

AI-Powered Traffic Monitoring System

Nasrin Mahian *

Bachelor of Civil Engineering, Transportation
and Traffic Orientation, University of Applied
Science and Technology for Municipalities.

Abstract

Manual traffic monitoring can be a daunting task, as traffic management centers operate a vast number of cameras installed across networks. Introducing a level of automation can reduce the workload of human operators performing manual monitoring and facilitate proactive decision-making. These measures help mitigate the impact of incidents and recurring congestion on roads. This paper presents a novel approach for automated real-time traffic video surveillance using convolutional neural networks (CNNs) and an independent graphical user interface. The authors describe the findings from research conducted during the development of models designed to function as an integrated framework for an AI-enabled active traffic monitoring system. The proposed system utilizes several advanced deep learning algorithms to automate various traffic surveillance tasks. Using a large-scale, annotated surveillance video dataset, deep learning models are trained to identify queues, track stationary vehicles, and tabulate vehicle counts. A pixel-level segmentation approach is applied to detect traffic queues and predict their severity. Real-time object detection algorithms, combined with various tracking systems, are deployed to automatically identify wandering vehicles and perform vehicle counting. At each stage of development, significant empirical results are presented to demonstrate the system's effectiveness. Overall, the results show that the proposed framework performs satisfactorily under diverse conditions, without being significantly affected by environmental challenges such as camera blur, low lighting, rain, or snow.

Keywords: traffic monitoring, intelligent transportation systems, traffic queues, vehicle counting, artificial intelligence, deep learning

Received: 15/April/2025

Accepted: 11/June/2025

eISSN: 3060-6144

ISSN: 2980-8936

سیستم نظارت بر ترافیک با قابلیت هوش مصنوعی

کارشناسی عمران، گرایش حمل‌ونقل ترافیک، دانشگاه جامع علمی کاربردی
شهرداری‌ها.

نسرین ماهیان *

چکیده

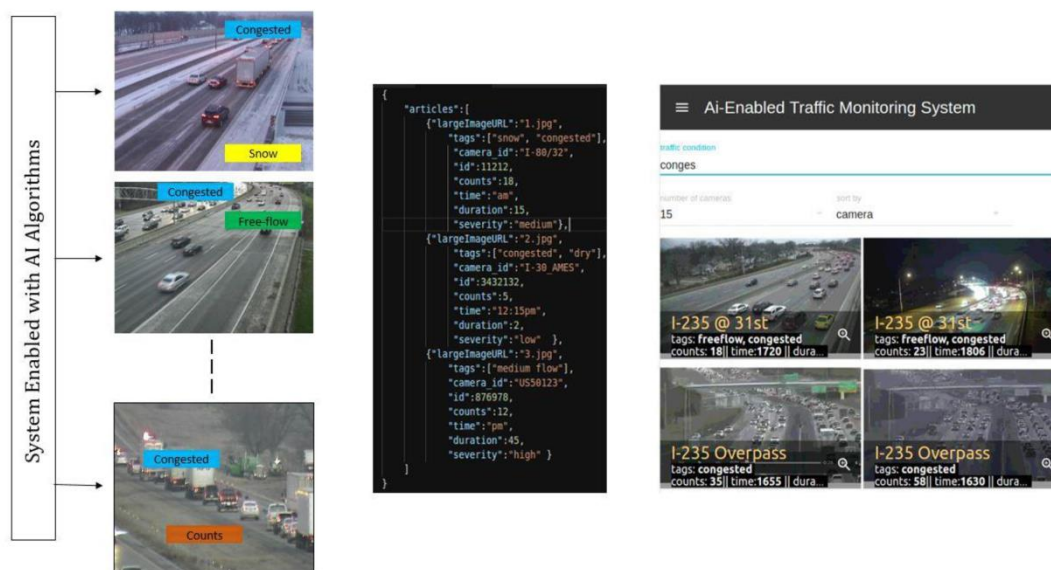
نظارت دستی ترافیک می‌تواند یک کار دلهره‌آور باشد؛ زیرا مراکز مدیریت ترافیک، تعداد بی‌شماری از دوربین‌های نصب‌شده روی یک شبکه را اجرا می‌کنند. تزریق سطحی از اتوماسیون می‌تواند به کاهش بار کار اپراتورهای انسانی که نظارت دستی را انجام می‌دهند، کمک کند و تصمیم‌گیری‌های پیشگیرانه را تسهیل نماید. این اقدامات، تأثیر حوادث و ازدحام مکرر در جاده‌ها را کاهش می‌دهد. این مقاله، یک رویکرد جدید برای نظارت خودکار فیلم‌های ترافیکی زمان واقعی با استفاده از شبکه‌های عصبی پیچیده و یک رابط کاربری گرافیکی مستقل را ارائه می‌کند. نویسندگان، نتایج تحقیقات دریافت‌شده در فرآیند توسعه مدل‌هایی را توصیف می‌کنند که به‌عنوان یک چارچوب یکپارچه برای یک سیستم نظارت بر ترافیک فعال با هوش مصنوعی عمل می‌کند. سیستم پیشنهادی، چندین الگوریتم یادگیری عمیق پیشرفته را جهت خودکارسازی نیازهای مختلف نظارت بر ترافیک به کار می‌گیرد. با استفاده از یک پایگاه داده بزرگ از داده‌های نظارت تصویری مشروح، مدل‌های مبتنی بر یادگیری عمیق برای شناسایی صف‌ها، ردیابی وسایل نقلیه ساکن و جدول‌بندی تعداد خودروها آموزش داده می‌شوند. یک رویکرد تقسیم‌بندی در سطح پیکسل به‌منظور شناسایی صف‌های ترافیک و پیش‌بینی شدت اعمال می‌شود. الگوریتم‌های تشخیص شیء در زمان واقعی همراه با سیستم‌های ردیابی مختلف برای شناسایی خودکار وسایل نقلیه سرگردان و همچنین، انجام شمارش وسایل نقلیه مستقر می‌شوند. در هر مرحله از توسعه، نتایج تجربی قابل توجهی برای نمایش اثربخشی سیستم پیشنهادی ارائه می‌شود. به‌طور کلی، نتایج نشان می‌دهد که چارچوب پیشنهادی در شرایط مختلف به‌طور رضایت‌بخشی عمل می‌کند؛ بدون اینکه به‌شدت تحت تأثیر خطرات محیطی مانند تاری نماهای دوربین، نور کم، باران یا برف قرار گیرد.

کلیدواژه‌ها: نظارت بر ترافیک، سیستم‌های حمل‌ونقل هوشمند، صف‌های ترافیکی، شمارش وسایل نقلیه، هوش مصنوعی، یادگیری عمیق

