

چارچوبی جامع مبتنی بر یادگیری تقویتی برای کنترل سیگنالهای ترافیکی در محیط های V2X  
تهیه و تنظیم: محمدامین خدادادی - ۴۰۰۱۲۳۴۰۱۱۹۱۶۲ - دکتر مهدی اسلامی - درس شبکه های مخابراتی  
نویسندگان:

Wang, Xi Xiong, Yuheng Kan, Chengcheng Xu, Man On Pun

IEEE Transactions on Vehicular Technology, 2024. DOI / (IEEE Xplore via DOI):  
<https://doi.org/10.1109/TVT.2024.3403879>

impact factor:6.1-8.6

### چکیده

ترافیک متراکم، به یک چالش مزمن و پرهزینه در مناطق شهری تبدیل شده است که ضرورت توسعه سیستم های مؤثر کنترل سیگنال ترافیکی (TSC) را بیش از پیش نمایان می سازد. در این زمینه، رویکردهای مبتنی بر یادگیری تقویتی (RL) پتانسیل بالایی برای بهینه سازی عملکرد TSC از طریق تعامل پویا با محیط ترافیکی نشان داده اند. با این حال، آموزش مؤثر این عامل ها در محیط های پیچیده و پویا، با چالش های متعددی از جمله نیاز به حجم عظیمی از داده های آموزشی متنوع و کیفیت پایین تعمیم پذیری به سناریوهای نادیده مواجه است.

در پاسخ به این چالش ها، این مقاله یک چارچوب نوآورانه و مقیاس پذیر مبتنی بر یادگیری تقویتی برای محیط های کنترل ترافیک با پشتیبانی از فناوری وسیله نقلیه به همه چیز (V2X) ارائه می دهد. هسته اصلی این چارچوب، طراحی یک معماری شبکه ی عصبی برای عامل (Agent) است که ماتریس تقاطع (Junction Matrix) را به عنوان یک بازنمایی غنی و ساختاریافته از حالت ترافیک پردازش می کند. این بازنمایی، اطلاعات کلیدی تقاطع ها، جریان های ترافیکی و ارتباطات بین آنها را در بر می گیرد و به عامل امکان می دهد تا تصمیمات بهینه تری برای زمان بندی فازهای چراغ راهنمایی اتخاذ نماید. طراحی ماتریس محور موجب می شود مدل پیشنهادی به راحتی به تقاطع های متنوع با توپولوژی های مختلف چندراهه،  $T$  شکل و غیره تعمیم یابد.

به منظور افزایش قابلیت تعمیم پذیری و مقابله با کمبود داده در سناریوهای ترافیکی پیچیده و نادر، در این چارچوب فوق اشاره از روش های افزایش داده Data Augmentation ویژه ای برای حالت های ترافیکی استفاده شده است. این تکنیک ها با ایجاد تغییرات ساختاریافته در داده های شبیه سازی (مانند تغییر الگوهای جریان ترافیک، نرخ ورودی خودروها یا پیکربندی های تقاطع)، فضای حالت آموزشی عامل را به طور مصنوعی گسترش داده و آن را برای مقابله با شرایط غیرمنتظره در دنیای واقعی مقاوم تر می سازند.

دسترسی به کد و داده ها: کد منبع کامل این پژوهش، مجموعه داده های شبیه سازی شده و دستورالعمل های لازم برای تکرار آزمایش ها، به صورت عمومی در مخزن گیت هاب زیر در دسترس است:

[https://github.com/wmn7/Universal\\_Light](https://github.com/wmn7/Universal_Light)

واژگان کلیدی: کنترل سیگنال ترافیکی \* مدل های مفهومی \* یادگیری تقویتی \*

• ترجمه شده توسط محمدامین خدادادی

<sup>1</sup> Traffic Signal Control

<sup>2</sup> Reinforcement learning

<sup>3</sup> Vehicle to everything

تراکم ترافیک به عنوان یک چالش حیاتی در مناطق شهری، پیامدهای گسترده‌ای از جمله اتلاف زمان شهروندان، افزایش مصرف سوخت و انتشار بیش از حد گازهای گلخانه‌ای به همراه دارد. [۱] رویکردهای سنتی کنترل ترافیک، مانند سیستم‌های چراغ‌های ثابت [۲] یا سامانه‌های مبتنی بر زمان‌بندی تطبیقی (مانند **SOTL**) [۳، ۴]، اگرچه تا حدودی مؤثر بوده‌اند، اما با رشد شهرها و افزایش پویایی شرایط جاده‌ها، اغلب در مواجهه با حجم بالای ترافیک ناکافی ظاهر می‌شوند. [۵]

ظهور فناوری‌های **V2X** (ارتباط وسایل نقلیه با همه چیز) در اکوسیستم‌های حمل‌ونقل هوشمند، نویدبخش راه‌حل‌های نوینی است. این فناوری امکان برقراری ارتباط و تبادل داده‌های بلادرنگ بین وسایل نقلیه، زیرساخت‌های جاده‌ای (مانند چراغ‌های راهنمایی) و سایر کاربران راه را فراهم می‌کند. [۶] با اتکا بر این داده‌های زنده و دقیق، مراکز مدیریت ترافیک می‌توانند حرکت وسایل نقلیه و عابران پیاده را در تقاطع‌ها، به صورت پویا و بر اساس شرایط لحظه‌ای جریان ترافیک، تنظیم و بهینه‌سازی کنند. [۷]

در این راستا، الگوریتم‌های مبتنی بر یادگیری تقویتی (**RL**) برای کنترل هوشمند چراغ‌های راهنمایی با توجه به شرایط لحظه‌ای ترافیک، پیشنهاد شده‌اند. این الگوریتم‌ها عمدتاً از سه رویکرد اصلی برای تنظیم چراغ‌ها استفاده می‌کنند:

۱. انتخاب فاز بعدی: در این روش، عامل **RL**، فاز بعدی چراغ را مستقیماً تعیین می‌کند و اجازه می‌دهد توالی فازها به جای ثابت بودن، کاملاً انعطاف‌پذیر باشد. [۸-۲۲] با این حال، اگرچه این رویکرد انعطاف‌پذیری بالایی ایجاد می‌کند، ممکن است برای رانندگان غیرقابل پیش‌بینی یا گیج‌کننده به نظر برسد، زیرا انتخاب فاز می‌تواند ظاهراً تصادفی باشد و این امر بالقوه می‌تواند منجر به افزایش خطاهای رانندگی و حوادث شود.

۲. نگه‌داشتن یا تغییر فاز: در این رویکرد، عامل **RL** تصمیم می‌گیرد که فاز کنونی ترافیک را حفظ کند یا به فاز بعدی

تغییر دهد. [۲۳-۲۵] این روش، پایداری بیشتری برای رانندگان فراهم می‌کند.

۳. تنظیم مدت زمان فاز: در این روش، عامل **RL**، طول مدت فاز جاری را از میان گزینه‌های از پیش تعیین‌شده انتخاب می‌کند. [۲۶-۲۸] از طریق تعامل مستقیم با محیط (شرایط ترافیکی)، عامل **RL** یاد می‌گیرد که بر اساس تجربیات و پاداش‌های دریافتی، تصمیمات خود را با تغییرات شرایط ترافیک وفق دهد.

علیرغم پیشرفت‌های قابل توجه در روش‌های مبتنی بر **RL**، یک محدودیت کلیدی این است که بسیاری از این مدل‌ها برای تقاطع‌های خاص طراحی شده‌اند. به عبارت دیگر، این مدل‌های **RL** باید هر بار که با تقاطعی جدید مواجه می‌شویم که از نظر تعداد خطوط، هندسه و فازبندی متفاوت است، از نو طراحی و آموزش داده شوند. این فرآیند نیازمند صرف زمان و منابع محاسباتی قابل توجهی برای جمع‌آوری داده، آموزش مدل و آزمایش است. [۲۹]

لذا، توسعه مدل‌های عام و قابل انتقال که بتوان آن‌ها را با حداقل تنظیمات در انواع مختلف تقاطع‌ها به کار گرفت، از اهمیت بالایی برخوردار است. چنین مدل‌هایی این قابلیت را دارند تا پیاده‌سازی راهکارهای مبتنی بر **V2X** را با نیاز به استقرار سریع و مقیاس‌پذیر در کل شهر، به طور مؤثرتری ممکن سازند و نیاز به طراحی مجدد برای هر تقاطع را برطرف کنند. [۳۰] این امر گامی اساسی در جهت تحقق شهرهای هوشمند و مدیریت یکپارچه ترافیک شهری به شمار می‌رود.

## ادبیات موجود و خلأ پژوهشی

در ادبیات موضوع، مدل‌های یادگیری تقویتی متعددی برای کنترل تقاطع‌های مختلف پیشنهاد شده‌اند [۱۰-۱۲، ۱۸]. با این حال، یک محدودیت اساسی این مدل‌های آموزش‌دیده، وابستگی شدید آنها به پیکربندی‌های خاصی است که در حین طراحی و آموزش مدل در نظر گرفته شده‌اند. در نتیجه، هنگامی که با پیکربندی‌های ناشناخته یا متفاوت مواجه می‌شوند، افت عملکرد قابل توجهی را تجربه می‌کنند.

### عامل کنترل چراغ راهنمایی یکپارچه (UniTSA): راهکار پیشنهادی

با توجه به این انگیزه، در این پژوهش، چارچوبی مبتنی بر یادگیری تقویتی با استفاده از یک عامل کنترل ترافیک یکپارچه (UniTSA) ارائه می‌دهیم. چارچوب پیشنهادی، امکان آموزش یک عامل عمومی با بهره‌گیری از داده‌های ترکیبی را برای کنترل چراغ‌های راهنمایی فراهم می‌سازد.

برای مدیریت تقاطع‌هایی با ساختارهای متنوع، از یک «ماتریس تقاطع» مبتنی بر داده استفاده می‌کنیم تا هندسه هر تقاطع را به‌طور یکپارچه توصیف کند. با استفاده از این ماتریس، می‌توان تقاطع‌های دارای ساختارهای مختلف را به ماتریس‌هایی با ابعاد ثابت نگاشت داد.

علاوه بر این، در طراحی فضای عمل (Action Space) این چارچوب، از رویکرد «نگه‌داشتن یا تغییر فاز فعلی» استفاده شده است تا پایداری مدل در سراسر تقاطع‌های مختلف حفظ شود.

برای مقابله مؤثر با تقاطع‌های نادیده (unseen) در مرحله آزمایش، پنج راهکار افزایشی حالت ترافیک ابداع شده‌اند تا داده‌های عامل را در طول فرآیند آموزش، غنی‌سازی کنند. این راهکارهای افزایش داده، آموزش جامع‌تری فراهم می‌آورند که در نهایت به عملکرد بهتر در تقاطع‌هایی منجر می‌شود که پیکربندی آن‌ها در مجموعه آموزشی وجود نداشته است. همچنین، از روش سازگارسازی کم‌رتبه [31] (LoRA) برای انجام بازآموزی سریع‌تر مدل در تقاطع‌های جدید استفاده می‌شود.

در نهایت، آزمایش‌های گسترده‌ای با استفاده از شبیه‌ساز SUMO بر روی تقاطع‌هایی با تعداد خطوط، ورودی‌ها و فازهای مختلف انجام گرفت. نتایج تجربی نشان می‌دهد که مدل UniTSA پیشنهادی، عملکرد بسیار مطلوبی حتی در تقاطع‌های ناشناخته ارائه می‌دهد.

