道路和出边道路。

3)进行感受野数据的结构化转换。为适配卷积神经网络的输入要求,将感受野中的道路数据规范化为感受野矩阵,通过赋予特征数据线性顺序,实现从路网拓扑空间到有序向量空间的转换。

4)计算特征矩阵。结合交通路网实时采集的数据,依据道路信息计算入边道路车辆密度、出边道路车辆密度、入边道路平均速度、出边道路平均速度四类特征矩阵,形成可直接输入卷积神经网络的状态,感受野矩阵与特征矩阵的关系如图 1(b)所示。

车辆密度 d 的计算公式为

$$d = mh/ql \tag{1}$$

车辆平均速度 S_{peed} 的计算公式为

$$S_{\text{peed}} = ml/t_m \tag{2}$$

式中:q 为每条道路的车道数;h 为每辆车的长度;l 为道路的长度;m 为 t_m 时间间隔内通过每条道路的车辆数。

1.2　动作定义

交通路网由道路与交叉口相互连接构成[14],相位决定交通流的走向。交通信号相位指在固定的信号时长内,特定方向的车流获得通行权,而其他方向的车流处于等待状态。

合理设置含信号灯的交叉口(信号交叉口)的相位对提高路网通行效率至关重要。相位持续的时间为一段固定的信号时长。本文中 0 相位代表车流在信号交叉口处东西向可通行、南北向禁止通行;1 相位代表车流在信号交叉口处南北向可通行、东西向禁止通行。当前相位结束后,系统依据实时交通状况自适应选择下一个相位,以提升整体交通流的通行效率。

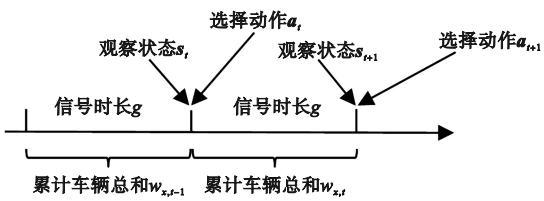


图 2　交通信号变换示意图

Fig. 2　Schematic diagram of traffic signals transformation

图 2 为交通信号变换示意图,在时间步 t,智能体根据当前路网状态 $\boldsymbol{s}_t$ 选择 $\boldsymbol{a}_t$ 动作,动作为由路网中所有信号交叉口相位组成的序列。动作执行固定信号时长 g 后,在下一个时间 $t+1$ 步,智能体观察新的状态 $\boldsymbol{s}_{t+1}$ 并选择新动作 $\boldsymbol{a}_{t+1}$。

1.3　奖励定义

在深度强化学习算法中,环境需向智能体反馈其过往行为的奖励[15]与状态信息,通过动态反馈引导智能体做出优化决策。将奖励定义为路网中所有信号交叉口 x 及其入边道路、出边道路中的车辆在相邻信号时长 g 内累计数的变化量序列。

信号交叉口 x 在第 t 个时间步的奖励 $r_{x,t}$ 为

$$r_{x,t} = w_{x,t-1} - w_{x,t} \tag{3}$$

式中:$w_{x,t-1}$ 和 $w_{x,t}$ 分别为信号交叉口 x 及其入边道路、出边道路中的车辆在时间步 $t-1$ 和 t 内的数量总和。当上一时间步的车辆数量总和大于当前时间步的车辆数量总和时,智能体获得正向奖励,反之,智能体受到负向奖励。

在时间步 t,奖励序列 $\boldsymbol{r}_t$ 用来指导不同的信号交叉口做出更明智的动作。

1.4　基于改进双重深度 Q 网络的交通信号控制算法

本文在基于经验策略的深度强化算法的基础上,提出一种基于改进双重深度 Q 网络的交通信号控制算法,通过在卷积神经网络中构建全局空间建模机制和时序记忆增强机制,提升卷积神经网络对远距离空间关联和长时序依赖的建模能力,增强算法在复杂环境下的鲁棒性。算法的流程如图 3 所示,算法中卷积神经网络的结构如图 4 所示。