۱- مقدمه

نظارت مؤثر بر ترافیک، از دیرباز یکی از مهم‌ترین تلاش‌ها در مهندسی حمل‌ونقل بوده است. تا به امروز، اکثر مراکز نظارت بر ترافیک برای ردیابی ماهیت جریان ترافیک و نظارت بر هر حادثه‌ای که در جاده‌ها اتفاق می‌افتد، به اپراتورهای انسانی متکی هستند. فرآیندهای مربوط به نظارت دستی وضعیت ترافیک می‌تواند چالش‌برانگیز و زمان‌بر باشد. از آنجاکه انسان‌ها مستعد خطا بوده و در معرض خستگی هستند، نتایج اغلب شامل اختلافات خاصی است؛ بنابراین، توسعه ابزارهای نظارت خودکار ترافیک برای کاهش بار کاری اپراتورهای انسانی و افزایش کارایی خروجی به نفع همه است. از این‌رو، جای تعجب نیست که سیستم‌های نظارت خودکار ترافیک، یکی از مهم‌ترین تلاش‌های تحقیقاتی در سیستم‌های حمل‌ونقل هوشمند بوده است. شایان ذکر است که بیشتر فعالیت‌های نظارت بر ترافیک امروزی در مراکز مدیریت ترافیک (TMCs) از طریق سیستم‌های دوربین مبتنی بر دید انجام می‌شود. با این حال، اکثر سیستم‌های مبتنی بر بینایی موجود توسط انسان نظارت می‌شوند که پیگیری دقیق تراکم، شناسایی وسایل نقلیه ساکن و هم‌زمان، پیگیری دقیق تعداد وسایل نقلیه را دشوار می‌کند. بدین ترتیب، TMCها تلاش‌هایی را برای تحقق سطوحی از اتوماسیون در مدیریت ترافیک انجام داده‌اند. سیستم‌های نظارت خودکار ترافیک با استفاده از هوش مصنوعی (AI)، نه تنها این قابلیت را دارند که ترافیک را به خوبی مدیریت کنند بلکه بر شرایط فعلی نظارت و دسترسی دارند که می‌تواند تعداد تصادفات جاده‌ای را کاهش دهد. به طور مشابه، یک سیستم مجهز به هوش مصنوعی می‌تواند هر وسیله نقلیه را شناسایی نموده و همچنین، ویژگی الگوی حرکت آن را برای شناسایی هرگونه رفتار خطرناک رانندگی مانند تغییر مسیر نامنظم، ردیابی کند. یکی دیگر از جنبه‌های مهم سیستم نظارت بر ترافیک با هوش مصنوعی، تشخیص صحیح هر وسیله نقلیه ساکن در جاده است. اغلب اوقات، وسایل نقلیه ثابتی وجود دارند که رها شده و جریان وسایل نقلیه قبلی را مختل می‌کنند که باعث ایجاد ترافیک می‌شود. این ازدحام، حرکت آزادانه وسایل نقلیه را مختل می‌کند؛ بنابراین، سیستم‌های نظارت هوشمند ترافیک، جزء لاینفک سیستم‌های موردنیاز برای شناسایی سریع و کاهش اثرات ازدحام ترافیک و عوامل انسانی هستند.

در چند سال اخیر، تحقیقات گسترده‌ای در مورد سیستم‌های نظارت بر ترافیک مبتنی بر یادگیری عمیق و ماشینی انجام شده است. اکثر TMCها به دنبال استقرار سیستم‌های خودکاری هستند که می‌توانند بار کاری اپراتورهای انسانی را کاهش دهند و به سیستم مدیریت ترافیک مؤثر منجر شوند. در عین حال، هزینه‌های مرتبط به دلیل صرفه‌جویی در ارتباط با عدم نیاز به ذخیره چندین ساعت داده‌های ویدیویی بزرگ، نسبتاً کمتر است. در این مطالعه، ما چندین الگوریتم یادگیری عمیق پیشرفته را بر اساس ماهیت برخی عملیات ترافیکی موردنیاز مستقر کردیم. الگوریتم‌های سنتی (Land, 1986; Rahman et al., 1996; He et al., 2010)، اغلب دقت کمتری را ثبت نموده و در ثبت الگوهای پیچیده در صحنه ترافیک شکست می‌خورند. از این‌رو، مدل‌های مبتنی بر یادگیری عمیق را که بر روی هزاران تصویر ترافیکی مشروح آموزش داده شده بودند، آزمایش و اجرا کردیم؛ بنابراین، سیستم پیشنهادی مطابق شکل ۱ می‌تواند موارد زیر را انجام دهد:



شکل ۱. سیستم مبتنی بر رابط کاربری گرافیکی front-end پیشنهادی با الگوریتم‌ها و پایگاه داده ترافیک پردازش شده در انتهای پستی. برای تجسم نمایش پلتفرم مبتنی بر رابط کاربری گرافیکی پیشنهادی، به (Git Hub, 2020) مراجعه کنید.

۱. نظارت بر تراکم ترافیک
۲. تصادفات رانندگی، شناسایی وسایل نقلیه ثابت یا سرگردان
۳. تشخیص و شمارش وسیله نقلیه
۴. مدیریت ترافیک با استفاده از رابط کاربری گرافیکی مستقل (GUI)
۵. مقیاس نظارت بر ترافیک به چندین دوربین ترافیک

۲- بررسی ادبیات

در چند سال گذشته، چندین سیستم مبتنی بر بینایی برای نظارت خود کار ترافیک مورد مطالعه قرار گرفته‌اند. ما به‌طور گسترده در مورد برخی از مقالات مرتبط با تمرکز بر پیش‌بینی تراکم، تعداد ترافیک و تشخیص ناهنجاری بحث می‌کنیم.

۲-۱- چارچوب‌های یادگیری عمیق برای تشخیص و طبقه‌بندی اشیاء

دو روش اصلی وجود دارد که از طریق آن، سیستم‌های نظارت بر تراکم مبتنی بر ویدیو کار می‌کنند. اولین مورد، روش مبتنی بر «استنتاج سه مرحله‌ای» و دیگری، رویکرد «طبقه‌بندی یک مرحله‌ای» است. ویلیس و همکاران در (Willis et al., 2017)، طبقه‌بندی صفوف ترافیک را با استفاده از شبکه‌های عصبی عمیق بر روی تصاویر ترافیک مورد مطالعه قرار دادند. محققان، یک شبکه دو فازی را با استفاده از GoogLeNet و یک زیرشبکه عمیق سفارشی آموزش دادند و آن را در فرآیند تشخیص تراکم شبکه ترافیک به کار گرفتند. چاکرابورتی و همکاران^۱ (۲۰۱۸) از تصاویر ترافیکی و شبکه‌های عصبی کانولوشنال عمیق (DCNN) و الگوریتم‌های You Only Look Once (YOLO) در تنظیمات محیطی مختلف استفاده کردند. به‌طور مشابه، برای رویکردهای مبتنی بر استنتاج، موریس و همکاران، یک سیستم قابل حمل را برای استخراج پارامترهای صف ترافیک در تقاطع‌های علامت‌دار از فیدهای ویدیویی پیشنهاد نمودند (Morris et al., 2011). برای این کار، آن‌ها تکنیک‌های پردازش تصویر مانند خوشه‌بندی،

1. Chakraborty et al.

تفریق پس‌زمینه و تقسیم‌بندی را جهت شناسایی وسایل نقلیه و درنهایت، جدول‌بندی طول صف برای دوربین‌های کالیبره‌شده در تقاطع‌های مختلف به کار بردند. فولادگر و همکاران^۱ (۲۰۱۷)، یک سیستم غیر متمرکز ساخته‌شده با یادگیری عمیق را پیشنهاد کردند که در آن، هر گره دقیقاً هر یک از حالت‌های تراکم خود را بر اساس ایستگاه‌های مجاور خود در شرایط زمان واقعی پیش‌بینی می‌کند. رویکرد آن‌ها مقیاس‌پذیر بود و می‌توانست برای پیش‌بینی ماهیت جریان ترافیک کاملاً غیر متمرکز باشد. به همین ترتیب، ما و همکاران^۲ (۲۰۱۵)، یک مدل کاملاً خودکار مبتنی بر شبکه عصبی عمیق برای تجزیه و تحلیل داده‌های ترافیک مکانی-زمانی پیشنهاد کردند. مدل آن‌ها ابتدا از شبکه عصبی کانولوشن برای یادگیری ویژگی‌های مکانی-زمانی استفاده می‌کند. سپس، یک شبکه عصبی بازگشتی با استفاده از خروجی مدل مرحله اول آن‌ها آموزش داده می‌شود که به دسته‌بندی توالی کامل کمک می‌کند. این مدل می‌تواند به‌طور عملی در مطالعه جریان‌های ترافیکی و پیش‌بینی تراکم استفاده شود. به‌طور مشابه، وانگ و همکاران^۳ (۲۰۱۶)، یک مدل یادگیری عمیق را معرفی کردند که از ساختار RCNN برای پیش‌بینی مداوم سرعت ترافیک استفاده می‌کند. آن‌ها با استفاده از مدل خود و ادغام اطلاعات ترافیکی مکانی-زمانی می‌توانند منابع ازدحام در جاده‌های کمربندی شهری را شناسایی کنند. کارلی و همکاران^۴ (۲۰۱۵)، یک تحلیل خودکار ازدحام ترافیک در خیابان‌های شهری را پیشنهاد نمودند. آن‌ها از داده‌های تولیدشده توسط GPS برای تعمیم ویژگی‌های ترافیک استفاده کردند. به همین ترتیب، در این مقاله، نویسندگان استفاده از یک سیستم نظارت بر تراکم مبتنی بر ویدیو را نشان داده‌اند که ممکن است به اندازه تکنیک مبتنی بر GPS دقیق نباشد؛ اما قابل اتکا است و هزینه‌های عملیاتی کمتری را به همراه دارد. علاوه بر این، از آنجاکه ازدحام اغلب در جاده‌های شهری رخ می‌دهد، شناسایی شاخص‌های مختلف برای برنامه‌ریزی مؤثر سیستم‌های حمل‌ونقل، مفید خواهد بود (Litman, 2007).