### مشارکت‌های این پژوهش را می‌توان به شرح زیر خلاصه کرد:

۱. ارائه چارچوبی تطبیقی برای کنترل چراغ راهنمایی با نام UniTSA در محیط‌های V2X که بر پایه یک مدل RL عمومی استوار است و از طراحی فضای عمل نوینی برای مدیریت ساختارهای متنوع تقاطع برخوردار است. همچنین، یک مکانیزم تنظیم دقیق (Fine-tuning) برای ارتقای عملکرد در تقاطع‌های کلیدی طراحی شده است.
۲. معرفی راهکارهای افزایشی حالت ترافیک برای چارچوب کنترل چراغ راهنمایی، به منظور بهبود درک عامل از تقاطع‌های متنوع و ارتقای عملکرد آن در هر دو مرحله آموزش و آزمایش.
۳. آزمایش‌های گسترده بر روی ۱۲ تقاطع با ساختارهای گوناگون نشان می‌دهد که مدل UniTSA

پیشنهادی، عملکرد قابل توجهی بهتر از مدل‌های عمومی مرسوم ارائه می‌دهد. علاوه بر این، برای یک تقاطع جدید، UniTSA تنها با بهره‌گیری از یک دوره بازآموزی کوتاه با استفاده از LoRA، می‌تواند به عملکرد بهینه نزدیک شود و نیاز به آموزش از صفر را مرتفع سازد. این ویژگی، گامی مهم در جهت قابلیت انتقال و مقیاس‌پذیری راهکارهای کنترل هوشمند ترافیک در شهرهای واقعی است.

مدل پیش‌آموزش‌دیده UniTSA می‌تواند با فرآیندی سریع از تنظیم دقیق (Fine-tuning) بر روی داده‌های تقاطع جدید، خود را تطبیق دهد و در زمان آموزشی به مراتب کوتاه‌تر نسبت به آموزش یک مدل جدید از ابتدا، به عملکردی قابل قبول یا حتی بهتر دست یابد.

ساختار کلی مقاله به صورت زیر است:

بخش ۲ مروری بر کارهای مرتبط با کنترل چراغ راهنمایی (TSC) ارائه می‌دهد. بخش ۳ اصطلاحات و مفاهیم مرتبط با راه‌ها و چراغ‌های ترافیکی را معرفی می‌کند. سپس بخش ۴ چارچوب پیشنهادی UniTSA و پنج راهکار افزایش حال ترافیکی را تشریح می‌نماید و بخش ۵ تنظیمات تجربی و نتایج ارزیابی UniTSA را ارائه می‌دهد. در نهایت، بخش ۶ مقاله را جمع‌بندی می‌کند.

---

<sup>1</sup> Universal Traffic State Augmentation

<sup>2</sup> Simulation of Urban MOBility

## ۲. کارهای مرتبط

پژوهش‌های گسترده‌ای در حوزه حمل‌ونقل برای حل مسئله کنترل چراغ راهنمایی (TSC) انجام شده است.

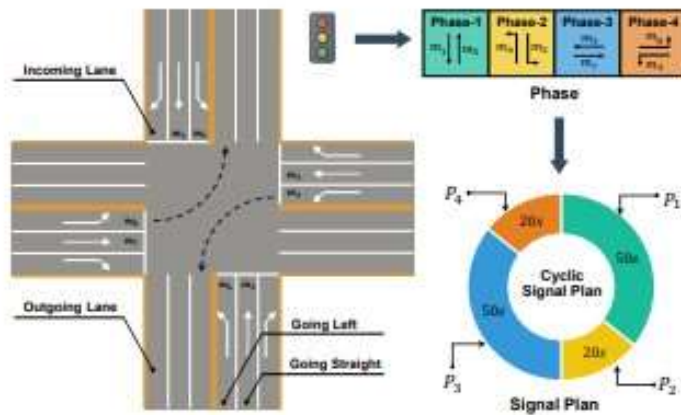
رویکرد سنتی کنترل زمان ثابت (Fixed-time)، یکی از قدیمی‌ترین و گسترده‌ترین روش‌های TSC است [۳۳]. این روش‌ها مبتنی بر زمان‌بندی‌های از پیش تعیین‌شده‌ای هستند که بر اساس الگوهای تاریخی ترافیک یا دستورالعمل‌های مهندسی تعیین می‌شوند. تکنیک‌های بهینه‌سازی مختلفی برای تعیین برنامه‌های زمان ثابت بهینه برای پیکربندی‌های تقاطع ابداع شده‌اند که روش وبستر (Webster) یکی از موفق‌ترین آن‌ها برای تقاطع‌های تکی است. این روش، طول چرخه و توزیع فاز را متناسب با حجم ترافیک در یک بازه زمانی (معمولاً ۱۵ یا ۳۰ دقیقه) محاسبه می‌کند. با این حال، چنین روش‌های زمان‌ثابتی، به دلیل عدم انعطاف در تطبیق با تغییرات پویای شرایط ترافیکی، اغلب عملکردی زیر حد بهینه دارند.

برای مقابله با این مشکل، روش‌های کنترلی سازگار (Adaptive) مانند سیستم هماهنگ‌شده و پویا سیدنی [SCATS] ۳۴، کنترل فشار حداکثر (Max-pressure) ۳۵ [و] SOTL ۴ [طراحی شده‌اند تا زمان‌بندی‌ها را بر اساس تقاضای ترافیکی بلادرنگ تنظیم کنند. با وجود مزایای متعدد، این روش‌های سازگار نیازمند تنظیمات تجربی برای هر تقاطع هستند و در سناریوهای پیچیده، عملکردشان کاهش می‌یابد.

در سال‌های اخیر، روش‌های مبتنی بر یادگیری تقویتی (RL) توجه زیادی را به خود جلب کرده‌اند؛ دلیل این امر، توانایی قابل توجه آن‌ها در سازگاری با شرایط ترافیکی بلادرنگ و قابلیت یادگیری سیاست‌های کنترلی بهینه در سناریوهای پیچیده است [۳۶]. به طور کلی، این روش‌ها را می‌توان به سه دسته تقسیم کرد: روش‌های مبتنی بر ارزش (Value-based) [۸، ۹، ۱۰، ۱۱، ۱۲، ۱۳، ۱۹، ۲۳، ۲۵]، روش‌های مبتنی بر سیاست (Policy-based) [۱۵، ۱۶، ۱۷، ۱۸، ۱۹] و روش‌های بازیگر-منتقد (Actor-Critic) [۲۰، ۲۱، ۲۲، ۲۸]. علیرغم عملکرد خوب، اکثر روش‌های RL موجود بر آموزش مدل‌هایی برای پیکربندی‌های تقاطع خاص متمرکزند.

تلاش‌هایی نیز برای انتقال (Transfer) مدل‌های آموزش‌دیده TSC انجام شده است. برای مثال :

در پژوهشی دیگر [۱۱]، یک مدل نسبتاً عمومی‌تر با بهره‌گیری از راهبرد فرا-یادگیری [37] (Meta-Learning) آموزش داده می‌شود. با این حال، این مدل همچنان نیازمند بازآموزی پارامترها برای هر تقاطع جدید است. برای غلبه بر این نقیصه، برخی پژوهش‌ها [۱۰، ۱۲، ۱۸] به ایجاد مدل‌های عمومی از طریق اشتراک‌گذاری پارامترها میان تقاطع‌های مختلف پرداختند؛ اما این روش‌ها معمولاً ساختار فزاینده اصلی چراغ راهنمایی را حفظ نمی‌کنند. در مقابل، روش پیشنهادی ما در این مقاله می‌تواند ساختار چراغ سیگنال موجود را حفظ کند و بطور همزمان، با تنظیم دقیق (Fine-tuning) بر روی تقاطع‌های حیاتی، عملکرد را به صورت چشمگیری بهبود بخشد.



شکل ۱) نقشه یک چهار راه استاندارد را نمایش می دهد.

### 3- مفاهیم اولیه

در این بخش، به توصیف و بررسی مدل‌هایی می‌پردازیم که در این پژوهش برای شبیه‌سازی و تحلیل تقاطع‌های ترافیکی مورد استفاده قرار گرفته‌اند. این مدل‌ها بر پایه یک تقاطع استاندارد چهارراهی (مانند آنچه در شکل ۱ نشان داده شده است) بنا شده‌اند. این رویکرد اجازه می‌دهد تا نتایج به دست آمده، قابل تعمیم به تقاطع‌های مشابه با ساختارهای متنوع باشند، از جمله تقاطع‌هایی با تعداد خطوط متفاوت، الگوهای ترافیکی پیچیده‌تر یا حتی تقاطع‌های چندسطحی. هدف از این مدل‌سازی، بهینه‌سازی جریان ترافیک، کاهش زمان انتظار وسایل نقلیه و افزایش ایمنی است. در ادامه، مفاهیم کلیدی این مدل را به صورت گام‌به‌گام توضیح می‌دهیم و با مثال‌های عملی گسترش می‌دهیم تا درک بهتری از کاربرد آن‌ها در دنیای واقعی فراهم شود.

#### الف) خط عبور (Lane)

خط عبور، بخشی از مسیر راه است که به صورت مشخص و تاریک‌شده (مانند خط‌کشی‌های جاده) طراحی شده تا مسیر ایمن و هدایت‌شده‌ای برای حرکت وسایل نقلیه در یک جهت خاص فراهم کند. این خطوط نقش اساسی در سازماندهی جریان ترافیک دارند و از برخوردهای احتمالی جلوگیری می‌کنند. در یک تقاطع معمولی چهارراهی، دو نوع اصلی خط عبور وجود دارد:

- **خط ورودی (Inbound Lane):** این خطوط جایی هستند که وسایل نقلیه از مسیرهای ورودی وارد تقاطع می‌شوند. برای مثال، در یک چهارراه شهری، وسایل نقلیه از خیابان‌های اطراف وارد این خطوط شده و منتظر سیگنال برای ادامه مسیر می‌مانند.

- **خط خروجی (Outbound Lane):** این خطوط برای خروج وسایل نقلیه از تقاطع به سمت مسیرهای خروجی طراحی شده‌اند. آن‌ها جریان ترافیک را به سمت مقصد هدایت می‌کنند.

در عمل، تعداد خطوط ورودی و خروجی می‌تواند بسته به حجم ترافیک تغییر کند. برای مثال، در تقاطع‌های پرترافیک مانند میدان‌های اصلی شهرها، ممکن است چندین خط ورودی برای جهت‌های مختلف (مانند مستقیم، چپ‌گرد یا راست‌گرد) وجود داشته باشد تا ظرفیت عبور افزایش یابد. این مدل‌سازی کمک می‌کند تا مهندسان ترافیک بتوانند با استفاده از نرم‌افزارهایی مانند VISSIM یا SUMO، سناریوهای مختلف را شبیه‌سازی کنند و بهترین طراحی را انتخاب نمایند.



## ب) حرکت ترافیکی (Traffic Movement)

حرکت ترافیکی به عنوان جریان وسایل نقلیه بین یک خط ورودی و یک خط خروجی تعریف می‌شود. این مفهوم پایه‌ای برای تحلیل رفتار ترافیک در تقاطع‌هاست و شامل انواع مختلفی از مانورها می‌شود. در یک تقاطع چهارراهی استاندارد (مانند آنچه در شکل ۱ نشان داده شده)، معمولاً ۱۲ حرکت ترافیکی اصلی وجود دارد که شامل ترکیبی از گردش‌های راست، چپ و حرکات مستقیم در چهار جهت جغرافیایی (شمال، جنوب، شرق و غرب) است. برای مثال:

- **حرکات مستقیم:** مانند حرکت از شمال به جنوب یا از شرق به غرب، که بیشترین حجم ترافیک را در ساعات پیک تشکیل می‌دهند.
  - **گردش‌های چپ:** که اغلب پیچیده‌تر هستند و ممکن است با حرکات مقابل تداخل داشته باشند، مانند گردش چپ از شرق به شمال.
  - **گردش‌های راست:** که معمولاً ساده‌تر و با تداخل کمتری همراه هستند.
- این ۱۲ حرکت می‌تواند بر اساس داده‌های واقعی (مانند شمارش وسایل نقلیه در ساعات مختلف روز) مدل‌سازی شود. گسترش این مفهوم شامل در نظر گرفتن عوامل خارجی مانند حضور عابران پیاده، دوچرخه‌سواران یا وسایل نقلیه سنگین است که می‌تواند زمان‌بندی سیگنال‌ها را تحت تأثیر قرار دهد. در مطالعات واقعی، ابزارهایی مانند دوربین‌های هوشمند برای ثبت این حرکات استفاده می‌شود تا مدل دقیق‌تری ساخته شود.