为了避免因数据不足、参数随机导致的训练震荡,将实验分为预训练和正式训练两个阶段。算法的动作决策和参数更新过程如下。

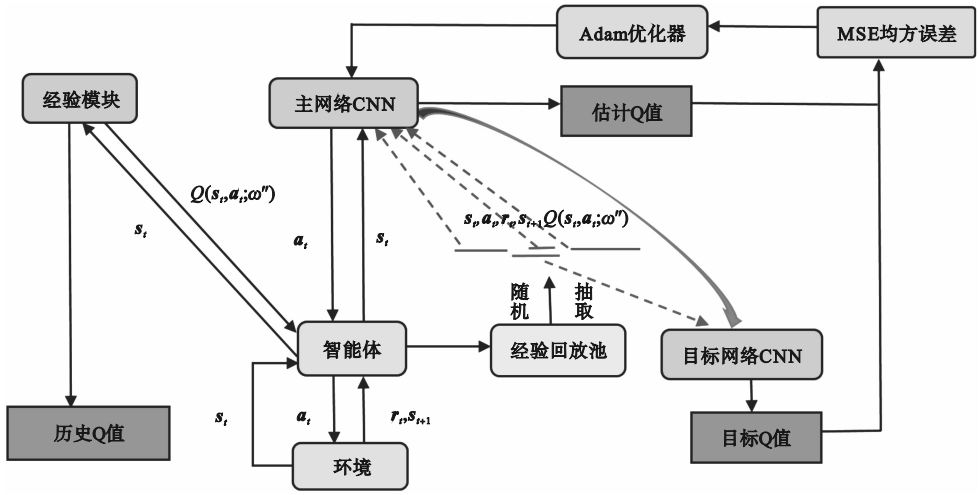
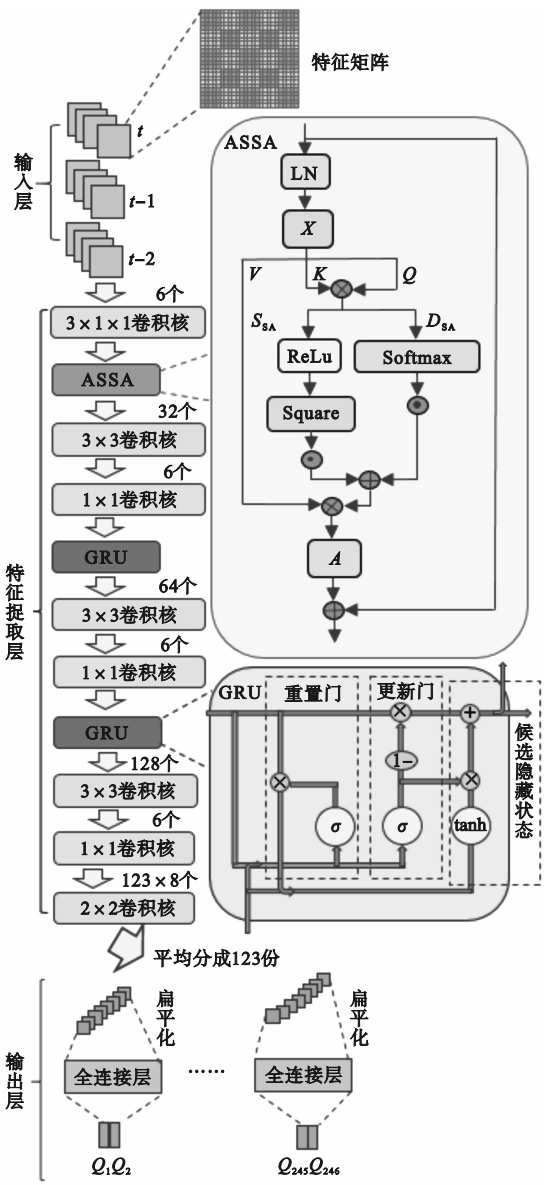