۲-۲- سیستم‌های تحلیل ترافیک مبتنی بر چشم‌انداز

اکثر روش‌های شمارش موجود را می‌توان به‌طور کلی به‌عنوان شمارنده نمونه تشخیص یا تخمین گر چگالی طبقه‌بندی کرد. شمارنده‌های نمونه تشخیص هر خودرو را منحصراً بومی‌سازی نموده و سپس، محلی‌سازی را شمارش می‌کنند. با این حال، این امر می‌تواند مشکلی را ایجاد کند؛ زیرا این فرآیند نیاز به بررسی دقیق پیکسل به پیکسل کل تصویر برای ایجاد محلی‌سازی دارد. به‌طور مشابه، انسداد می‌تواند مانع دیگری ایجاد کند؛ زیرا آشکارسازها ممکن است اشیاء هم‌پوشانی را ادغام کنند. در مقابل، تخمین گرهای چگالی به شیوه‌ای غریزی کار می‌کنند و تلاش می‌کنند تقریبی از چگالی را برای وسایل نقلیه قابل شمارش ایجاد کنند و سپس، آن‌ها را در آن ناحیه متراکم جذب نمایند. تخمین گرهای چگالی معمولاً به مقادیر زیادی از نمونه‌های داده آموزشی نیاز ندارند؛ اما عموماً در کاربرد در همان صحنه‌ای که داده‌های آموزشی در آن جمع‌آوری می‌شوند، محدود می‌گردند.

چیو و همکاران، یک ترافیک خودکار ارائه کردند؛ سیستم مانیتورینگ IC که یک الگوریتم تقسیم‌بندی اشیاء را پیاده‌سازی می‌کند که قادر به تشخیص، ردیابی و تشخیص وسیله نقلیه از تصاویر ترافیک است. رویکرد آن‌ها وسایل نقلیه متحرک را از وسایل ثابت با استفاده از تکنیک قطعه‌بندی شیء متحرک که از ویژگی‌های هندسی وسایل نقلیه برای طبقه‌بندی نوع خودرو استفاده می‌کند، جدا نمود. به همین ترتیب، ژوانگ و همکاران، یک روش آماری پیشنهاد کردند که یک تخمین مبتنی بر همبستگی را برای شمارش وسایل نقلیه شهری با استفاده از دوربین‌های ترافیکی انجام می‌دهد. برای این کار، آن‌ها دو تکنیک را معرفی کردند؛ اولی با استفاده از رویکرد یادگیری ماشین آماری مبتنی بر

1. Fouladgar et al.

2. Ma et al.

3. Wang et al.

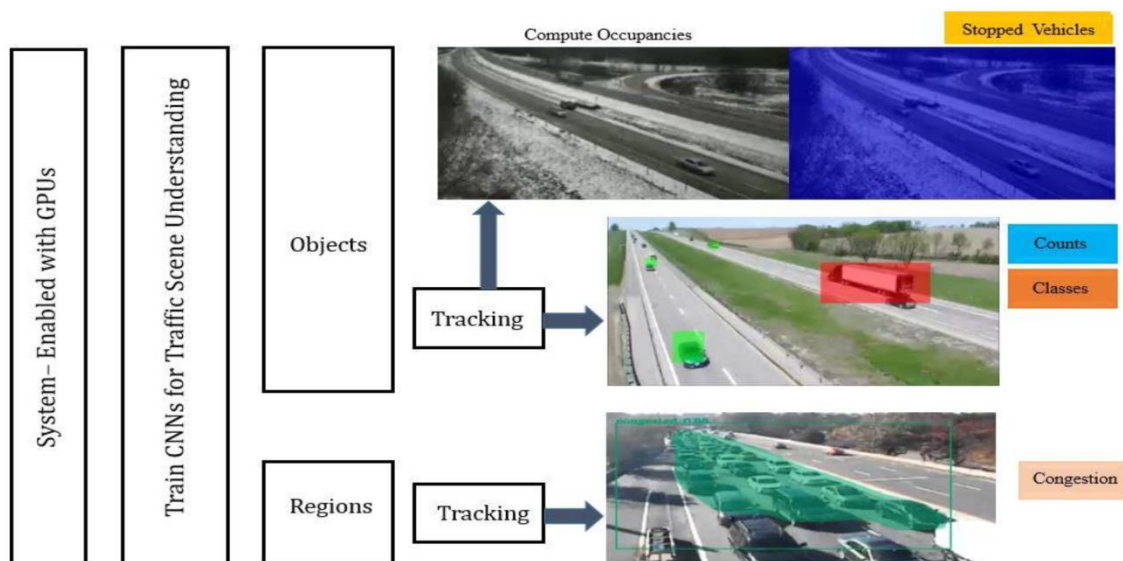
4. Carli et al.

مدل‌های گاوسی و دومی با استفاده از رویکرد انحراف تحلیلی بر اساس جفت ماتریس مبدا-مقصد. موندنک و همکاران، مجموعه داده‌ای از اتومبیل‌های بالای سر ایجاد کردند و یک شبکه عصبی عمیق را برای طبقه‌بندی، شناسایی و شمارش تعداد اتومبیل‌ها به کار گرفتند. آن‌ها برای شناسایی و طبقه‌بندی وسایل نقلیه از یک شبکه عصبی به نام ResCeption استفاده کردند. این شبکه، یادگیری باقیمانده را با لایه‌های Inception ادغام می‌کند که می‌تواند تعداد خودروها را در یک نگاه شناسایی و شمارش نماید. رویکرد آن‌ها در به دست آوردن شمارش دقیق وسایل نقلیه در مقایسه با شمارش‌های انجام‌شده با تخمین محلی‌سازی یا تراکم برتر است.

جدا از تشخیص تراکم و شمارش وسایل نقلیه، مقالات مختلفی برای مطالعه سیستم‌های تشخیص ناهنجاری بررسی شده است. کامیجو و همکاران، یک الگوریتم ردیابی وسیله نقلیه بر اساس میدان‌های تصادفی مارکوف مکانی-زمانی به منظور تشخیص تصادفات ترافیکی در تقاطع‌ها ایجاد کردند. مدل ارائه‌شده در مطالعه آن‌ها قادر به ردیابی قدرتمند وسایل نقلیه فردی بود؛ بدون اینکه دقت آن‌ها به شدت تحت تأثیر اثرات انسداد و درهم و برهم قرار گیرند؛ دو ویژگی بسیار رایج در اکثر تقاطع‌های شلوغ که برای اکثر مدل‌ها مشکل ایجاد می‌کند. اگرچه به‌طور سنتی، سنسورهای نقطه‌ای عمدتاً برای تشخیص حادثه استفاده می‌شدند؛ اما دامنه استفاده از آن‌ها برای سیستم‌های تشخیص ناهنجاری‌ها نسبتاً بی‌اهمیت بود؛ بنابراین، رویکردهای مبتنی بر چشم‌انداز به دلیل قابلیت تشخیص رویداد برتر، بسیار و فراتر از آن مورد استفاده قرار گرفته‌اند. اطلاعاتی مانند ترافیک، تخلفات رانندگی، تصادفات و غیره را می‌توان به راحتی از سیستم‌های مبتنی بر دید استخراج کرد. روجاس و همکاران و زنگ و همکاران، تکنیک‌هایی را برای تشخیص وسایل نقلیه در بزرگراه با استفاده از دوربین مداربسته ثابت ارائه کردند؛ در حالی که Ai و همکاران، روشی را برای تشخیص تخلفات ترافیکی در تقاطع‌ها پیشنهاد نمودند.