## ج) سیگنال حرکت (Movement Signal)

- سیگنال حرکت، کنترل‌کننده‌ای است که بر روی هر حرکت ترافیکی اعمال می‌شود و وضعیت مجاز یا ممنوع بودن آن را تعیین می‌کند. این سیگنال‌ها معمولاً با چراغ‌های راهنمایی نشان داده می‌شوند:
- **چراغ سبز:** نشان‌دهنده مجاز بودن حرکت است و وسایل نقلیه می‌توانند ادامه دهند.
  - **چراغ قرمز:** نشان‌دهنده ممنوع بودن حرکت است و وسایل باید توقف کنند.
- در بسیاری از کشورها، گردش راست (Right Turn on Red) مستقل از چراغ قرمز مجاز است، به شرطی که ایمنی رعایت شود. این امر باعث می‌شود که از ۱۲ حرکت ترافیکی ممکن در یک چهارراه، تنها ۸ سیگنال حرکت اصلی نیاز به کنترل مستقیم داشته باشند.

این ۸ حرکت به صورت زیر نام گذاری و تعریف می شوند (با استفاده از جهت های جغرافیایی استاندارد):

- **N (North Through):** حرکت مستقیم از شمال به جنوب.
- **NL (North Left):** گردش چپ از شمال به غرب.
- **E (East Through):** حرکت مستقیم از شرق به غرب.
- **EL (East Left):** گردش چپ از شرق به شمال.
- **W (West Through):** حرکت مستقیم از غرب به شرق.
- **WL (West Left):** گردش چپ از غرب به جنوب.
- **S (South Through):** حرکت مستقیم از جنوب به شمال.
- **SL (South Left):** گردش چپ از جنوب به شرق.

برای مثال، سیگنال M1 ممکن است نشان دهنده این باشد که وسایل نقلیه می توانند از شمال مستقیم به جنوب حرکت کنند. این نام گذاری کمک می کند تا در الگوریتم های بهینه سازی، تداخل بین حرکات (مانند گردش چپ با حرکت مستقیم مقابل) شناسایی و مدیریت شود. گسترش این بخش شامل ادغام فناوری های هوشمند مانند سیگنال های تطبیقی است که بر اساس حجم واقعی ترافیک (از طریق سنسورها) زمان بندی را تغییر می دهند، که می تواند زمان انتظار را تا ۳۰ درصد کاهش دهد.

#### د) فاز (Phase)

فاز، ترکیبی از چندین سیگنال حرکت است که به طور همزمان فعال می شوند. هر فاز، مجموعه ای خاص از حرکات ترافیکی را مجاز می کند در حالی که دیگران را محدود یا ممنوع می سازد. این رویکرد اجازه می دهد تا تداخل ها به حداقل برسد و ایمنی افزایش یابد. در قسمت بالا-راست شکل ۱، چهار فاز استاندارد برای یک چهارراه نشان داده شده است.

برای مثال:

- **فاز ۱:** شامل حرکات M1 و M5 است و اجازه می دهد وسایل نقلیه در جهت شمال-جنوب (حرکات مستقیم) ادامه دهند، در حالی که سایر حرکات متوقف هستند. این فاز مناسب برای ساعاتی است که ترافیک شمال-جنوب غالب است.

- فازهای بعدی ممکن است حرکات شرق-غرب یا گردش های چپ را پوشش دهند.

گسترش این مفهوم شامل طراحی فازهای پیچیده تر برای تقاطع های نامتقارن است، جایی که فازهای اضافی برای عابران یا وسایل عمومی (مانند اتوبوس) اضافه می شود. در مدل سازی، از الگوریتم هایی مانند Webster's Method برای محاسبه زمان بهینه هر فاز استفاده می شود تا تعادل بین ظرفیت و ایمنی برقرار شود.

#### ه) طرح سیگنال (Signal Plan)

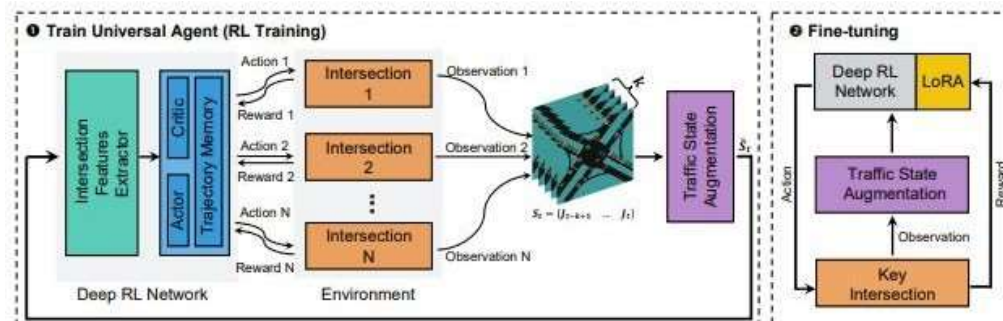
طرح سیگنال، دنباله ای چرخه ای از فازها همراه با مدت زمان اختصاص یافته به هر فاز است که برای کنترل کلی چراغ های تقاطع به کار می رود. این طرح به صورت مجموعه ای از جفت ها نمایش داده می شود،  $\{(P_1, T_1), \dots, (P_n, T_n)\}$ ، جایی که  $P_i$  فاز  $i$  ام و  $T_i$  مدت زمان آن (به ثانیه) است. معمولاً این دنباله چرخه ای است،

یعنی پس از آخرین فاز، به فاز اول بازمی‌گردد.

در قسمت پایین-راست شکل ۱، یک طرح چرخه‌ای سیگنال نشان داده شده با زمان‌بندی زیر:

- $T_1 = 50$  ثانیه (برای فاز ۱، تمرکز روی حرکات مستقیم شمال-جنوب).
- $T_2 = 20$  ثانیه (برای فاز ۲، مثلاً حرکات شرق-غرب).
- $T_3 = 20$  ثانیه.
- $T_4 = 20$  ثانیه (فرض بر چهار فاز استاندارد).

این زمان‌بندی می‌تواند بر اساس داده‌های واقعی تنظیم شود؛ برای مثال، در ساعات پیک، زمان فازهای پرتراфик افزایش یابد. گسترش این بخش شامل ادغام هوش مصنوعی است، جایی که طرح سیگنال به صورت پویا تغییر می‌کند (مانند سیستم‌های SCATS در استرالیا) تا با شرایط واقعی مانند ترافیک سنگین یا حوادث تطبیق یابد. در نهایت، این مدل‌ها نه تنها به کاهش congestion کمک می‌کنند، بلکه مصرف سوخت و آلودگی را نیز پایین می‌آورند.



شکل ۲) نمایش کلی ساختار UniTSA

### ۳- ویژگی‌های سیگنال ترافیکی

این بخش شامل سه پارامتر باینری است:

- $I_{i,t}^{cg}$ : آیا حرکت فعلاً چراغ سبز دارد یا نه
- $I_{i,t}^{ng}$ : آیا حرکت در فاز بعد چراغ سبز خواهد داشت یا نه
- $I_{i,t}^{mg}$ : آیا مدت زمان حداقل چراغ سبز فعلی به پایان رسیده یا نه

این ۸ ویژگی به سادگی قابل استخراج هستند، که مدل را برای استفاده عملی قابل اجرا می‌سازد.

بردار  $m_i^t$  به صورت زیر تعریف می‌شود:

$$m_i^t = [F_i, t, O_i, tmax, O_i, tmean, l_{is}, L_i, l_i, t_{cg}, l_i, t_{ng}, l_i, t_{mg}]T \quad (2)$$

### ۴- روش شناسی

#### ۴-۱ چارچوب پیشنهادی

همان‌طور که در شکل ۲ نشان داده شده است

چارچوب پیشنهادی **UniTSA (Unified Traffic Signal Agent)** از دو ماژول اصلی تشکیل شده است:

**(الف) ماژول رمزگذاری حالت محیط (State Encoding Module)**

این ماژول، یک مدل یادگیری تقویتی **Reinforcement Learning** یا **RL** آگاه از محیط را با استفاده از تقاطع‌های متنوع، رمزگذاری (**encode**) می‌کند. این ماژول از طراحی نوینی الهام گرفته شده که قادر است تقاطع‌هایی با توپولوژی‌ها و راه‌های سیگنالی متفاوت را با یک ساختار ثابت مدل کند. این قابلیت، از طریق بهره‌گیری از ویژگی‌های مشترک حرکات ترافیکی و اعمال اصل فاز بعدی یا هیچ (**Next-Phase-or-None**) و همچنین پنج نوروں اضافی برای افزایش حل‌پذیری مسائل ترافیکی پیچیده محقق شده است.

مدل نهایی که برای تقاطع‌های کوچک در این پژوهش به کار گرفته شده، در واقع مدل اولیه (**Initial Model**) است که برای تقاطع‌های خاص با پیچیدگی بالاتر، به صورت تحلیلی ارتقا یافته است. جزئیات دقیق هر مرحله از این ماژول در بخش‌های بعدی توضیح داده خواهد شد.

**(ب) طراحی مدل به صورت ماتریس تقاطع (Intersection Matrix Representation)**

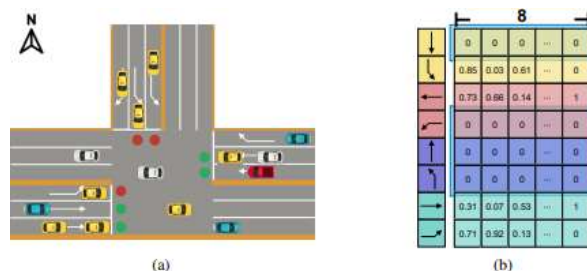
پیش‌نیاز: در هر مسیر تقاطع‌ها ممکن است که دارای خطوط متفاوتی داشته باشند که در نتیجه آن، ویژگی‌های حالت

**(State Features)** آن‌ها در صحت، ابعاد و فضای حالت با یکدیگر متفاوت خواهد بود. همان‌طور که در بخش ۳ اشاره شد، در یک چهارراه استاندارد تنها هشت سیگنال حرکت معتبر وجود دارد، بدون توجه به تعداد خطوط ورودی. با توجه به این مشاهده کلیدی، ما پیشنهاد می‌کنیم که حالت تقاطع در زمان  $t$  به صورت یک **ماتریس تقاطع (Intersection Matrix)** نمایش داده شود. این ماتریس به شکل زیر تعریف می‌شود:

$$\mathbf{S}_t = \begin{bmatrix} \mathbf{S}_N & \mathbf{S}_{NL} & \mathbf{S}_{NE} \\ \mathbf{S}_W & \mathbf{S}_C & \mathbf{S}_E \\ \mathbf{S}_{SW} & \mathbf{S}_S & \mathbf{S}_{SE} \end{bmatrix} \in \mathbb{R}^{3 \times 3 \times d}$$

که در آن:

- هر خانه از ماتریس  $3 \times 3$  معادل یکی از جهت‌های جغرافیایی یا ترکیبی از آن‌هاست:
    - $\mathbf{S}_N$ : ویژگی‌های مربوط به جهت شمال (حرکت مستقیم از شمال).
    - $\mathbf{S}_{NL}$ : ویژگی‌های مربوط به گردش چپ از شمال.
    - $\mathbf{S}_{NE}$ : ویژگی‌های مربوط به گردش راست از شمال (یا سایر حرکات فرعی).
    - $\mathbf{S}_W, \mathbf{S}_S, \mathbf{S}_E$ : به ترتیب جهت‌های غرب، شرق و جنوب.
    - $\mathbf{S}_{SW}, \mathbf{S}_{SE}$ : گردش‌های فرعی جنوب‌غرب و جنوب‌شرق.
    - $\mathbf{S}_C$ : خانه مرکزی که می‌تواند برای ویژگی‌های مشترک تقاطع (مانند زمان سپری‌شده از چرخه، وضعیت عابران پیاده یا اطلاعات کلی) استفاده شود.
  - بعد سوم ( $d$ ) نشان‌دهنده بردار ویژگی‌های هر جهت است مانند تعداد وسایل نقلیه در صف، سرعت متوسط، وضعیت چراغ فعلی و غیره. این نمایش ماتریسی مزایای زیر را دارد:
  - **ثابت بودن ابعاد**: صرف‌نظر از تعداد خطوط واقعی در هر جهت، همیشه یک ماتریس  $3 \times 3$  داریم که باعث یکنواختی ورودی شبکه عصبی می‌شود.
  - **حفظ روابط فضایی**: موقعیت نسبی جهت‌ها (مانند شمال در ردیف بالا، شرق در ستون راست) حفظ می‌شود و شبکه می‌تواند روابط جغرافیایی را به طور طبیعی یاد بگیرد.
  - **قابلیت تعمیم‌پذیری**: برای تقاطع‌های غیراستاندارد (مانند سه‌راه یا پنج‌راه)، می‌توان خانه‌های غیرمرتبط را با صفر پر کرد یا از ماسک توجه (**attention mask**) در مدل‌های ترانسفورمر استفاده نمود.
- این ساختار ماتریسی، پایه اصلی ماژول رموزگذاری است و اجازه می‌دهد تا عامل **RL (Agent)** بدون نیاز به طراحی جداگانه برای هر توپولوژی، روی انواع تقاطع‌ها آموزش ببیند و عملکرد بالایی داشته باشد. در بخش‌های بعدی، نحوه استخراج ویژگی‌های هر خانه و چگونگی ادغام این ماتریس در شبکه عصبی عامل توضیح داده خواهد شد.