图 3　交通信号控制算法

Fig. 3　Traffic signals control algorithm

1.4.1 动作决策

在预训练阶段,算法在每个时间步 t 随机输出动作 $\boldsymbol{a}_t$ 和历史 Q 值;进入正式训练阶段后,算法将当前状态 $\boldsymbol{s}_t$ 输入主网络,经训练后主网络通过贪婪策略的探索 - 利用机制输出动作 $\boldsymbol{a}_t$ 到环境,环境反馈奖励 $\boldsymbol{r}_t$,与此同时,算法将 $\boldsymbol{s}_t$ 输入经验网络和目标网络,经训练后经验网络和目标网络分别输出历史 Q 值和目标网络 Q 值。算法将每个时间步 t 的状态 $\boldsymbol{s}_t$、动作 $\boldsymbol{a}_t$、奖励 $\boldsymbol{r}_t$、历史 Q 值及其下一时间步 $t+1$ 的状态 $\boldsymbol{s}_{t+1}$ 存储至经验回放池。



注:⊙为逐元素相乘运算;⊕为残差连接运算;
⊗为矩阵乘法运算

图 4 卷积神经网络结构

Fig. 4 Convolutional neural network architecture

在正式训练阶段,算法中的主网络、目标网络、经验网络模块使用结构相同的卷积神经网络对路网环境的空间特征和时间特征进行提取。网络由输入层、特征提取层、输出层构成,如图 4 所示。

输入层的状态为来自 3 个连续时间片段的 12 个特征矩阵,每个时间片段中的特征矩阵收集了所有感受野内交叉口入边道路的车辆密度、出边道路的车辆密度、入边道路的车辆平均速度和出边道路的车辆平均速度 4 类数据信息。

特征提取层是卷积神经网络的关键部分,用于深度挖掘路网中的时空信息。针对卷积操作易导致的空间特征割裂问题,本文在该层引入 ASSA 构建全局空间建模机制。

全局空间建模机制对经过归一化(LN)运算的特征 $\boldsymbol{X}$ 进行窗口分割并生成查询矩阵 $\boldsymbol{Q}$、键矩阵 $\boldsymbol{K}$、值矩阵 $\boldsymbol{V}$,再通过双分支注意力融合关键路网信息得到特征矩阵 $\boldsymbol{A}$。计算过程如下。

$$\boldsymbol{Q}=\boldsymbol{X}\boldsymbol{W}_{Q},\boldsymbol{K}=\boldsymbol{X}\boldsymbol{W}_{K},\boldsymbol{V}=\boldsymbol{X}\boldsymbol{W}_{V} \tag{4}$$

$$\boldsymbol{D}_{\mathrm{SA}}=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}+\boldsymbol{B}\right) \tag{5}$$

$$\boldsymbol{S}_{\mathrm{SA}}=\left[\mathrm{ReLu}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}+\boldsymbol{B}\right)\right]^{2} \tag{6}$$

$$\boldsymbol{A}=(\omega_1\boldsymbol{S}_{\mathrm{SA}}+\omega_2\boldsymbol{D}_{\mathrm{SA}})\boldsymbol{V} \tag{7}$$

式中:$\boldsymbol{W}_Q$、$\boldsymbol{W}_K$、$\boldsymbol{W}_V$ 为可学习的权重矩阵;$\boldsymbol{B}$ 代表可学习的相对位置偏差;d 为特征维度;ω_1 和 ω_2 是用于自适应调节双分支的归一化权重;$\boldsymbol{D}_{\mathrm{SA}}$为密集注意力矩阵;$\boldsymbol{S}_{\mathrm{SA}}$为稀疏注意力矩阵。特征矩阵 $\boldsymbol{A}$ 经窗口合并后与原特征进行残差连接,形成了强化空间关联的特征。