۳- روش پیشنهادی

متدولوژی اتخاذشده برای پیاده‌سازی یک سیستم نظارت خودکار ترافیک در شکل ۲ نشان داده شده است. اجزاء اصلی، ابتدا شامل یک باطن فعال با GPU (در فرض) است که برای اطمینان از اینکه مدل‌های بسیار عمیق را می‌توان به سرعت آموزش داد و در یک سیستم پیاده‌سازی کرد، طراحی شده است. مجموعه گسترده‌ای از دوربین‌ها در زمان واقعی در قلب سیستم نظارت بر ترافیک پیشنهادی مبتنی بر هوش مصنوعی، توسعه و آموزش چندین مدل شبکه عصبی کانولوشنال عمیق است که قادر به شناسایی و طبقه‌بندی اشیاء مختلف یا تقسیم‌بندی صحنه ترافیک به اشیاء تشکیل‌دهنده آن هستند. تصاویر ترافیکی با حاشیه‌نویسی دستی به عنوان منبع اصلی مجموعه داده مورد استفاده برای آموزش این مدل‌ها استفاده می‌شود. برای فعال‌سازی سیستم برای آگاهی از موقعیت، الگوریتم‌های ردیابی شیء مختلف برای ایجاد مسیر برای هر شیء شناسایی شده در صحنه ترافیک در همه زمان‌ها پیاده‌سازی می‌شوند. سپس، مراحل قبل برای استخراج متغیرهای مختلف جریان ترافیک (به عنوان مثال، حجم ترافیک و اشغال) و نظارت بر شرایط مختلف ترافیک مانند صف‌بندی، تصادفات و سایر ناهنجاری‌های صحنه ترافیک ترکیب می‌شوند. سیستم نظارت بر ترافیک با هوش مصنوعی قادر به ردیابی کلاس‌های مختلف وسایل نقلیه، جدول‌بندی تعداد آن‌ها، شناسایی و تشخیص تراکم و ردیابی وسایل نقلیه ساکن در زمان واقعی است.



شکل ۲. نمایش بصری سیستم پیشنهادی مجهز به هوش مصنوعی

در ادامه، برخی از الگوریتم‌های یادگیری عمیق مورد استفاده در مطالعه توضیح داده شده‌اند.

۳-۱- R-CNN سریع‌تر

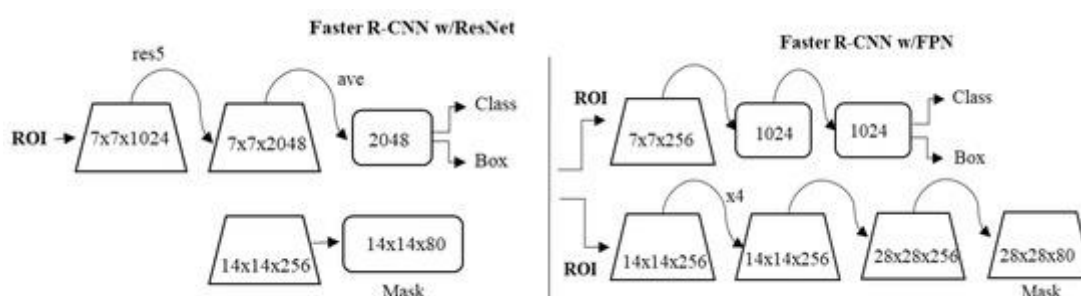
R-CNN سریع‌تر، یک الگوریتم تشخیص هدف دومرحله‌ای است. در Faster-RCNN، یک شبکه پیشنهادی منطقه (RPN)، ویژگی‌های کانولوشنال تصویر کامل را به همراه یک شبکه تشخیص که پیشنهادها را بدون هزینه را امکان‌پذیر می‌کند، به اشتراک می‌گذارد. در اینجا، RPN به‌طور هم‌زمان محدود شده و مقادیر امتیاز معادل آن‌ها را در هر موقعیت پیش‌بینی می‌کند. آموزش انتها به پایان RPN، پیشنهادها را منطقه‌ای با کلاس بالا را ارائه می‌دهد که توسط Faster R-CNN برای دستیابی به پیش‌بینی اشیاء استفاده می‌شود. در مقایسه با R-CNN، Fast R-CNN، سریع‌تر با جایگزینی روش جست‌وجوی انتخابی با RPN، تشخیص شیء با کیفیت بالا را تولید می‌کند. این الگوریتم، هر تصویر را به بخش‌های متعددی از مناطق فشرده تقسیم می‌کند و سپس، هر ناحیه را از روی آرایش فیلترهای کانولوشنال عبور می‌دهد تا توصیف‌گرهای ویژگی با کیفیت بالا را استخراج کند که بعد از آن از طریق یک طبقه‌بندی‌کننده عبور داده می‌شود. پس از آن، طبقه‌بندی‌کننده احتمال اشیاء را در هر بخش از یک تصویر تولید می‌کند. برای دستیابی به دقت پیش‌بینی بالاتر در فیدهای دوربین ترافیک، این مدل برای پنج کلاس عابر پیاده، دوچرخه‌سوار، اتوبوس، کامیون و ماشین آموزش داده شده است. آموزش در GPU NVIDIA GTX 1080Ti تقریباً ۸ ساعت طول کشید. این مدل، فیدهای ویدیویی را با سرعت ۵ فریم در ثانیه پردازش می‌کرد.

۳-۲- ماسک R-CNN

Mask R-CNN اختصار Mask-region Based Convolutional Neural Network، توسعه R-CNN سریع‌تر است (He et al., 2017). علاوه بر انجام وظایفی معادل R-CNN سریع‌تر، Mask R-CNN آن را با افزودن ماسک‌های برتر تکمیل نموده و منطقه موردنظر را پیکسل به پیکسل تقسیم می‌کند. مدل مورد استفاده در این مطالعه مبتنی بر شبکه هرمی ویژگی (FPN) بوده و با ستون فقرات resnet101 اجرا شده است. ResNet101 به‌عنوان استخراج‌کننده ویژگی برای مدل عمل می‌کند. هنگام استفاده از FPN، با معرفی هرم دیگری که ویژگی‌های سطح بالاتری را از هرم اول گرفت و در نتیجه، آن‌ها را به لایه‌های فرعی منتقل کرد، در هرم استاندارد استخراج ویژگی

بهبود یافت. این ویژگی‌ها را در هر سطح فعال می‌کند تا در شخصیت‌های سطح بالاتر و پایین‌تر پذیرش بگیرند. در این مطالعه، حداقل نرخ اطمینان تشخیص ۹۰ درصد تعیین شد و اعتبارسنجی در ۵۰ مرحله اجرا شد. یک رویکرد آموزشی تصویرمحور دنبال شد که در آن، هر تصویر به شکل مربع برش داده شد.

تصاویر از $1024 \times 1024 \times 3$ پیکسل (RGB) به یک نقشه ویژگی به شکل $32 \times 32 \times 2048$ در عبور از شبکه ستون فقرات تبدیل شدند. هر دسته دارای یک تصویر واحد در هر GPU بود و در مجموع، هر تصویر دارای ۲۰۰ منطقه مورد علاقه آموزش دیده (ROI) بود. با استفاده از نرخ یادگیری 0.001 و اندازه دسته‌ای ۱، این مدل بر روی GPU NVIDIA GTX 1080Ti آموزش داده شد. نرخ یادگیری ثابت در طول تکرار استفاده شد. به همین ترتیب، کاهش وزن 0.001 و حرکت یادگیری 0.9 استفاده شد. کل زمان آموزش برای آموزش مدل روی مجموعه داده نمونه تقریباً ۳ ساعت بود. در شکل ۳، چارچوب Mask-RCNN نشان داده شده است.