شکل ۳) یک تقاطع سه راهی به همراه ماتریس متقاطع آن

### ۳- افزودن اطلاعات زمانی

مشاهده یک ماتریس  $t$  در یک زمان واحد برای درک کامل دینامیک ترافیک کافی نیست. برای رفع این مشکل، ما آخرین  $K$  مشاهده را وارد وضعیت می‌کنیم تا عامل بتواند الگوها و روندهای ترافیکی را بهتر تشخیص دهد. در نتیجه، وضعیت عامل در زمان  $t$  به صورت زیر تعریف می‌شود:

$$S_t \in \mathbb{R}^{K \times 8 \times 8} \quad (3)$$

شایان ذکر است که هرگاه تعداد حرکت‌ها در یک تقاطع کمتر از ۸ باشد، برای یکنواخت‌سازی ابعاد ورودی، از روش **Zero Padding** استفاده می‌شود. به عنوان نمونه، شکل ۳-الف یک تقاطع سه‌راهی متداول را نمایش می‌دهد که در آن تنها حرکت‌های شرق (E)، گردش چپ به شرق (EL)، غرب (W) و گردش چپ به جنوب (SL) فعال هستند. همان‌طور که در شکل ۳-ب مشاهده می‌شود، سطرهای مربوط به حرکت‌های غیرفعال با مقدار صفر پر می‌شوند تا ابعاد ماتریس حاصل با ماتریس مربوط به یک تقاطع چهارراهی یکسان باقی بماند.

طراحی فضای اعمال و انتخاب اقدامات باید به گونه ای انجام شود که ایمنی ترافیکی نیز در نظر گرفته شود. اگرچه استفاده از اقدام «انتخاب فاز بعدی» (Choose Next Phase) می تواند بازدهی تقاطع را به طور قابل توجهی افزایش دهد، اما این رویکرد ترتیب منطقی و از پیش تعریف شده چراغ های راهنمایی را برهم می زند و در نتیجه ممکن است ایمنی رانندگان را تحت تأثیر قرار دهد.

در مقابل، در این پژوهش از اقدام «حفظ یا تغییر فاز» استفاده شده است. این رویکرد ضمن پایبندی به توالی استاندارد چراغ ها، فازها را به صورت ترتیبی اجرا می کند؛ به عنوان مثال: فاز ۱، فاز ۲، فاز ۳، فاز ۴ و سپس مجدداً فاز ۱ و فاز ۲ و به همین ترتیب. عامل تصمیم گیرنده (Agent) بر اساس وضعیت  $t$  تصمیم می گیرد که:

- فاز فعلی را حفظ کند، یا
- به فاز بعدی منتقل شود.

با توجه به وجود شاخص فعال بودن فاز فعلی  $I_{cg}$  و شاخص فعال بودن فاز بعدی  $I_{ng}$  در ماتریس تقاطع، در ادامه نشان داده می شود که این طراحی از فضای اعمال، مقیاس پذیری بسیار مناسبی برای تقاطع هایی با تعداد متفاوتی از مسیرها و فازهای سیگنال ارائه می دهد.

---

### پاداش (Reward)

میانگین طول صف در هر حرکت  $q_i$  به عنوان تابع پاداش انتخاب شده است. معیارهایی مانند زمان انتظار، زمان سفر یا تأخیر مورد استفاده قرار نگرفته اند، زیرا محاسبه آنها نیازمند تجهیزات و سامانه های پیشرفته تشخیص ترافیک است که در همه تقاطع ها در دسترس نیستند. بر این اساس، تابع پاداش پیشنهادی به صورت زیر تعریف می شود:

$$r_t = - \left( \sum_{i=1}^8 q_i \right) - \frac{\mu}{\sigma + \varepsilon} \quad (4)$$

که در آن  $\varepsilon$  عددی کوچک است تا از تقسیم بر صفر جلوگیری شود. علاوه بر این،  $\mu$  و  $\sigma$  به ترتیب میانگین و انحراف معیار اولین  $R$  مقدار پاداش هستند. این مقادیر به صورت زیر محاسبه می شوند:

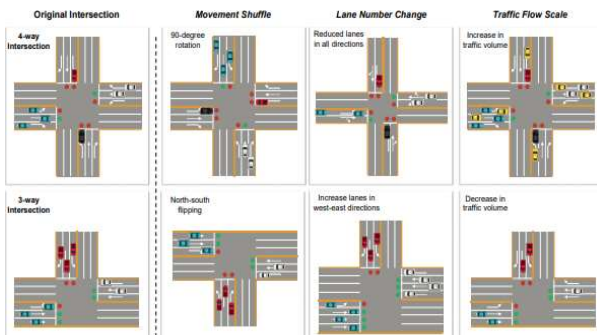
$$\mu = \frac{1}{R-1} \sum_{j=1}^{R-1} r_j \quad (5)$$

$$\sigma = \sqrt{\frac{1}{R-1} \sum_{j=1}^{R-1} (r_j - \mu)^2} \quad (6)$$

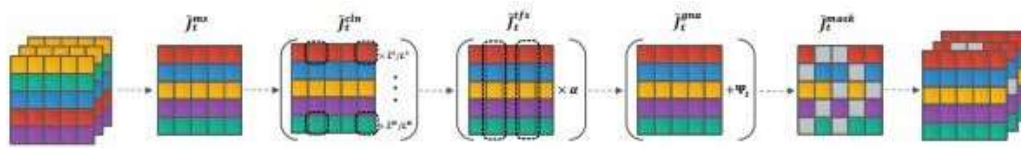
پاداش نرمال سازی می شود تا روند آموزش سریع تر و کارآمدتر گردد.

## افزایش داده ترافیکی

در پژوهش های پیشین نشان داده شده است که به کارگیری روش های افزایش داده می تواند نقش مؤثری در ارتقای اثربخشی و بهبود فرآیند آموزش مدل های یادگیری تقویتی ایفا کند. [38-40] با در اختیار داشتن مجموعه داده ای متنوع تر و غنی تر که از طریق تکنیک های افزایش داده تولید شده است، عامل های یادگیری تقویتی قادر خواهند بود الگوهای رفتاری پایدارتری را فراگرفته و عملکرد قابل اعتمادتر و بهتری را در شرایط مختلف و حتی سناریوهای دیده نشده از خود نشان دهند. شکل ۴ نمایی شماتیک از فرآیند افزایش وضعیت ترافیکی را نشان می دهد که برای تقاطع های چهارراهی و سه راهی به کار گرفته شده است.







شکل ۵ مراحل دقیق اجرای فرآیند افزایش وضعیت ترافیکی را نمایش می‌دهد.

افزایش داده با هدف گسترش دامنه سناریوهای دیده‌نشده و ارتقای توان تعمیم‌پذیری مدل انجام می‌شود. متداول‌ترین روش‌های افزایش داده که در ادبیات پژوهشی گزارش شده‌اند شامل افزودن نویز گاوسی و ماسک‌گذاری هستند. در این پژوهش، سه روش نوآورانه دیگر برای افزایش وضعیت‌های ترافیکی، به‌طور خاص متناسب با مسائل کنترل سیگنال ترافیکی مبتنی بر یادگیری تقویتی (RL-TSC) پیشنهاد می‌شود؛ این روش‌ها شامل جابجایی حرکت‌ها، تغییر تعداد حرکت‌ها و مقیاس‌بندی جریان ترافیک هستند که در شکل ۴ به‌صورت شماتیک نمایش داده شده‌اند.

شکل ۵ فرآیند گام‌به‌گام اعمال پنج روش مختلف افزایش داده بر وضعیت ترافیکی  $K_t$  را نشان می‌دهد که در نهایت منجر به تولید وضعیت‌های افزوده‌شده  $K_t^*$  می‌شود. در طول فرآیند آموزش، یک مینی‌بتچ از داده‌ها به‌صورت تصادفی از حافظه بازپخش یا از مسیرهای افزوده‌شده نمونه‌برداری می‌شود. هرچند عملیات افزایش داده به‌صورت تصادفی بر روی مینی‌بتچ‌ها اعمال می‌گردد، اما در میان فریم‌های متوالی، ثبات زمانی حفظ می‌شود.

همچنین شایان ذکر است که این پنج روش افزایش وضعیت ترافیکی می‌توانند به‌صورت مستقیم بر روی ماتریس تقاطع اعمال شوند. این ویژگی باعث می‌شود عامل یادگیرنده بتواند سناریوها و ساختارهای متنوع‌تری از تقاطع‌ها را تجربه کرده و توانایی یادگیری و سازگاری بهتری با شرایط مختلف به‌دست آورد.

### جا به جایی حرکت‌ها

در این روش، سطرهای ماتریس تقاطع جابجا می‌شوند تا چرخش‌ها یا تقارن‌های مختلف یک تقاطع شبیه‌سازی شوند. به‌عبارت دیگر، جابجایی حرکت‌ها را می‌توان نوعی «چرخش مؤثر» تقاطع در نظر گرفت، همان‌گونه که در شکل ۴ نشان داده شده است. فرض اصلی این روش آن است که اقدام انتخاب‌شده توسط عامل پس از چرخش تقاطع تغییر نکند.

به‌صورت ریاضی، این خاصیت به شکل زیر مدل‌سازی می‌شود:

$$J_t^{ms} = P \cdot J_t, \quad (7)$$

که در آن،  $J_t$  ماتریس اصلی و  $\tilde{J}_t^{ms}$  ماتریس افزوده شده هستند. ماتریس  $P$  یک ماتریس جایگشتی است که سطرهای  $J_t$  را جابجا می کند.

با اعمال «جابجایی حرکتها» بر تمام ماتریسهای موجود در  $S_t$ ، حالت جدید به صورت زیر تعریف می شود:

$$\tilde{S}_t^{ms} = [\tilde{J}_{t-K+1}^{ms}, \tilde{J}_{t-K+2}^{ms}, \dots, \tilde{J}_t^{ms}]. \quad (8)$$

تغییر تعداد خطوط

همانطور قرار دادن عامل در معرض ترکیبهای متنوع تری از ساختارهای جاده ای، تعداد خطوط  $L_i$  در هر بردار حرکت  $\tilde{m}_{ms}$  به صورت تصادفی تغییر داده می شود. این روش موجب می شود عامل در طول فرآیند آموزش با پیکربندیهای مختلف خطوط آشنا شده و توانایی سازگاری بهتری با انواع گوناگون تقاطعها به دست آورد. افزون بر این، ویژگیهای ترافیکی نظیر جریان و اشغال نیز به صورت متناسب با حفظ نسبتها، مقیاس بندی می شوند.