感知交通状态是时空交织的复杂过程,除了挖掘路网空间关联信息外,捕捉时间特征的长期依赖关系对精准刻画路网环境同样关键。

尽管输入数据已经包含了 3 个历史时间片段的交通特征矩阵,能够提供一定的时间信息,但仅依靠卷积层无法有效捕捉到时间序列中的长期依赖关系。针对长时序依赖信息缺失问题,本文在卷积神经网络的特征提取层引入 GRU 构建时序记忆增强机制。

时序记忆增强机制使用两层 GRU 捕捉交通状态时序关联。GRU 的重置门筛选历史信息,更新门调节筛选后的历史信息与当前输入特征的融合权重,候选隐藏状态根据融合权重生成新特征,更新门再融合历史信息与候选隐藏状态,得到输出特征。计算过程如下。

$$\boldsymbol{r}_t=\sigma(\boldsymbol{W}_{\mathrm{r}}[\boldsymbol{x}_t,\boldsymbol{h}_{t-1}]+\boldsymbol{b}_{\mathrm{r}}) \tag{8}$$

$$\boldsymbol{z}_t=\sigma(\boldsymbol{W}_{\mathrm{z}}[\boldsymbol{x}_t,\boldsymbol{h}_{t-1}]+\boldsymbol{b}_{\mathrm{z}}) \tag{9}$$

$$\widetilde{\boldsymbol{h}}_t=\tanh(\boldsymbol{W}_{\mathrm{h}}[\boldsymbol{r}_t\odot\boldsymbol{x}_t,\boldsymbol{h}_t-1]+\boldsymbol{b}_{\mathrm{h}}) \tag{10}$$

$$\boldsymbol{h}_t=(1-z_t)\odot\boldsymbol{h}_{t-1}+\boldsymbol{z}_t\odot\widetilde{\boldsymbol{h}}_t \tag{11}$$

式中:σ 为 Sigmoid 激活函数;$\odot$ 为按元素相乘操作;$\boldsymbol{x}_t$ 为 t 时刻的输入数据;$\boldsymbol{h}_{t-1}$ 和 $\boldsymbol{h}_t$ 为 $t-1$ 和 t 时刻的隐藏状态;$\boldsymbol{W}_{\mathrm{r}}$、$\boldsymbol{W}_{\mathrm{z}}$、$\boldsymbol{W}_{\mathrm{h}}$ 为权重矩阵,分别用于计算 t 时刻的重置门 $\boldsymbol{r}_t$、更新门 $\boldsymbol{z}_t$、候选隐藏状态;$\widetilde{\boldsymbol{h}}_t$、$\boldsymbol{b}_{\mathrm{r}}$、$\boldsymbol{b}_{\mathrm{z}}$、$\boldsymbol{b}_{\mathrm{h}}$ 为偏置项。

经过全局空间建模机制与时序记忆增强机制的处理,特征提取层输出的特征已充分融合了时空信息。输出层将这些特征转换为 123 组 Q 值,每组 Q 值包含两个 Q 值。

主网络输出的每组 Q 值中,一个 Q 值对应 0 相位,另一个 Q 值对应 1 相位,每个信号交叉口选择较大 Q 值对应的相位,123 个信号交叉口相位组成的序列即为路网环境执行的实际动作。

目标网络和经验模块不直接参与决策,输出的 Q 值用于更新参数和最小化目标函数,使主网络做出更准确的决策。

1.4.2　参数迭代

交通信号控制算法中的主网络、目标网络以及经验模块卷积神经网络的结构及初始参数设置均相同,权重参数通过不同的方法进行迭代[16]。

在正式训练阶级,算法从经验回放池中随机采样,将时间步 t 的主网络输出值与历史 Q 值 $\boldsymbol{Q}_{\mathrm{h}}$ 进行平均得到估计 Q 值,提取实际动作对应的估计 Q 值。利用时间步 $t+1$ 的目标网络输出 Q 值、历史 Q 值和时间步 t 的奖励 $\boldsymbol{r}_t$ 计算目标 Q 值 $\boldsymbol{y}$,计算过程为

$$\boldsymbol{a}_{\mathrm{m}}=\underset{\boldsymbol{a}}{\operatorname{argmax}}\ Q(\boldsymbol{s}_{t+1},\boldsymbol{a};\boldsymbol{\omega}) \tag{12}$$

$$y=\boldsymbol{r}_t+\gamma[0.5Q(\boldsymbol{s}_{t+1},\boldsymbol{a}_{\mathrm{m}};\boldsymbol{\omega}')+0.5Q(\boldsymbol{s}_{t+1},\boldsymbol{a}_{\mathrm{m}};\boldsymbol{\omega}'')] \tag{13}$$