شکل ۳. چارچوب شبکه عصبی کانولوشنال مبتنی بر منطقه ماسک (Mask R-CNN)

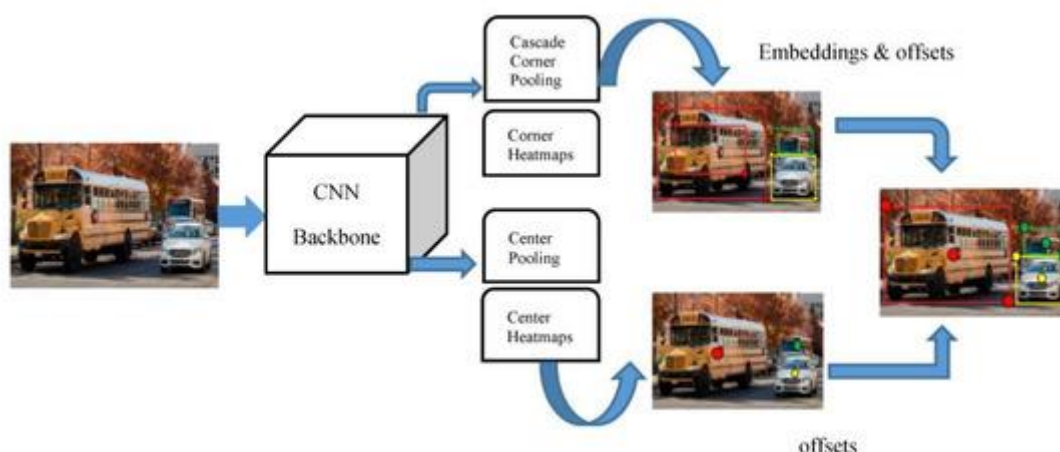
۳-۳-۳- یولو

You Only Look Once (YOLO) پیشرفته‌ترین الگوریتم تشخیص اشیاء است. برخلاف سیستم‌های سنتی تشخیص اشیاء، YOLO تنها یک بار تصویر را بررسی می‌کند و تشخیص می‌دهد که آیا اشیاء در آن وجود دارد یا خیر؟ در این مطالعه، YOLOv4 برای انجام تشخیص وسیله نقلیه، شمارش و مقایسه نتایج برای تولید صف‌های ترافیکی استفاده شد. بیشتر الگوریتم‌های تشخیص اشیاء معاصر، طبقه‌بندی‌کننده‌های CNN را با هدف انجام تشخیص‌ها تغییر کاربری می‌دهند. به عنوان مثال، برای انجام تشخیص شیء، این الگوریتم‌ها از یک طبقه‌بندی‌کننده برای آن شیء استفاده می‌کنند و آن را در مکان‌ها و مقیاس‌های مختلف در تصویر آزمایشی آزمایش می‌کنند. با این حال، YOLO تشخیص شیء را دوباره قاب می‌کند؛ یعنی به جای اینکه هزار بار به یک تصویر نگاه کند تا تشخیص دهد، فقط یک بار به تصویر نگاه می‌کند و پیش‌بینی دقیق شیء را انجام می‌دهد. یک شبکه سی‌ان‌ان به شکل هم‌زمان باکس‌های مرزی متعدد و احتمالات کلاس را برای آن جعبه‌های تولید شده پیش‌بینی می‌کند. برای ساخت مدل‌های YOLO، زمان معمول تقریباً ۲۰-۳۰ ساعت بود. YOLO از منابع سخت‌افزاری مشابه Mask R-CNN برای آموزش استفاده کرد.

۳-۴- CenterNet

CenterNet 16، الگوهای بصری را در هر بخش از یک تصویر برش‌خورده با هزینه‌های محاسباتی کمتر کشف می‌کند. به جای تشخیص اشیاء به عنوان یک جفت نقطه کلیدی، CenterNet آن‌ها را به عنوان یک سه‌گانه تشخیص داده و بدین ترتیب، دقت و مقادیر فراخوان را افزایش می‌دهد. این چارچوب بر روی اشکالاتی است که CornerNet 37 با آن مواجه شده و از یک جفت نقطه کلید گوشه برای انجام تشخیص اشیاء استفاده می‌کند. با این حال،

CornerNet در ساختن یک چشم‌انداز جهانی تر از یک شیء، CenterNet با داشتن یک نقطه کلید اضافی برای به دست آوردن اطلاعات مرکزی تر از یک تصویر، این کار را انجام می‌دهد. CenterNet بر اساس شهود عمل می‌کند که اگر یک جعبه مرزی شناسایی شده دارای یک تقاطع بالاتر از اتحادیه (IOU) با جعبه حقیقت زمین باشد، احتمال اینکه آن نقطه کلید مرکزی در منطقه مرکزی خود باشد و در همان کلاس برچسب گذاری شود، زیاد است. از این رو، دانش داشتن یک سه‌قلو به جای یک جفت، برتری CenterNet را نسبت به CornerNet یا هر رویکرد تشخیص مبتنی بر لنگر دیگر افزایش می‌دهد. با وجود استفاده از یک سه‌گانه، CenterNet هنوز یک آشکارساز تک‌مرحله‌ای است؛ اما تا حدی قابلیت‌های RoI Pooling را دریافت می‌کند. شکل ۴، معماری CenterNet را نشان می‌دهد که در آن از یک ستون فقرات CNN استفاده می‌کند و ادغام گوشه‌های آبخاری و ادغام مرکزی را انجام می‌دهد تا نقشه حرارتی دو گوشه و نقطه کلید مرکزی را ارائه دهد. در اینجا، ادغام گوشه‌های آبخاری، ماژول ادغام گوشه اصلی را قادر می‌سازد تا اطلاعات داخلی را دریافت کند؛ در حالی که ادغام مرکزی به نقاط کلیدی در مرکز کمک می‌کند تا به الگوی بصری قابل شناسایی بیشتری در داخل اشیاء دست یابند که آن را قادر می‌سازد قسمت مرکزی منطقه را درک کند. به همین ترتیب، مشابه CornerNet، از یک جفت گوشه شناسایی شده و جاسازی‌های آشنا برای پیش‌بینی یک جعبه محدود استفاده می‌شود. سپس، کادرهای مرزبندی نهایی با استفاده از نقاط کلیدی مرکز شناسایی شده تعیین می‌شوند.



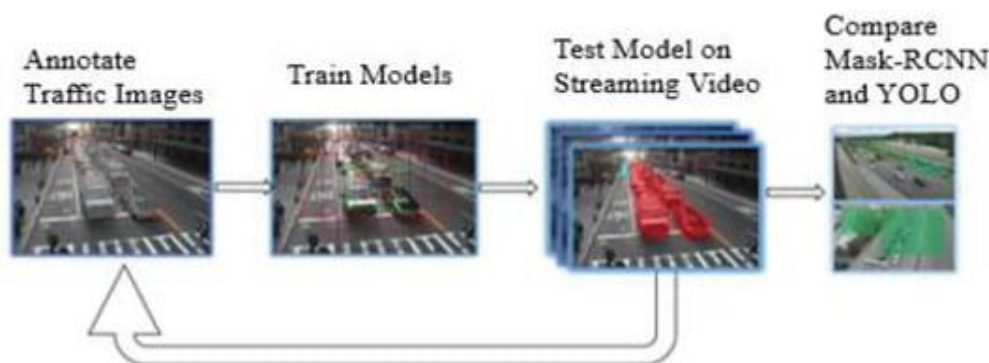
شکل ۴. معماری CenterNet

بخش‌های زیر، شرح مختصری از چندین عملیات ترافیکی را ارائه می‌دهد که می‌توانند به‌طور یکپارچه خودکار شوند.