نمونه ای از این روش در شکل ۴ ارائه شده است؛ به گونه ای که در یک تقاطع چهارراهی، تعداد خطوط هر جهت به میزان ۲ کاهش یافته و در یک تقاطع سه راهی، تعداد خطوط ورودی شرق و غرب به میزان ۴ افزایش داده شده است. در این فرآیند، نسبت تعداد وسایل نقلیه در هر جهت ثابت نگه داشته می شود؛ بنابراین انتظار می رود رفتار آموزشی عامل پیش و پس از این تغییرات یکسان باقی بماند.

این ویژگی به صورت زیر مدل سازی می شود:

$$\tilde{m}_{cln} = f(\tilde{m}_{ms}, \tilde{L}_i), i = 1, 2, \dots, 8$$

(9)

که در آن:

- $\tilde{L}_i$  متغیرهای تصادفی با توزیع یکنواخت هستند که افزایش یا کاهش تعداد خطوط را نشان می دهند؛
- تابع  $f$  وظیفه تنظیم ویژگیهای ترافیکی را بر عهده دارد.

تابع  $f$  به صورت زیر تعریف می شود:

$$f(\tilde{m}_{ms}, \tilde{L}_i) = \begin{cases} \frac{[\tilde{m}_{ms}]_k}{L_i} \cdot \tilde{L}_i, & k = 1, 2, 3, 5 \\ [\tilde{m}_{ms}]_k, & \text{در غیر این صورت} \end{cases}$$

(10)

که در آن  $k$  مؤلفه  $k$ -ام بردار را نشان می‌دهد. پس از اعمال این روش، وضعیت جدید به صورت زیر تعریف می‌شود:

$$\tilde{S}_{cln} = [\tilde{J}_{cln}^t, \dots, \tilde{J}_{cln}^{t-K+1}]$$

(11)

و ماتریس تقاطع متناظر برابر است با:

$$\tilde{J}_{cln}^t = [\tilde{m}_{cln}^1, \tilde{m}_{cln}^2, \dots, \tilde{m}_{cln}^8]^T$$

(12)

مقیاس‌بندی جریان ترافیک

برای کاهش وابستگی عامل به مقادیر مطلق جریان ترافیک و افزایش تمرکز آن بر روابط نسبی بین حرکات، از یک ضریب تصادفی مقیاس‌بندی جریان استفاده می‌شود. در این روش، مقدار جریان و اشغال هر حرکت در ماتریس تقاطع با یک ضریب تصادفی یکنواخت  $\alpha$  ضرب می‌شود.

ضریب  $\alpha$  برای تمامی حرکات‌های یک وضعیت ثابت در نظر گرفته می‌شود تا نسبت میان حرکات حفظ گردد. این روش به عامل کمک می‌کند الگوهای نسبی ترافیک را بهتر یاد بگیرد و وابستگی کمتری به مقادیر مطلق جریان داشته باشد. لازم به ذکر است که در این روش، تعداد خطوط تغییری نمی‌کند.

این فرآیند به صورت زیر مدل‌سازی می‌شود:

$$\tilde{m}_{tfs} = g(\tilde{m}_{cln}, \alpha)$$

(13)

که تابع  $g$  به شکل زیر تعریف می‌شود:

$$g(\tilde{m}_{cln}, \alpha) = \begin{cases} [\tilde{m}_{cln}]_k \cdot \alpha, & k = 1, 2, 3 \\ [\tilde{m}_{cln}]_{k'}, & \text{در غیر این صورت} \end{cases}$$

(14)

در نهایت، وضعیت حاصل از این مرحله با نماد  $\tilde{S}_{tfs}$  نمایش داده می‌شود.

### افزودن نویز گاوسی

به منظور افزایش تصادفی بودن داده های آموزشی، نویز گاوسی به صورت مستقیم به ماتریس تقاطع افزوده می شود. این نویز می تواند بر تمامی مؤلفه های ماتریس، از جمله ویژگی های ترافیکی، ویژگی های حرکتی و مشخصه های سیگنال تاثیر بگذارد. به کارگیری این روش سبب می شود عامل یادگیرنده در برابر شرایط نامطمئن و اختلالات پیش بینی نشده، پایداری و مقاومت بیشتری از خود نشان دهد. فرآیند افزودن نویز به صورت زیر مدل سازی می شود:

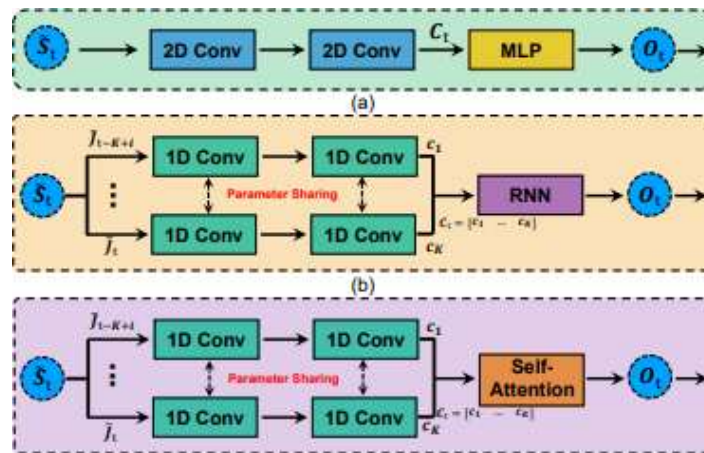
$$\tilde{J}_{gna}^t = \tilde{J}_{tfs}^t + \Psi_t$$

(15)

که در آن  $\Psi_t \sim \mathcal{N}(0, I)$  نویز گاوسی با میانگین صفر است. پس از اعمال این مرحله، وضعیت حاصل با نماد  $\tilde{J}_{gna}^t$  نمایش داده می شود.

### نقاب گذاری (Masking)

به منظور تشویق عامل به یادگیری پویایی های جریان ترافیک، بخشی از مقادیر موجود در ماتریس تقاطع در یک لحظه زمانی به صورت تصادفی صفر می شوند. این فرآیند نقاب گذاری، عامل را وادار می کند تا برای استنباط وضعیت ترافیکی، به اطلاعات گذشته و آینده و همچنین الگوهای زمانی ترافیک اتکا کند. استفاده از این روش، توانایی عامل در درک نوسانات و تغییرات زمانی ترافیک را به طور مؤثری افزایش می دهد.



شکل ۶: انواع ساختارهای استخراج ویژگی تقاطع

- (a) ساختار مبتنی بر CNN
- (b) ساختار مبتنی بر RNN
- (c) ساختار مبتنی بر ترانسفورمر

#### ۴-۴ استخراج ویژگی‌های تقاطع

در این بخش، سه ساختار شبکه عصبی برای استخراج ویژگی‌های تقاطع از وضعیت‌های ترافیکی افزایش یافته به کار گرفته می‌شوند:

۱. شبکه کانولوشنی چندبعدی (CNN)

۲. شبکه بازگشتی (RNN)

۳. شبکه مبتنی بر ترانسفورمر (Transformers)

#### ساختار مبتنی بر CNN

همان‌طور که در شکل ۶ (a) نشان داده شده است، در این ساختار از یک شبکه CNN چندلایه برای استخراج ویژگی‌های سری‌های زمانی موجود در ماتریس تقاطع استفاده می‌شود. نمایش نهایی استخراج شده  $C_t$  به شکل زیر تعریف می‌شود:

$$C_t = \text{ReLU}(W_2^d W_1^d \tilde{S}_t)$$

(16)

که در آن  $W_1^d$  و  $W_2^d$  پارامترهای قابل آموزش لایه‌های کانولوشن هستند. لایه اول  $W_1^d$  ویژگی‌های مربوط به حرکت‌های ترافیکی را استخراج می‌کند و لایه دوم  $W_2^d$  ویژگی‌های کل تقاطع را در نظر می‌گیرد. تابع ReLU نیز برای فعال‌سازی غیرخطی به کار می‌رود. بردار ویژگی نهایی  $O_t$  از طریق یک MLP چندلایه به دست می‌آید:

$$O_t = W_u C_t + b_u$$

(17)

که  $W_u$  و  $b_u$  پارامترهای قابل آموزش لایه MLP هستند.

#### ساختار مبتنی بر RNN

بر اساس ویژگی‌های استخراج شده توسط CNN، ساختار مبتنی بر RNN که در شکل ۶ (b) نشان داده شده است، از یک لایه کانولوشن ۱D با اشتراک پارامترها برای استخراج ویژگی از هر ماتریس تقاطع استفاده می‌کند.

هر ماتریس تقاطع  $\tilde{J}_{t-K+i}$  برای  $i = 1, 2, \dots, K$  به ترتیب وارد لایه کانولوشن می‌شود و ویژگی‌های استخراج شده به صورت زیر تعریف می‌شوند:

$$c_i = \text{ReLU}(W_2^d W_1^d \tilde{J}_{t-K+i}), i = 1, 2, \dots, K$$

(18)

پارامترهای  $W_1^d$  و  $W_2^d$  قابل آموزش هستند و خروجی‌های کانولوشن به ماژول RNN داده می‌شوند:

$$h_i = \tanh(W_x c_i + h_{i-1} W_h + b_h), i = 1, 2, \dots, K$$

(19)

که در آن  $W_x, W_h$  و  $b_h$  پارامترهای لایه مخفی RNN هستند.

حالت مخفی نهایی  $h_K$  برای محاسبه ویژگی کل تقاطع در بازه زمانی مورد نظر به کار می‌رود:

$$O_t = W_v h_K + b_v$$

(20)

که  $W_v$  و  $b_v$  پارامترهای لایه خروجی RNN می‌باشند

## ساختار مبتنی بر ترانسفورمر (Transformer)

شکل ۶ (c)-ساختار مبتنی بر ترانسفورمر را نشان می‌دهد. در این ساختار، ابتدا از یک شبکه CNN با اشتراک پارامترها برای استخراج ویژگی از ماتریس تقاطع در هر بازه زمانی استفاده می‌شود (مشابه معادله ۱۸). سپس به جای استفاده از RNN، از رمزگذار ترانسفورمر برای مدل‌سازی وابستگی‌های زمانی بین توالی ویژگی‌ها بهره گرفته می‌شود. توالی ویژگی‌ها را به صورت زیر تعریف می‌کنیم:

$$\tilde{C}_t = [c_1, c_2, \dots, c_K]$$

برای وارد کردن ویژگی‌های زمان‌بندی شده، یک توکن کلاس  $c_{class}$  به توالی اضافه می‌شود:

$$C_t = \langle c_{class}, \tilde{C}_t \rangle = [c_{class}, c_1, c_2, \dots, c_K]$$

(21)

خروجی رمزگذار ترانسفورمر بر اساس این توالی به عنوان ویژگی وضعیت ترافیکی  $O_t$  استفاده می‌شود.

مکانیزم Self-Attention در رمزگذار ترانسفورمر به صورت زیر محاسبه می‌شود:

ابتدا پرجکشن‌های Query، Key و Value محاسبه می‌شوند:

$$Q_C = C_t W_Q, K_C = C_t W_K, V_C = C_t W_V$$

(22)

سپس خروجی attention محاسبه می‌شود:

$$Z_t = \phi \left( \frac{Q_C K_C^T}{\sqrt{C_d}} \right) V_C$$

(23)

که در آن  $W_Q, W_K, W_V$  پارامترهای پرجکشن و  $\phi(\cdot)$  تابع SoftMax است. خروجی  $O_t$  برداری است که معمولاً مربوط به توکن  $c_{class}$  می‌باشد.