式中:$Q(\cdot)$ 为状态-动作对的价值函数;$\boldsymbol{a}_{\mathrm{m}}$ 为主网络选择的下一时间步 $t+1$ 价值最大的动作;γ 为折扣因子,值为 0.99;$\boldsymbol{s}_{t+1}$ 为下一个时间步的状态;$\boldsymbol{a}$ 为样本动作空间中的任意动作;$\boldsymbol{\omega}$ 为主网络的权重;$\boldsymbol{\omega}'$为目标网络的权重;$\boldsymbol{\omega}''$为经验网络的权重。

算法通过实际动作对应的估计 Q 值与目标 Q 值计算出均方误差(MSE)M_{SE},计算公式为

$$M_{\mathrm{SE}}(x_{i,j},y_{i,j})=\frac{\sum_{i=1}^{123}\sum_{j=1}^{32}(x_{i,j}-y_{i,j})^2}{123\times32} \tag{14}$$

式中:123 为交通路网中信号交叉口的数量;32 为每次采样的样本数;$x_{i,j}$为第 j 个样本中第 i 个信号交叉口对应实际动作的估计 Q 值;$y_{i,j}$为第 j 个样本中第 i 个信号交叉口的目标 Q 值。$M_{\mathrm{SE}}(x_{i,j},y_{i,j})$为算法的第一个目标函数,算法使用 Adam 优化器最小化目标函数以更新主网络参数。

目标网络每隔 4 秒通过指数加权平均的方式学习主网络参数,参数更新的方式为

$$\boldsymbol{\omega}'\leftarrow(1-\alpha)\boldsymbol{\omega}'+\alpha\boldsymbol{\omega} \tag{15}$$

式中:α 为平滑系数,值为 0.001。

算法提取样本中第 t 个时间步的实际动作对应的历史 Q 值 $Q(\boldsymbol{s}_t,\boldsymbol{a}_t;\boldsymbol{\omega}'')$用于计算经验模块的目标值 y',公式为

$$y'=(1-\eta)Q(\boldsymbol{s}_t,\boldsymbol{a}_t;\boldsymbol{\omega}'')+\eta Q(\boldsymbol{s}_t,\boldsymbol{a}_t;\boldsymbol{\omega}) \tag{16}$$

式中:η 为衰减系数,值为 0.05。算法利用经验模块的目标值计算出损失函数 M'_{SE},公式为

$$M'_{\mathrm{SE}}(y'_{u,v},y''_{u,v})=\frac{\sum_{u=1}^{123}\sum_{v=1}^{32}(y'_{u,v}-y''_{u,v})^2}{123\times32} \tag{17}$$

式中:$y'_{u,v}$为第 j 个样本中第 i 个交叉口对应的经验模块目标值;$y''_{u,v}$为第 j 个样本中第 i 个信号交叉口对应实际动作的历史 Q 值。$M_{\mathrm{SE}}(y'_{u,v},y''_{u,v})$为算法的第二个目标函数,算法使用 Adam 优化器最小化目标函数以更新经验模块的参数值。

1.5　算法复杂度分析

设 N 为单个时间步下,单个输入特征图的空间位置总数,为可变量。

算法的时间复杂度为 $O(N^2)$。ASSA 注意力处理的是全局特征,其矩阵乘法的维度与单个特征图尺寸一致,时间复杂度为 $O(N^2)$;卷积操作的计算量与 N 成正比,时间复杂度为 $O(N)$;从缓冲区中抽取固定大小的样本,该操作的时间复杂度为 $O(1)$;全连接层的输入维度与输出维度固定,时间复杂度为 $O(1)$;GRU 的输入序列长度与状态维度固定,时间复杂度为 $O(1)$。因此算法整体时间复杂度为 $O(N^2)$。