۳-۵- نظارت بر صف‌های ترافیکی

در شکل ۵، روش اتخاذ شده برای یک سیستم نظارت خودکار صف نشان داده شده است. در اولین مرحله، انجام حاشیه‌نویسی با استفاده از VGG Image Annotator به دست آمد. در ادامه، از تصاویر حاشیه‌نویسی برای آموزش هر دو مدل Mask R-CNN و YOLO استفاده شد. زمان تمرین برای Mask R-CNN و YOLO، به ترتیب تقریباً ۳،۵ و ۲۲ ساعت بود. پس از انجام آموزش، این مدل‌ها بر روی ویدیوهای ترافیکی بلادرنگ اجرا شدند تا عملکرد آن‌ها ارزیابی گردد. دلیل اصلی استفاده از Mask R-CNN به دلیل توانایی آن در به دست آوردن ماسک‌های تقسیم‌بندی سطح پیکسل بود که تشخیص صف را دقیق می‌کرد. از آنجاکه YOLO از یک جعبه محدود کننده برای

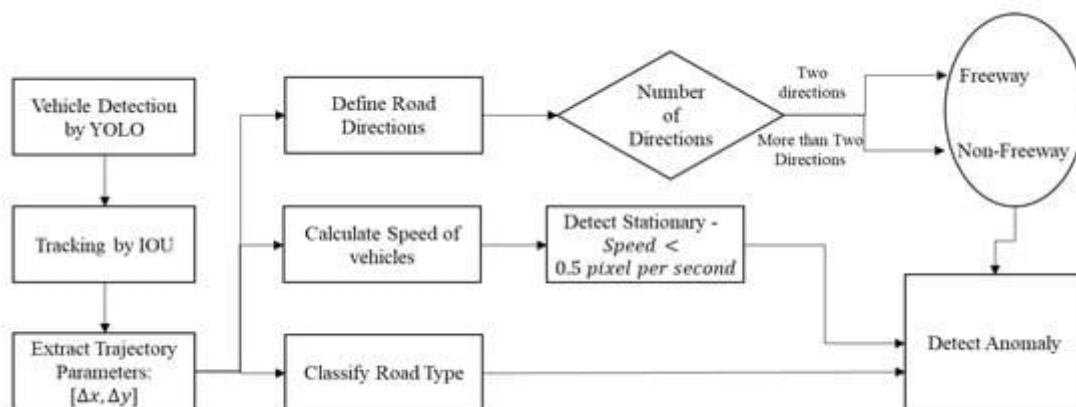
انجام تشخیص استفاده می‌کند، مناطقی را پوشش می‌دهد که هم شلوغ و هم غیر متراکم هستند؛ بنابراین، Mask-RCNN نسبت به YOLO در پیش‌بینی دقیق مناطق طبقه‌بندی شده مورد علاقه مزیت دارد.



شکل ۵. فلوچارت عملیات گام‌به‌گام

۳-۶- تشخیص وسایل نقلیه ساکن

شکل ۶، روش پیشنهادی برای تشخیص وسایل نقلیه ساکن یا سرگردان را نشان می‌دهد. برای شروع فرآیند، یک مدل YOLO برای انجام تشخیص وسیله نقلیه آموزش داده می‌شود. سپس، ردیابی‌ها با استفاده از یک فرآیند تقاطع روی اتحاد (IOU) ردیابی می‌شوند و هر مسیر وسیله نقلیه از صحنه‌های ترافیک ترسیم می‌شود. در ادامه از نتایج ردیابی برای ترسیم جهت‌های خاص سفر (شرق، غرب، شمال یا جنوب)، نوع جاده مورد تجزیه و تحلیل (تقاطع یا آزادراه) و سرعت پیش‌بینی شده وسایل نقلیه ردیابی شده استفاده می‌گردد. نتایج ردیابی، بعداً برای بیان مسیرهای گسسته سفر، نوع جاده و سرعت تخمینی وسیله نقلیه استفاده می‌شود. برای انواع خاصی از جاده‌ها، اگر سرعت وسیله نقلیه برای مدت زمان معینی زیر یک آستانه خاص قرار گیرد، مدل می‌تواند ثابت کند وسیله نقلیه را تشخیص دهد.



شکل ۶. فلوچارت برای تشخیص ناهنجاری ترافیک

۴- شرح داده‌ها

تصاویر دوربین‌های ترافیکی به عنوان منبع اصلی مجموعه داده مورد استفاده در این مطالعه استفاده می‌شود. تصاویر از Iowa DOT Open Data، RTIS، New York State DOT، Iowa 511 و وزارت حمل و نقل و توسعه لوئیزیانا به دست آمده است. در مجموع، ۱۸۵۰۹ تصویر برای اهداف آموزشی و اعتبارسنجی استفاده شد. مجموعه داده‌ها شامل تصاویری است که در زمان‌های مختلف روز در شرایط محیطی مختلف گرفته شده است. تصاویر تقاطع، آزادراه و

منطقه کاری در هر دو نمونه آموزشی و آزمایشی گنجانده شد. این تصاویر برای آموزش و اعتبارسنجی مدل‌های یادگیری عمیق به منظور انجام فرآیندهای تشخیص تراکم، ردیابی خودروی ثابت و شمارش وسایل نقلیه مورد استفاده قرار گرفتند. برای تشخیص ناهنجاری، از ویدیوهای ترافیکی NVIDIA AI City Challenge جهت آزمایش اثربخشی مدل پیشنهادی استفاده شد. در نهایت، مدل بر روی ۱۰۰ فید ویدیویی دوربین مداربسته با انواع مختلف ناهنجاری در ترافیک نامنظم و الگوهای آب‌وهوا مورد ارزیابی قرار گرفت.

۵- نتایج

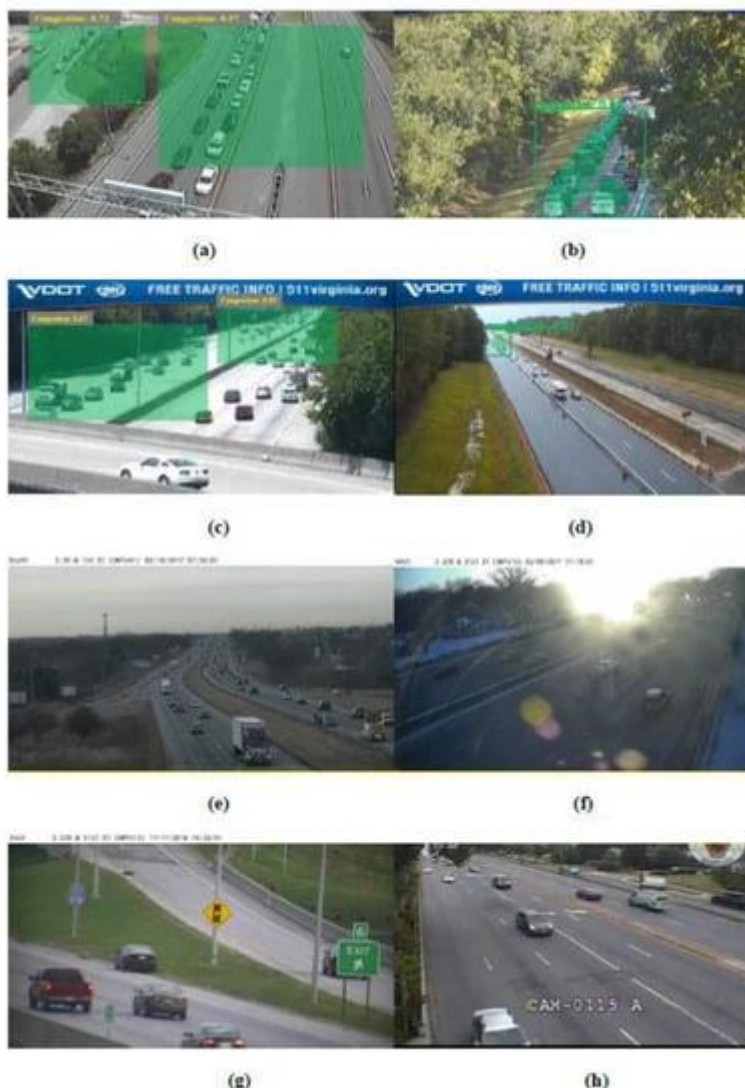
در این بخش، عملکرد صفوف ترافیکی، سیستم تشخیص ناهنجاری و شمارش خودکار وسایل نقلیه را ارزیابی می‌کنیم.

۵-۱- تشخیص صف‌های ترافیکی

عملکرد Mask R-CNN بر روی ۱۰۰۰ تصویر دوربین ترافیک (۵۰۰ تصویر شلوغ و ۵۰۰ بدون تراکم)، آزمایش و یک تحلیل مقایسه‌ای با چارچوب کلاسیک YOLO انجام شد. معیارهای عملکرد استاندارد دقت، یادآوری و دقت، همان‌طور که به ترتیب در معادلات (۱)، (۲) و (۳) نشان داده شده است، برای آزمایش مدل‌ها استفاده شد. سپس، نتایج اجرای بلادرنگ Mask R-CNN در یک تقاطع نشان داده شد. دلیل آن می‌تواند دور بودن دسته وسایل نقلیه از دوربین (شکل e7) و مشکلات آشکار (شکل f7) باشد. شکل g,h7 به درستی تصاویر بدون تراکم را به عنوان نگاتیو واقعی (TN) پیش‌بینی کرد. جدول ۱، مقادیر Precision، Recall و Accuracy را نشان می‌دهد که برای مدل‌های Mask R-CNN و YOLO در پیش‌بینی صحیح صف‌های ترافیک به دست آمده است.