## آموزش RL و Fine-tuning

برای آموزش مدل UniTSA از الگوریتم Proximal Policy Optimization (PPO) استفاده می‌شود. همان‌طور که در شکل ۲ نشان داده شده است، عامل در تعامل با سناریوهای ترافیکی متنوع، مسیرهایی (trajectories) از اقدامات و پاداش‌ها ایجاد می‌کند. این مسیرها مبنای محاسبه تابع‌های هزینه هستند:

Loss سیاست (Policy Loss): نشان‌دهنده فاصله بین سیاست فعلی و سیاست گذشته بر اساس مسیرهای جمع‌آوری شده است و عامل را تشویق می‌کند تا اقدامات منجر به پاداش بالاتر را بیشتر انتخاب کند.

$$L_{pf}(\theta) = \hat{\mathbb{E}}_t [\min(r_t A_t, \text{clip}(r_t, 1 - \epsilon, 1 + \epsilon) A_t)]$$

(24)

که  $r_t = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$  است.

Loss ارزش: (Value Loss) فاصله بین مقدار پیش‌بینی شده تابع ارزش  $V_\theta(s_t)$  و مجموع پاداش‌های مورد انتظار  $\hat{R}_t$  را اندازه‌گیری می‌کند:

$$L_{vf}(\theta) = \mathbb{E}_t[(V_\theta(s_t) - \hat{R}_t)^2]$$

(26)

تابع هدف نهایی به صورت ترکیبی از این دو تعریف می‌شود:

$$L(\theta) = -L_{pf}(\theta) + \lambda L_{vf}(\theta)$$

(27)

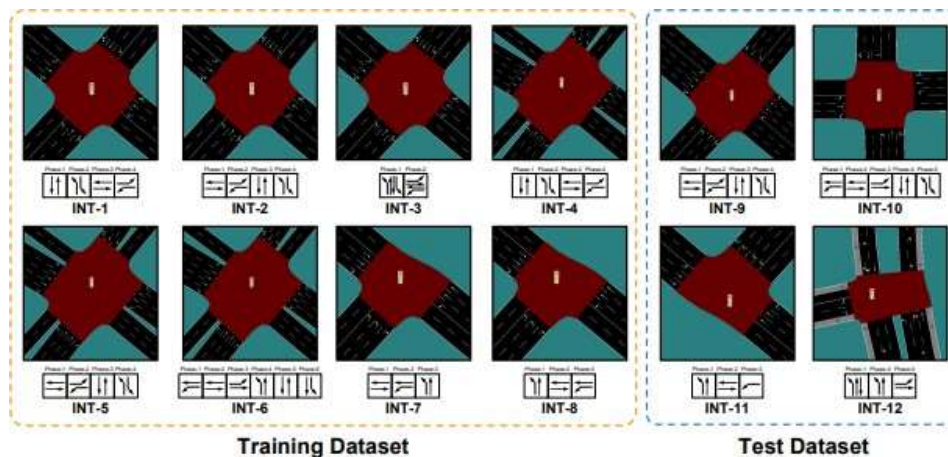
که  $\lambda$  وزن‌دهی بین دو تابع را مشخص می‌کند.

---

## LoRA با Fine-tuning

برای تسهیل فاین‌تیون بر روی تقاطع‌های بحرانی، از ماژول‌های LoRA استفاده می‌شود. این ماژول‌ها به لایه‌های Dense شبکه Actor و Critic اضافه شده و تنها این بخش‌ها در فرآیند فاین‌تیون به‌روزرسانی می‌شوند. این روش امکان افزایش پارامترهای قابل آموزش بدون افزایش بیش از حد پیچیدگی مدل را فراهم می‌کند و در عین حال عملکرد شبکه اصلی حفظ می‌شود.





شکل ۷: توپولوژی تقاطع‌ها و فازهای موجود

شکل ۷ نمایی از توپولوژی‌های تقاطع مورد استفاده در آزمایش‌ها و فازهای موجود آن‌ها را نشان می‌دهد. این توپولوژی‌ها برای آموزش و ارزیابی مدل پیش‌بینی‌کننده به کار گرفته شدند.

## ۵. آزمایش‌ها

### ۵-۱. تنظیمات آزمایش

برای ارزیابی مدل پیشنهادی، شبیه‌سازی‌های گسترده با استفاده از نرم‌افزار SUMO انجام شد. SUMO یک شبیه‌ساز ترافیک میکروسکوپی متن‌باز است که امکان مدیریت شبکه‌های بزرگ را فراهم می‌کند. این نرم‌افزار از رابط TraCI (Traffic Control Interface) پشتیبانی می‌کند و اجازه کنترل چراغ‌های ترافیکی و استخراج وضعیت ترافیکی هر تقاطع را می‌دهد. جریان ترافیک و میزان اشغال (occupancy) هر مسیر با تحلیل موقعیت‌ها و مسیرهای خودروها محاسبه می‌شود.

تنها خودروهایی که در فاصله ۱۵۰ متری تقاطع قرار دارند، برای شبیه‌سازی در نظر گرفته شدند.

چراغ زرد پس از خاموش شدن چراغ سبز با مدت ۳ ثانیه اعمال شد تا رفتار رانندگان به صورت طبیعی شبیه‌سازی شود.

زمان انتظار هر خودرو به عنوان معیار عملکرد استفاده شد؛ مقدار کمتر نشان‌دهنده عبور سریع‌تر خودروها از تقاطع است.

برای اجرای مدل از کتابخانه Stable-Baselines3 و الگوریتم PPO استفاده شد. محیط شبیه‌سازی شامل ۳۰ پردازنده موازی بود و تعداد کل گام‌های محیط برای آموزش برابر با ۱۰ میلیون تعیین شد. هر حالت شامل ۸ نمونه ماتریس تقاطع بود و فاصله زمانی بین اقدامات متوالی ۵ ثانیه در نظر گرفته شد.

### جدول ۱: تنظیمات ابرپارامترها

شبکه‌های Actor و Critic از لایه‌های کاملاً متصل (Dense) تشکیل شده‌اند. ابعاد ورودی، مخفی و خروجی برای Actor به ترتیب 64, 32, 32 و برای Critic 32, 1 تنظیم شد.

### جدول ۲: پیکربندی تقاطع‌ها

تعداد و نوع مسیرها، تعداد فازها و توپولوژی هر تقاطع در این جدول مشخص شده است.

## ۵-۲. مجموعه داده‌ها

برای آموزش مدل، ۱۲ تقاطع با توپولوژی‌های متنوع در نظر گرفته شد تا مدل یکپارچه‌ای برای پیش‌بینی اقدامات در تمام تقاطع‌ها ایجاد شود.

INT-1 تا INT-3: تقاطع‌های ۴ راهه کوچک با یک فاز در هر مسیر.

INT-4 تا INT-6: تقاطع‌های ۴ راهه بزرگ با بیش از چهار مسیر در هر مسیر.

INT-7 تا INT-8: تقاطع‌های ۳ راهه با توپولوژی خاص.

INT-9 تا INT-12: تقاطع‌های جدید برای ارزیابی مدل در محیط‌های دیده‌نشده.

برای هر تقاطع، ۱۰۰ مسیر منحصر به فرد تولید شد و طول هر مسیر تقریباً ۳۰،۰۰۰ ثانیه شبیه‌سازی شد.

### ۳-۵ روش‌های مقایسه‌ای

برای ارزیابی عملکرد مدل UniTSA، با روش‌های پیشرفته RL و روش‌های سنتی کنترل ترافیک مقایسه شد:

[2]: FixTime کنترل زمان ثابت چراغ‌ها (۳۰ و ۴۰ ثانیه).

[3]: Webster زمان‌بندی فازها بر اساس حجم ترافیک و کاهش زمان انتظار.

[4]: SOTL کنترل دینامیک چراغ‌ها بر اساس تعداد خودروهای منتظر.

[10]: MPLight استفاده از شبکه MLP با اشتراک پارامتر و مکانیزم پاداش مبتنی بر فشار برای کنترل فازها.

[18]: AttendLight مدل توجه‌محور برای مدیریت تقاطع‌های با جریان متغیر.

مدل‌های UniTSA شامل:

Single: استفاده از RNN برای استخراج ویژگی وضعیت ترافیک در یک محیط.

Multi: همزمان در چند محیط با توپولوژی‌های مختلف، استفاده از CNN، RNN و Transformer.

Multi+TSA: استفاده از تکنیک Traffic State Augmentation برای افزایش تنوع داده‌ها و بهبود عملکرد.

### ۴-۵ نتایج تقاطع‌های آموزش

نتایج متوسط زمان انتظار هر خودرو در جدول ۳ ارائه شده است.

روش‌های مبتنی بر RL در اکثر تقاطع‌ها عملکرد بهتری نسبت به روش‌های سنتی داشتند.

SOTL در INT-2 زمان انتظار بالایی داشت، اما نیاز به اطلاعات دقیق ترافیکی داشت.

UniTSA نسبت به MPLight و AttendLight بهبود قابل توجهی داشت: میانگین کاهش زمان انتظار حدود ۱۵٪ و ۱۲٪ نسبت به MPLight و AttendLight مشاهده شد.

مزایای UniTSA:

با اشتراک پارامترها و جایگزینی RNN یا Transformer، ویژگی‌های زمانی ترافیک به‌صورت بهتری استخراج می‌شوند.

امکان مواجهه با تنوع بیشتری از وضعیت‌های تقاطع در حین آموزش فراهم است.

بررسی ساختار شبکه‌ها نشان می‌دهد که RNN نسبت به CNN عملکرد بهتری دارد.

جدول ۱) تنظیمات ابر پارامترها

Hyper-parameter	Value
Learning rate	0.0001
Trajectory memory size	3000
Clipping range $\epsilon$	0.2
Discount factor $\gamma$	0.99
Value function coefficient $\lambda$	0.9
Scale hyperparameter $\alpha$	1
Rank of LoRA module	8

ابر پارامترها نیز مطابق جدول ۱ تنظیم شدند. شبکه‌های Actor و Critic به صورت دو لایه کاملاً متصل طراحی شدند. ابعاد ورودی  $\{۳۲, ۶۴\}$  و ابعاد خروجی  $\{۲, ۳۲\}$  برای Actor و  $\{۱, ۳۲\}$  برای Critic تنظیم شد.

جدول ۲) تمام تنظیمات تقاطع‌ها

Intersection ID	Training Dataset								Test Dataset			
	INT-1	INT-2	INT-3	INT-4	INT-5	INT-6	INT-7	INT-8	INT-9	INT-10	INT-11	INT-12
roads	4	4	4	4	4	4	3	3	4	4	3	3
lanes per road	(3,3,3,3)	(3,3,3,3)	(3,3,3,3)	(3,4,4,5)	(3,4,4,5)	(3,4,4,5)	(3,3,3)	(3,3,3)	(3,4,3,4)	(3,3,3,3)	(4,3,3)	(2,3,2)
phases	4	4	2	4	4	6	3	3	4	5	3	3

جدول ۳) نتایج کمی (میانگین زمان انتظار به ازای هر وسیله نقلیه) تقاطع‌های آموزشی برای مدل‌های جهانی. مقدار کمتر نشان دهنده عملکرد بهتر است و کمترین مقادیر با حروف پررنگ مشخص شده‌اند.