算法的空间复杂度为 $O(N^2)$。ASSA 注意力计算所有空间位置对的注意力权重,需存储注意力矩阵,空间复杂度为 $O(N^2)$;各层卷积输出的特征所占空间与空间位置数 N 成比例,空间复杂度为 $O(N)$;在经验回放池中,样本数量固定,每个样本的状态大小与 N 成正比,缓存样本的空间复杂度为 $O(N)$;全连接层的输入、输出维度固定,空间复杂度为 $O(1)$;GRU 的参数为固定大小,空间复杂度为 $O(1)$。因此算法整体空间复杂度为 $O(N^2)$。

本研究中,N 的值设为 625,鉴于现代 CPU/

GPU 并行计算架构的强大处理能力，N^2 的绝对运算量级相对较小，该复杂度在现实场景下具有充分的计算可行性。

2 仿真实验及指标分析

2.1 实验过程

本文使用 SUMO 平台[17]进行实验。仿真实验的路网由 160 个交叉口组成，其中 123 个交叉口含交通信号灯。路网感受野数量为 25，单个感受野大小为 25。算法中经验池容量为 500，每次随机采样 32 个样本，衰减率 γ 为 0.99，Adam 优化器学习率为 0.001。实验预训练时间为 5 000 s，正式训练时间为 15 000 s。固定信号时长为 10 s，路网中共 11 个出车对，各出车对有固定起止点，起点以 6.67 s/辆的固定频率发车。

在仿真实验中，采用两个指标评估算法的性能。第一个指标是车辆平均等待时间，其计算方法是从车辆进入交叉口的入边道路时开始计时，直到其完全驶离出边道路为止，记录所有车辆的通行时间并计算平均值，该指标能够反映车辆在路网中的整体通行效率。第二个指标是车辆等待数，即在固定的信号周期内统计整个交通路网中处于等待状态的车辆总数，该指标可用于衡量路网的拥堵程度。

2.2 对比实验分析

为了验证本文算法（用“OURS”表示）的可行性和有效性，将 OURS 与基于 SENet 的交通信号控制算法[18]（用“S-EDDQN”表示）以及基于 CBAM 注意力机制的智能交通信号控制算法[19]（用“C-EDDQN”表示）进行对比实验，实验采用相同的参数，实验结果如图 5、图 6、表 1 所示。