جدول ۱. عملکرد Mask R-CNN در مقابل YOLO در تشخیص صف

مدل	دقت، درستی	حافظه	دقت
Mask R-CNN	۹۲/۸	۹۵/۶	۹۰/۵
YOLO	۹۵/۵	۹۴/۸	۹۳/۷



شکل ۷. طبقه‌بندی صف‌های پیش‌بینی‌شده: مثبت واقعی-(a,b)، مثبت کاذب-(c,d)، منفی کاذب-(e,f)، منفی واقعی-(g,h) توسط You Only Look Once (YOLO) و ماسک R-CNN

مطالعه موردی برای مطالعه صف‌های ترافیک

یک مطالعه موردی انجام شد که در آن، مدل Mask R-CNN در زمان واقعی برای نظارت بر صف‌ها در یک تقاطع پیاده‌سازی شد. توجه به این نکته ضروری است که تغییرات در چشم‌انداز دوربین فیلم‌برداری، اغلب استخراج پارامترهای صف ترافیک از صحنه‌های فریم را به چالش می‌کشد. یک دوره معمولی در این مورد تنظیم دوربین بر روی یک ارتفاع خاص، زاویه مشاهده، سطح زوم و غیره بود. اگرچه این ممکن است مؤثر باشد؛ اما مقیاس‌پذیر نیست. جایگزین دیگر این رویکرد می‌تواند استفاده مستقیم از مقادیر پیکسل تصویر به منظور مشخص کردن پارامترهای صف باشد. در حین استفاده از این روش، اطلاعات صف از یک نقطه را نمی‌توان با مکان دیگری مقایسه کرد؛ زیرا ممکن است ویژگی‌های هندسی دوربین متفاوت باشد. در مراحل توضیح داده‌شده در زیر، یک رویکرد ساده و استاندارد شده بدون کالیبراسیون برای استخراج پارامترهای طول صف از فیدهای ویدیوی ترافیک ایجاد می‌کنیم. این رویکرد مقیاس‌پذیر است و در مقایسه سطوح صف در مکان‌های مختلف مفید است.

مرحله ۱: استخراج مناطق صف از فیدهای ویدیوی ترافیک با Mask RCNN.

- مرحله ۲: طول پیکسل هر ماسک صف شناسایی شده را محاسبه کنید.
- مرحله ۳: طول را در طول زمان جمع کنید (حداقل مدت ۱ هفته است).
- مرحله ۴: از آستانه تطبیقی (شکل ۸) برای قرار دادن طول صف در سطوح مختلف شدت استفاده کنید: کم، متوسط و زیاد.
- مرحله ۵: ایجاد نقشه حرارتی سطوح صف و در نهایت، مقایسه.

Steps shown for Adaptive Thresholding

```

Initialize: L, M, H
Input: PL—pixel lengths
for each location do
  for each [day, hour, minute] in [30 days, 24 h, 60 min] do
    % extract first, second, third quartile pixel lengths
    Q = percentile[PL, {Q1, Q2, Q3}]

  end

  L = Q[{Q1, Q2, Q3}].mean.max + k * Q[{Q1}].std
  M = Q[{Q1, Q2, Q3}].mean.max + k * Q[{Q2}].std
  H = Q[{Q1, Q2, Q3}].mean.max + k * Q[{Q3}].std

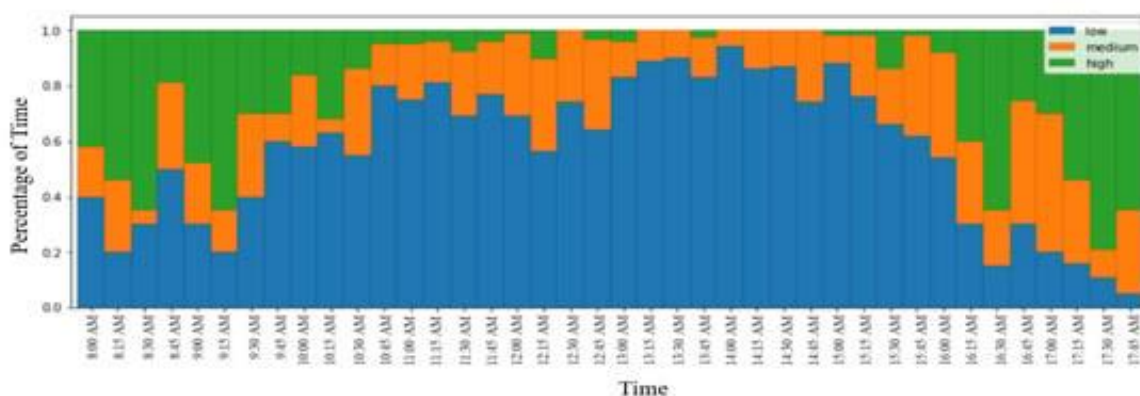
end

Output: L, M, H

```

شکل ۸. مراحل آستانه سازی تطبیقی

چارچوب Mask R-CNN برای تعیین کمیت سطوح صف در یک تقاطع استفاده شد. نقشه حرارتی نشان داده شده در شکل ۹، به وضوح شروع و از بین رفتن صف ها را نشان می دهد. نقشه حرارتی تقاطع می تواند ساعت اوج AM و PM را تشخیص دهد.



شکل ۹. نقشه حرارتی شدت صف ترافیک در یک تقاطع

۵-۲- تشخیص خودروی ثابت

الگوریتم های R-CNN سریع تر و YOLO برای مطالعه وسایل نقلیه ساکن به کار گرفته شدند. برای درک و مقایسه نتایج آزمون برای هر دو چارچوب R-CNN سریع تر و YOLO، از ماتریس های سردرگمی و نمرات F-1 استفاده

شد. ماتریس سردرگمی، سطوح دقت را برای بخش‌های مختلف طبقه‌بندی تصویر نشان می‌دهد. به‌طور کلی، ۲۵ نتیجه آزمون در یک جدول ۵×۵ نشان داده شده است که ماتریس سردرگمی نامیده می‌شود. در اینجا، هر ردیف تعداد واقعی پیش‌بینی‌ها را نشان می‌دهد و تعداد کل هر ردیف، نشان‌دهنده تعداد اهداف پیش‌بینی شده برای آن کلاس است. به همین ترتیب، هر ستون نشان‌دهنده تعداد واقعی اهداف است؛ درحالی‌که تعداد کل هر ستون، نشان‌دهنده تعداد واقعی اهداف برای آن کلاس است. به‌طور مشابه، امتیاز F-1 نشان داده شده در معادله (۴) برای مقایسه عملکرد هر دو مدل R-CNN سریع‌تر و YOLO استفاده می‌شود. در جداول ۲ و ۳، نتایج به‌دست‌آمده برای امتیازات ماتریس سردرگمی و F-1 نشان داده شده است.

different sections of image classification. Overall, 25 test results are shown in a 5 × 5 table that is referred to as a confusion matrix. Here, each row shows the actual number of predictions and total number of each row implies the number of targets predicted for that class. Likewise, every column signifies the true number of targets while the total number of each column represents the actual number of targets for that class. Similarly, the F-1 score shown in Equation (4) is used to compare the performance of both Faster R-CNN and YOLO models. The results obtained for confusion matrix and F-1 scores are shown in Table 2 and Table 3.

$$F-1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Table 2. Confusion Matrix of YOLO and Faster R-CNN.

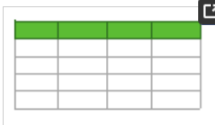
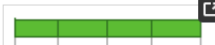


Table 3. F-1 Scores of YOLO and Faster R-CNN.