	INT-1	INT-2	INT-3	INT-4	INT-5	INT-6	INT-7	INT-8
Fix-30 [2]	39.458	38.862	8.323	35.123	35.031	51.923	18.920	18.595
Fix-40 [2]	50.696	52.108	10.667	44.888	45.540	61.743	23.411	24.759
Webster [3]	26.466	26.889	5.751	25.413	24.541	40.128	12.755	13.574
SOTL [4]	16.048	16.561	<b>2.764</b>	28.169	27.902	27.404	8.639	7.925
MPLight [10]	19.111	14.659	4.469	16.067	19.925	19.115	6.654	7.523
AttendLight [18]	16.483	13.893	3.860	16.903	18.915	20.795	6.532	8.104
UniTSA (Multi+CNN)	13.776	13.580	3.265	14.790	15.437	18.751	6.592	6.894
UniTSA (Multi+RNN)	13.692	13.437	3.495	15.198	14.896	18.612	6.393	6.625
UniTSA (Multi+Trans)	14.976	20.459	2.811	16.489	16.481	23.793	7.552	7.175
UniTSA (Multi+CNN+TSA)	14.071	14.150	3.036	15.990	16.472	24.383	6.329	6.328
UniTSA (Multi+RNN+TSA)	13.450	13.578	3.007	<b>14.470</b>	<b>14.462</b>	18.689	<b>6.242</b>	<b>6.164</b>
UniTSA (Multi+Trans+TSA)	<b>13.335</b>	<b>13.314</b>	3.311	14.648	14.566	<b>18.579</b>	6.643	6.571

جدول ۴) نتایج کمی تقاطع‌های آزمون برای مدل‌های جهانی

	INT-9	INT-10	INT-11	INT-12
Fix-30	40.341	38.360	17.136	17.440
Fix-40	60.043	50.580	23.504	20.570
Webster	26.191	27.766	11.981	13.280
SOTL	23.031	28.066	7.452	8.070
MPLight	23.674	21.475	8.447	15.047
AttendLight	18.080	18.501	7.393	12.982
UniTSA (Multi+CNN)	17.877	18.244	7.063	12.820
UniTSA (Multi+RNN)	16.771	15.450	6.807	10.140
UniTSA (Multi+Trans)	17.640	16.269	6.598	9.630
UniTSA (Multi+CNN+TSA)	13.075	14.245	6.794	7.550
UniTSA (Multi+RNN+TSA)	<b>12.677</b>	<b>14.208</b>	<b>5.914</b>	6.730
UniTSA (Multi+Trans+TSA)	13.054	16.186	5.983	<b>6.190</b>

قابل توجه است که استفاده از تکنیک‌های تقویت داده‌های ترافیکی باعث بهبود عملکرد مدل UniTSA در تقاطع‌های مختلف شده است. بر اساس نتایج آزمایش‌ها:

UniTSA (Multi+RNN+TSA)، UniTSA (Multi+Trans+TSA) و UniTSA (Multi+RNN+TSA) به ترتیب بهبودهای ۲۳.۴٪، ۱۹.۸٪ و ۱۷.۹٪ را نسبت به نسخه پایه UniTSA (Multi+RNN) نشان دادند.

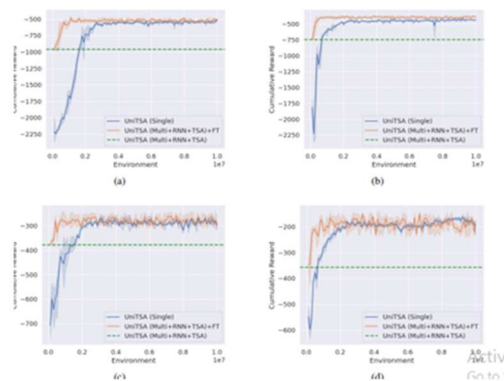
این بهبودها ناشی از افزایش تنوع در سناریوهای تقاطع در داده‌های آموزش است. تکنیک‌های تقویت داده ترافیکی مانند «جابجایی کل مجموعه فازها» امکان تولید ترکیب‌های متنوع از پیکربندی‌های فاز را فراهم می‌کند و مدل می‌تواند با سناریوهای بیشتری مواجه شود.

**جدول ۴** عملکرد روش‌های مختلف را نشان می‌دهد و مقایسه بین الگوریتم‌های RL و روش‌های سنتی کنترل ترافیک را برجسته می‌کند. نتایج نشان می‌دهد که: مدل‌های مبتنی بر RL عملکرد قابل توجهی نسبت به روش‌های سنتی دارند.

در میان روش‌های RL، مدل‌های UniTSA در تمام تقاطع‌های آزمایشی عملکرد بهتری ارائه کردند.

بهترین عملکرد در کل آزمایش‌ها توسط UniTSA (Multi+Trans+TSA) و پس از آن UniTSA (Multi+RNN+TSA) به دست آمد.

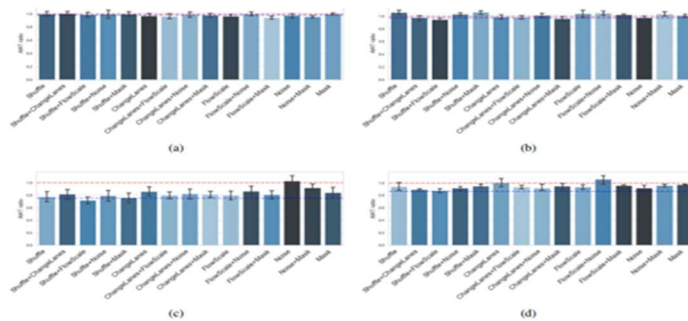
شکل ۸ نمایی از تقاطع‌های INT-9، INT-10، INT-11 و INT-12 را نشان می‌دهد که برای ارزیابی و مقایسه عملکرد مدل استفاده شدند.



شکل ۸ (مراحل محیطی روش‌های مختلف در چهار تقاطع آزمون). (الف) INT-9. (ب) INT-10. (ج) INT-11. (د) INT-12.

جدول ۵) نتایج کمی تنظیم دقیق در تقاطع‌های آزمایشی

	INT-9	INT-10	INT-11	INT-12
UniTSA (Multi+RNN)	16.771	15.450	6.807	10.140
UniTSA (Multi+RNN+TSA)	12.677	14.208	5.914	6.730
<i>IM Environment Steps</i>				
UniTSA (Single)	23.683	14.863	8.417	5.024
UniTSA (Multi+RNN+TSA) + FT	<b>10.995</b>	<b>12.553</b>	<b>4.475</b>	<b>3.534</b>
<i>IOM Environment Steps</i>				
UniTSA (Single)	10.353	12.254	4.359	<b>3.089</b>
UniTSA (Multi+RNN+TSA) + FT	<b>10.292</b>	<b>11.081</b>	<b>4.153</b>	3.110



شکل ۹) تحلیل مقایسه‌ای روش‌های تقویت وضعیت ترافیک در تقاطع‌های آموزشی و آزمایشی انتخاب‌شده (الف) INT-1. (ب) INT-7. (ج) INT-9. (د) INT-11

## ۵-۶ نتایج فایننتیون در تقاطع‌های آزمون

در سناریوهای پیچیده تقاطع‌ها، به دلیل نیاز به توجه ویژه به رفتارهای ترافیکی، ارزیابی مدل بر روی داده‌های واقعی اهمیت دارد. برای این منظور، مدل UniTSA (Multi+RNN) به عنوان مدل پایه انتخاب شد، زیرا ساختار RNN در استخراج ویژگی‌های زمانی ترافیک نسبت به شبکه‌های CNN و Transformer عملکرد بهتری در تقاطع‌های کم‌تجربه نشان می‌دهد.

در مقایسه با UniTSA (Single) که فقط بر یک سناریو آموزش دیده است، مدل Multi+RNN توانست عملکردی مشابه یا حتی بهتر را با مواجهه با چند مرحله آموزش نشان دهد.

شکل ۸ روند اجرای مدل در مراحل مختلف فایننتیون برای تقاطع‌های آزمایشی را نشان می‌دهد. نتایج نشان می‌دهد که مدل یادگیری شده بدون هیچ فایننتیون اضافی نیز قادر به ارائه عملکرد قابل قبول در تقاطع‌های جدید است. نکته جالب آن است که حتی پس از آموزش محدود، مدل می‌تواند نتایج امیدوارکننده‌ای ارائه دهد.

مطابق شکل، مدل فایننتیون‌شده (با رنگ نارنجی) عملکرد بهتری نسبت به مدل پایه دارد. مشاهده شد که مدل تک‌محیطی تقریباً پس از ۳ میلیون مرحله آموزش همگرا می‌شود، در حالی که مدل فایننتیون‌شده تنها با ۱ میلیون مرحله آموزش، عملکرد مشابه را با حدود ۶۶٪ کاهش زمان محاسبات به دست می‌آورد.

جدول ۵ جزئیات عملکرد مدل پس از فایننتیون را نشان می‌دهد. در تنها ۱ میلیون مرحله، مدل فایننتیون‌شده بهبود متوسط ۳۶٪ نسبت به UniTSA (Single) در چهار تقاطع آزمایشی نشان داد. حتی پس از ۱۰ میلیون مرحله آموزش، مدل فایننتیون‌شده همچنان حدود ۳٪ بهتر از مدل‌های بدون فایننتیون عمل کرد. این ویژگی برای کاربردهای بلادرنگ بسیار مهم است، زیرا در شبکه‌های بزرگ با بیش از ۱۰۰۰ تقاطع،

زمان آموزش و سازگاری با محیط به شکل قابل توجهی بهبود می‌یابد و عملکرد سیستم کنترل ترافیک پایدار باقی می‌ماند.

## ۵-۷ تحلیل مقایسه‌ای روش‌های تقویت داده ترافیکی

در این بخش، تأثیر روش‌های تقویت داده‌های ترافیکی بر بهبود عملکرد مدل **UniTSA (Multi+RNN)** بررسی شد. آزمایش‌ها با استفاده از ترکیب‌های متنوعی از تکنیک‌های تقویت داده بر روی تقاطع‌های آزمون انجام شد. برای ارزیابی تأثیر این روش‌ها، نسبت میانگین زمان انتظار (**AWT ratio**) بین مدل‌ها با و بدون تقویت داده محاسبه شد. مقادیر بالاتر از ۱ نشان‌دهنده بهبود عملکرد ناشی از استفاده از روش‌های تقویت داده است. شکل ۹ نتایج مقایسه‌ای را برای تقاطع‌های **INT-1**، **INT-7**، **INT-9** و **INT-11** نشان می‌دهد. در این نمودارها، میله‌ها نسبت میانگین **AWT** برای مدل با تقویت داده را نشان می‌دهند و نوارهای خطای ۹۵٪ اطمینان، پراکندگی مقادیر را مشخص می‌کنند. نتایج به وضوح نشان می‌دهد که مدل **UniTSA (Multi+RNN+TSA)**، که از تکنیک‌های تقویت داده ترافیکی استفاده می‌کند، عملکرد بهتری نسبت به نسخه بدون تقویت داده دارد و میانگین زمان انتظار در تمامی تقاطع‌ها کاهش یافته است.

## تحلیل نتایج شکل ۹

شکل ۹ (a) و (b) نتایج مربوط به تقاطع‌های **INT-1** و **INT-7** را نشان می‌دهد. بررسی این نتایج نشان می‌دهد که بهبود متوسط عملکرد ناشی از استفاده از روش‌های تقویت داده ترافیکی حدود ۲٪ بوده است. همچنین استفاده از روش‌های **Noise** و **Mask** در **INT-7**، تأثیر قابل توجهی بر بهبود عملکرد نداشته است. این موضوع می‌تواند به دلیل آن باشد که مدل قبلاً الگوها و ویژگی‌های پایدار تقاطع‌های تمرینی را به خوبی یاد گرفته و افزودن داده‌های مصنوعی از طریق تقویت داده، مزیت قابل توجهی برای آن ایجاد نکرده است. با این حال، روش‌های تقویت داده ترافیکی در ترکیب با ویژگی‌های تقاطع دیدنشده عملکرد قابل توجهی ارائه دادند. شکل ۹ (c) و (d) نسبت‌های **AWT** برای تقاطع‌های آزمایشی **INT-9** و **INT-11** را نشان می‌دهد. نتایج حاکی از آن است که اغلب روش‌های تقویت داده، عملکرد سیاست پایدار مدل را در تقاطع‌های آزمون بهبود بخشیده‌اند. تنوع داده‌های ایجادشده از طریق تقویت داده، به مدل اجازه داد تا از سناریوهای بیشتری یاد گرفته و رفتار خود را تطبیق دهد، در نتیجه عملکرد در مجموعه داده‌های آزمون افزایش یافته است. روش‌های **Shuffle Movement** و **Scale Flow Traffic** در میان تکنیک‌های مورد استفاده مؤثرترین بودند، زیرا امکان یادگیری و تطبیق بهتر مدل با تنوع سناریوها را فراهم کردند و در نتیجه عملکرد در مجموعه آزمون بهبود یافت.