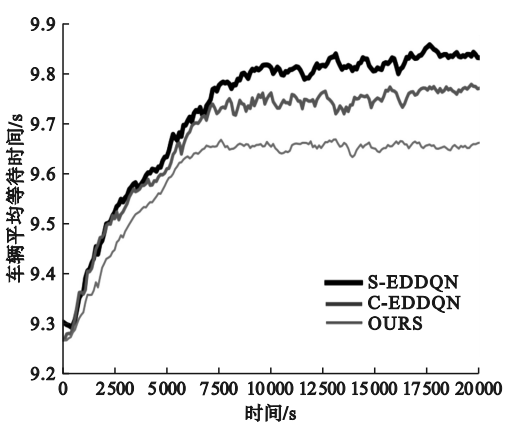


图 5 车辆平均等待时间

Fig. 5 Average waiting time of vehicle

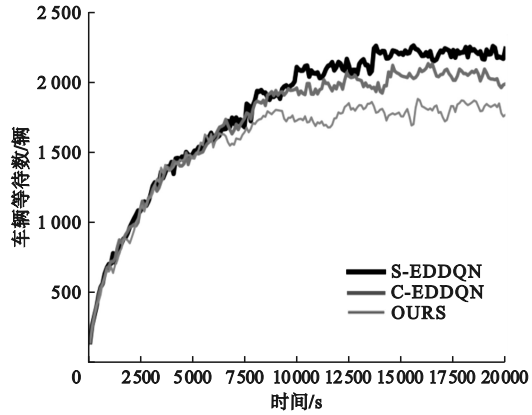


图 6 车辆等待数

Fig. 6 Number of waiting vehicles

表 1 对比实验结果

Table 1 Results of comparative experiments

算法	车辆平均等待时间/s	车辆等待数/辆
S-EDDQN	9.72	1 788
C-EDDQN	9.68	1 705
OURS	9.60	1 564

根据图 5 和图 6 的实验结果可知，OURS 在车辆平均等待时间和车辆等待数两个指标上均优于对比算法。由表 1 可知，与 S-EDDQN 相比，OURS 使得车辆平均等待时间缩短了 0.04 s，车辆等待数减少了 83 辆；与 C-EDDQN 相比，OURS 使得车辆平均等待时间缩短了 0.08 s，车辆等待数减少了 141 辆。

2.3 消融实验分析

本文以基于经验策略的深度强化学习算法（用“EDDQN”表示）为基线算法，依次添加 ASSA 机制和 GRU 机制，以相同的实验参数完成消融实验，结果如图 7、图 8 所示，具体数据如表 2 所示。

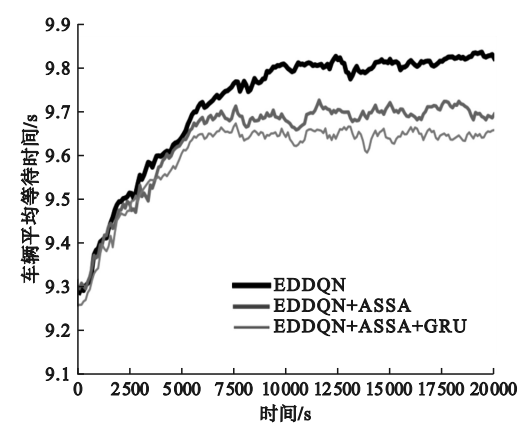


图 7 消融实验车辆平均等待时间

Fig. 7 Average waiting time of vehicles for ablation experiments

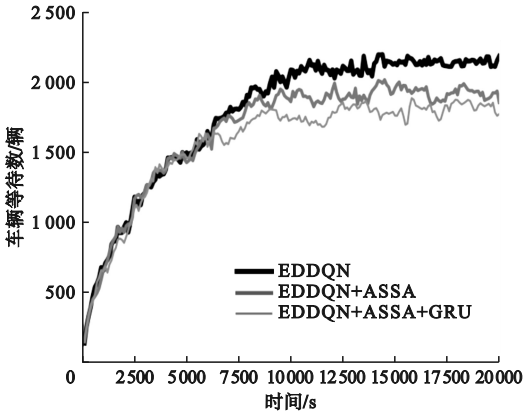


图 8　消融实验车辆等待数

Fig. 8　Number of waiting vehicles for ablation experiments

表 2　消融实验结果

Table 2　Results of the ablation experiments

算法	车辆平均等待时间/s	车辆等待数/辆
EDDQN	9.72	1 770
EDDQN + ASSA	9.64	1 651
EDDQN + ASSA + GRU	9.60	1 564

由图 7 和图 8 的实验结果可知,在交通信号控制算法中依次添加 ASSA 机制和 GRU 机制后,算法表现出了优越性和稳定性。表 2 的数据显示,全局空间建模机制的设计使得车辆平均等待时间缩短了 0.08 s,车辆等待数减少了 119 辆;在全局空间建模机制的基础上,时序记忆增强机制的设计使得车辆平均等待时间缩短了 0.04 s,车辆等待数减少了 87 辆。

3　结论

为了缓解交通拥堵压力,本文提出了一种基于改进双重深度 Q 网络的交通信号控制算法。该算法利用全局空间建模机制和时序记忆增强机制提升卷积神经网络对时空特征的提取能力。实验结果表明,本文算法可有效减少车辆等待数、缩短车辆平均等待时间,提高了交通系统运行效率。

参考文献(References):

[1] Pi M Y, Yeon H, Son H, et al. Visual cause analytics for traffic congestion[J]. IEEE Transactions on Visualization and Computer Graphics, 2021, 27(3): 2186 – 2201.

[2] Zhao D B, Dai Y J, Zhang Z. Computational intelligence in urban traffic signal control: a survey[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2012, 42(4): 485 – 494.

[3] Jin Q. Automatic control of traffic lights at multiple intersections based on artificial intelligence and ABST light[J]. IEEE Access, 2024, 12: 103004 – 103017.

[4] Rasheed F, Yau K A, Noor R M, et al. Deep reinforcement learning for traffic signal control: a review[J]. IEEE Access, 2020, 8: 208016 – 208044.

[5] Li L L, Zhu R J, Wu S N, et al. Adaptive multi-agent deep mixed reinforcement learning for traffic light control[J]. IEEE Transactions on Vehicular Technology, 2024, 73(2): 1803 – 1816.

[6] Gu H K, Wang S B, Ma X G, et al. Large-scale traffic signal control using constrained network partition and adaptive deep reinforcement learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(7): 7619 – 7632.

[7] Xu D W, Yu Z F, Liao X W, et al. A graph deep reinforcement learning traffic signal control for multiple intersections considering missing data[J]. IEEE Transactions on Vehicular Technology, 2024, 73(12): 18307 – 18319.

[8] 刘智敏,叶宝林,朱耀东,等. 基于深度强化学习的交通信号控制方法[J]. 浙江大学学报(工学版),2022,56(6):1249 – 1256.
Liu Z M, Ye B L, Zhu Y D, et al. Traffic signal control method based on deep reinforcement learning[J]. Journal of Zhejiang University (Engineering Science), 2022, 56(6): 1249 – 1256. (in Chinese)

[9] 黄思萌. 基于经验策略深度强化学习的交通灯配时优化技术研究[D]. 沈阳:沈阳理工大学,2023.

[10] Yang C, Gao Q X, Chen G H, et al. DBD-CR: dual-branch diffusion residual reconstruction for cloud removal in optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1 – 13.

[11] Yang Y H, Zhao J F, Du H F, et al. Shear-wave velocity prediction by CNN-GRU fusion network based on the self-attention mechanism[J]. IEEE Geoscience and Remote Sensing Letters, 2025, 22: 1 – 5.

[12] Ikotun A M, Ezugwu A E, Abualigah L, et al. K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data[J]. Information Sciences, 2023, 622: 178 – 210.

[13] Tripathy H K, Mishra S, Thakkar H K, et al. CARE: a collision-aware mobile robot navigation in grid environment using improved breadth first search[J]. Computers & Electrical Engineering, 2021, 94: 107327.

[14] 文峰,张可新. 基于深度强化学习的交通信号配时优化研究[J]. 沈阳理工大学学报,2019,38(1):48 – 52,63.
Wen F, Zhang K X. Research on optimization of traffic signal timing based on deep reinforcement learning[J]. Journal of Shenyang Ligong University, 2019, 38(1): 48 – 52, 63. (in Chinese)

[15] 杨思明,单征,丁煜,等. 深度强化学习研究综述[J]. 计算机工程,2021,47(12):19 – 29.
Yang S M, Shan Z, Ding Y, et al. Survey of research on deep reinforcement learning[J]. Computer Engineering, 2021, 47(12): 19 – 29. (in Chinese)

[16] 刘思萌,文峰. 交通灯配时与车辆诱导协同控制技术研究[J]. 沈阳理工大学学报,2025,44(4):43 – 49.
Liu S M, Wen F. Research on coordinated control of traffic signal timing and vehicle guidance[J]. Journal of Shenyang Ligong University, 2025, 44(4): 43 – 49. (in Chinese)

(下转第 54 页)