## ۶- نتیجه‌گیری

در این مقاله، یک چارچوب کنترل چراغ‌های ترافیک مبتنی بر یادگیری تقویتی (**RL**) با توانایی تعمیم‌پذیری (**UniTSA**) برای ساختارهای مختلف تقاطع در محیط‌های **V2X** ارائه شد.

مدل **UniTSA** امکان آموزش یک عامل **RL** هوشمند را با استفاده از ماتریس‌های تقاطع فراهم می‌کند و قادر است اقدامات بهینه در تقاطع‌های ناشناخته را پیش‌بینی کند. برای مقابله با تقاطع‌های دیده‌نشده، روش‌های تقویت داده ترافیکی ارائه شد که داده‌های آموزشی را گسترش داده و قابلیت تعمیم مدل را افزایش می‌دهد.

نتایج نشان می‌دهد که **UniTSA** نیاز به آموزش مجدد برای هر تقاطع منفرد را حذف کرده و در عین حال، معماری پویا و باز را برای کنترل

چراغ‌های ترافیک مبتنی بر RL ارائه می‌دهد. تجربیات نشان داد که **UnitSA** کمترین زمان انتظار متوسط را در تقاطع‌های مختلف ارائه کرده و عملکرد بهتری نسبت به روش‌های موجود دارد، حتی نسبت به مدل‌های آموزش‌دیده با داده‌های محدود و بدون فاین‌تونیگ، عملکرد بهتری نشان می‌دهد.

---

#### قدردانی (Acknowledgments)

این تحقیق تحت حمایت برنامه ملی چین در حوزه تحقیقات و توسعه (92020YFB1807700) و پروژه 21PJD092 انجام شده است. تقدیر ویژه از **Shanghai Pujiang Lab** بابت همکاری و حمایت علمی ارائه می‌شود.

## References

- [1] Fehda Malik, Hasan Ali Khattak, and Munam Ali Shah. Evaluation of the impact of traffic congestion based on sumo. In 2019 25th International Conference on Automation and Computing (ICAC), pages 1–5. IEEE, 2019.
- [2] Alan J Miller. Settings for fixed-cycle traffic signals. *Journal of the Operational Research Society*, 14(4):373–386, 1963.
- [3] Thomas Urbanik, Alison Tanaka, Bailey Lozner, Eric Lindstrom, Kevin Lee, Shaun Quayle, Scott Beaird, Shing Tsoi, Paul Ryus, Doug Gettman, et al. Signal timing manual, volume 1. Transportation Research Board Washington, DC, 2015.
- [4] Carlos Gershenson. Self-organizing traffic lights. arXiv preprint nlin/0411066, 2004.
- [5] Ishu Tomar, S Indu, and Neeta Pandey. Traffic signal control methods: Current status, challenges, and emerging trends. *Proceedings of Data Analytics and Management: ICDAM 2021, Volume 1*, pages 151–163, 2022.
- [6] Wang Tong, Azhar Hussain, Wang Xi Bo, and Sabita Maharjan. Artificial intelligence for vehicle- to-everything: A survey. *IEEE Access*, 7:10823–10843, 2019.
- [7] Tamás Wágner, Tamás Ormándi, Tamás Tettamanti, and István Varga. Spat/map v2x communication between traffic light and vehicles and a realization with digital twin. *Computers and Electrical Engineering*, 106:108560, 2023.



[8] Yit Kwong Chin, Lai Kuan Lee, Nurmin Bolong, Soo Siang Yang, and Kenneth Tze Kin Teo. Exploring q-learning

optimization in traffic signal timing plan management. In 2011 third international conference on computational

intelligence, communication systems and networks, pages 269–274. IEEE, 2011.

[9] Guanjie Zheng, Yuanhao Xiong, Xinshi Zang, Jie Feng, Hua Wei, Huichu Zhang, Yong Li, Kai Xu, and Zhenhui

Li. Learning phase competition for traffic signal control. In Proceedings of the 28th ACM international conference

on information and knowledge management, pages 1963–1972, 2019.

[10] Chacha Chen, Hua Wei, Nan Xu, Guanjie Zheng, Ming Yang, Yuanhao Xiong, Kai Xu, and Zhenhui Li. Toward a

thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control.

In Proceedings of

the AAAI Conference on Artificial Intelligence, volume 34, pages 3414–3421, 2020.

[11] Xinshi Zang, Huaxiu Yao, Guanjie Zheng, Nan Xu, Kai Xu, and Zhenhui Li. Metalight: Value- based metareinforcement learning for traffic signal control. In Proceedings of the AAAI Conference on Artificial Intelligence,

volume 34, pages 1153–1160, 2020.

[12] Enming Liang, Zicheng Su, Chilin Fang, and Renxin Zhong. Oam: An option-action reinforcement learning framework for universal multi-intersection control.

In Proceedings of the AAAI Conference on Artificial Intelligence,

volume 36, pages 4550–4558, 2022.

[13] Azzedine Boukerche, Dunhao Zhong, and Peng Sun. A novel reinforcement learning-based cooperative traffic

signal system through max-pressure control. IEEE Transactions on Vehicular Technology, 71(2):1187–1198,

2022.

[14] Liang Zhang, Qiang Wu, Jun Shen, Linyuan Lü, Bo Du, and Jianqing Wu. Expression might be enough:

representing pressure and demand for reinforcement learning based traffic signal control. In International

Conference on Machine Learning, pages 26645–26654. PMLR, 2022.

[15] Yuanhao Xiong, Guanjie Zheng, Kai Xu, and Zhenhui Li. Learning traffic signal control from demonstrations.

In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pages 2289–2292, 2019.

[16] Stefano Giovanni Rizzo, Giovanna Vantini, and Sanjay Chawla. Time critic policy gradient methods for traffic

signal control in complex and congested scenarios. In Proceedings of the 25th ACM SIGKDD International

Conference on Knowledge Discovery and Data Mining, KDD ’19, page 1654–1664, New York, NY, USA, 2019.

Association for Computing Machinery.

[17] Tianshu Chu, Jie Wang, Lara Codecà, and Zhaojian Li. Multi-agent deep reinforcement learning for large-scale

traffic signal control. IEEE Transactions on Intelligent Transportation Systems, 21(3):1086–1095, 2019.

[18] Afshin Oroojlooy, Mohammadreza Nazari, Davood Hajinezhad, and Jorge Silva. Attendlight: Universal attentionbased reinforcement learning model for traffic signal control. Advances in Neural Information Processing Systems,

33:4079–4090, 2020.

[19] Zian Ma, Chengcheng Xu, Yuheng Kan, Maonan Wang, and Wei Wu. Adaptive coordinated traffic control for

arterial intersections based on reinforcement learning. In 2021 IEEE International Intelligent Transportation

Systems Conference (ITSC), pages 2562–2567. IEEE, 2021.

[20] Seyed Sajad Mousavi, Michael Schukat, and Enda Howley. Traffic light control

using deep policy-gradient and  
value-function-based reinforcement learning. IET Intelligent Transport Systems,  
11(7):417–423, 2017.

[21] Mohammad Aslani, Mohammad Saadi Mesgari, Stefan Seipel, and Marco Wiering.

Developing adaptive traffic

signal control by actor–critic and direct exploration methods. In Proceedings of the Institution of Civil EngineersTransport, volume 172, pages 289–298. Thomas Telford Ltd, 2019.

[22] Haoran Su, Yaofeng D Zhong, Joseph YJ Chow, Biswadip Dey, and Li Jin. Emvlight: A multi- agent reinforcement learning framework for an emergency vehicle decentralized routing and traffic signal control system.

Transportation Research Part C: Emerging Technologies, 146:103955, 2023.

[23] Elise Van der Pol and Frans A Oliehoek. Coordinated deep reinforcement learners for traffic light control.

Proceedings of learning, inference and control of multi-agent systems (at NIPS 2016), 8:21–38, 2016.

[24] Patrick Mannion, Jim Duggan, and Enda Howley. An experimental review of reinforcement learning algorithms

for adaptive traffic signal control. Autonomic road transport support systems, pages 47–66, 2016.

[25] Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. Intellilight: A reinforcement learning approach for

intelligent traffic light control. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge

Discovery & Data Mining, pages 2496–2505, 2018.

[26] Lun-Hui Xu, Xin-Hai Xia, and Qiang Luo. The study of reinforcement learning for traffic self- adaptive control

under multiagent markov game environment. Mathematical Problems in Engineering, 2013, 2013.

[27] Mohammad Aslani, Mohammad Saadi Mesgari, and Marco Wiering. Adaptive traffic signal control with actorcritic methods in a real-world traffic network with different traffic disruption events. *Transportation Research Part*

*C: Emerging Technologies*, 85:732–752, 2017.

[28] Mohammad Aslani, Stefan Seipel, Mohammad Saadi Mesgari, and Marco Wiering. Traffic signal optimization

through discrete and continuous reinforcement learning with robustness analysis in downtown tehran. *Advanced*

*Engineering Informatics*, 38:639–655, 2018.

[29] Halit Bugra Tulay and Can Emre Koksak. Road state inference via channel state information.

*IEEE Transactions*

on Vehicular Technology, pages 1–14, 2023.

[30] Mohamed MG Farag, Hesham A Rakha, Emadeldin A Mazied, and Jayanthi Rao. Integration large-scale modeling

framework of direct cellular vehicle-to-all (c-v2x) applications. *Sensors*, 21(6):2127, 2021.

[31] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu

Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

[32] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich,

Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner.

Microscopic traffic simulation using

sumo. In 2018 21st international conference on intelligent transportation systems (ITSC), pages 2575–2582.

IEEE, 2018.

[33] Peter Koonce and Lee Rodegerdts. Traffic signal timing manual. Technical report, United States. Federal Highway

Administration, 2008.



[34] Arthur G Sims and Kenneth W Dobinson. The sydney coordinated adaptive traffic (scat) system philosophy and

benefits. *IEEE Transactions on vehicular technology*, 29(2):130–137, 1980.

[35] Pravin Varaiya. The max-pressure controller for arbitrary networks of signalized intersections. *Advances in*

dynamic network modeling in complex transportation systems, pages 27–66, 2013.

[36] Hua Wei, Guanjie Zheng, Vikash Gayah, and Zhenhui Li. A survey on traffic signal control methods. arXiv

preprint arXiv:1904.08117, 2019.

[37] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A

survey. IEEE transactions on pattern analysis and machine intelligence, 44(9):5149–5169, 2021.

[38] Misha Laskin, Kimin Lee, Adam Stooke, Lerrel Pinto, Pieter Abbeel, and Aravind Srinivas.

Reinforcement

learning with augmented data. Advances in neural information processing systems, 33:19884–19895, 2020.

[39] Ilya Kostrikov, Denis Yarats, and Rob Fergus. Image augmentation is all you need: Regularizing deep reinforcement learning from pixels. arXiv preprint arXiv:2004.13649, 2020.

[40] Nicklas Hansen and Xiaolong Wang. Generalization in reinforcement learning by soft data augmentation. In 2021

IEEE International Conference on Robotics and Automation (ICRA), pages 13611–13617. IEEE, 2021.

[41] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization

algorithms. arXiv preprint arXiv:1707.06347, 2017.

[42] Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stablebaselines3: Reliable reinforcement learning implementations. The Journal of Machine Learning Research,

22(1):12348–12355, 2021.

[43] James Ault and Guni Sharon. Reinforcement learning benchmarks for traffic

signal control. In Thirty-fifth

Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1), 2021.

[44] Seung-Bae Cools, Carlos Gershenson, and Bart D’Hooghe. Self-organizing traffic lights: A realistic simulation.

Advances in applied self-organizing systems, pages 45–55, 2013.

[45] Hua Wei, Chacha Chen, Guanjie Zheng, Kan Wu, Vikash Gayah, Kai Xu, and Zhenhui Li.

Presslight: Learning

max pressure control to coordinate traffic signals in arterial network. In Proceedings of the 25th ACM SIGKDD

International Conference on Knowledge Discovery & Data Mining, pages 1290–1298, 2019