Back to Top

جدول ۲. ماتریس سردرگمی YOLO و R-CNN سریع‌تر

		YOLO				
True	Pred	Ped	Cyclist	Car	Bus	Truck
Ped		0.9928	0.0053	0.0008	0	0.0008
Cyclist		0.0228	0.9726	0	0	0.0045
Car		0	0	0.9947	0.0001	0.0050
Bus		0	0	0	0.9942	0.0057
Truck		0	0	0.0457	0.0074	0.9467

		Faster R-CNN				
True	Pred	Ped	Cyclist	Car	Bus	Truck
Ped		0.9973	0.0026	0	0	0
Cyclist		0.0401	0.9553	0.0044	0	0
Car		0.0002	0.0001	0.9943	0.0003	0.0047
Bus		0	0	0.0103	0.9792	0.0103
Truck		0	0	0.0367	0.0079	0.9553

Back to Top

جدول ۳. امتیازات F-1 YOLO و R-CNN سریع تر

YOLO			
Class	Precision	Recall	F-1 Score
Ped	0.9216	0.7367	0.8188
Cyclist	0.9424	0.8658	0.9025
Car	0.9276	0.7990	0.8585
Bus	0.9508	0.8571	0.9015
Truck	0.9160	0.8400	0.8764
Total	0.9269	0.7975	0.8573
Faster R-CNN			
Ped	0.8754	0.8838	0.8796
Cyclist	0.9380	0.8514	0.8926
Car	0.8312	0.8788	0.8543
Bus	0.8952	0.8663	0.8805
Truck	0.8928	0.8972	0.8950
Total	0.8417	0.8798	0.8604

مطابق جدول ۲، عملکرد هر دو مدل R-CNN سریع تر و YOLO مشابه بود. R-CNN سریع تر در تشخیص دوچرخه سواران و اتوبوس ها نسبتاً پایین تر عمل کرد؛ اما در مقایسه با عملکرد YOLO در تشخیص کامیون ها بهتر بود. هر دو مدل، خودروها و عابران پیاده را با سطح دقت ۹۹ درصد پیش بینی کردند. از جدول ۳ مشخص شده است که امتیاز تجمعی F-1 YOLO کمتر از امتیاز R-CNN سریع تر است. علاوه بر این، مقدار یادآوری برای YOLO کمتر بود که نشان می دهد در مقایسه با R-CNN سریع تر، اشیاء کمتری را در صحنه ترافیک تشخیص می دهد. پس از مقایسه نتایج در جدول ۳، به نظر می رسد Faster-RCNN کمی بهتر است؛ اما قابل مقایسه با YOLO است؛ بنابراین، هر یک از آنها می تواند به عنوان یک آشکارساز شیء مورد استفاده قرار گیرد.

به طور مشابه، پس از اینکه آشکارساز شیء، موقعیت هر وسیله نقلیه را در یک صحنه ترافیک تشخیص داد، ردیاب برای ردیابی وضعیت وسایل نقلیه از دنباله ای از فریم های ویدیویی ترافیک وارد می شود. تقاطع بیش از اتحادیه (IOU) و سیستم های ردیابی مبتنی بر ویژگی مستقر گشته که به شرح زیر هستند:

تشخیص ردیابی توسط IOU و Feature Tracker

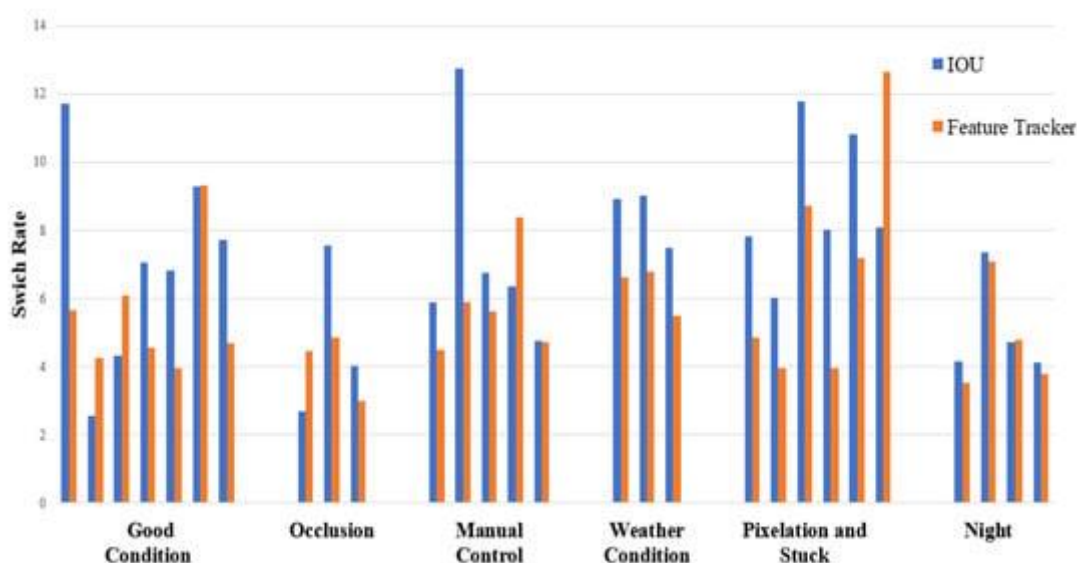
سیستم های تشخیص ناهنجاری، نه تنها به یک آشکارساز برای تشخیص صحیح وسایل نقلیه در قاب ها بلکه به یک ردیاب نیز برای تشخیص وضعیت و حرکت هر وسیله نقلیه نیاز دارند. پس از اینکه آشکارساز، موقعیت وسیله نقلیه را در هر فریم پیش بینی کرد، ردیاب، مسئول ردیابی مسیر وسیله نقلیه بر اساس یک سری فریم های متوالی در یک فایل ویدیویی است. پس از محاسبه هم پوشانی فضایی جعبه های تشخیص اشیاء در هر فریم ویدیویی متوالی، یک IOU تشخیص ها را اختصاص می دهد. IOU اریک و همکاران در این مطالعه اجرا شد. از آنجا که ردیاب های IOU هزینه محاسباتی پایین تری دارند، دستیابی به مسیرهای وسایل نقلیه و ادغام با سایر ردیاب های سطح بالاتر بدون تأثیر بر سرعت محاسبات آسان است. با IOU می توان به نرخ فریم حتی تا ۵۰۰۰۰ فریم در ثانیه دست یافت. توجه به این نکته ضروری است که ردیاب IOU به شدت به چگونگی پیش بینی دقیق توسط مدل های تشخیص شیء وابسته است. نوع

جاده بر اساس تعداد جهت‌های خیابان شناسایی شده طبقه‌بندی می‌شود. برای بیش از دو جهت شناسایی شده، نوع جاده به‌عنوان یک تقاطع یا یک تقاطع طبقه‌بندی می‌شود. به همین ترتیب، دقیقاً برای دو جهت شناسایی شده، جاده به‌عنوان یک آزادراه یا به‌سادگی، یک خیابان دو خطه طبقه‌بندی می‌شود. در شکل ۱۰، تصویر اول، یک آزادراه و تصویر دوم، یک تقاطع طبقه‌بندی شده است.



شکل ۱۰. ردیابی وسیله نقلیه و طبقه‌بندی نوع جاده

در ردیابی شیء مبتنی بر ویژگی، اطلاعات ظاهری برای ردیابی اشیاء در صحنه ترافیک استفاده می‌شود. این روش در ردیابی وسایل نقلیه در محیطی که انسداد مکرر رخ می‌دهد، مفید است. این سیستم، ویژگی‌های شیء را از یک فریم استخراج نموده و سپس، اطلاعات ظاهری را با فریم‌های بعدی بر اساس سطح شباهت مطابقت می‌دهد. حداقل مقدار فاصله کسینوس برای محاسبه هر شباهت بین برخی از ویژگی‌های مشخصه مناسب بوده که برای ردیابی وسیله نقلیه راحت است. علاوه بر این، نتایج بین IOU و Feature Tracker بر اساس میانگین نرخ سوئیچ برای شرایط مختلف محیطی و کیفیت ویدئو مقایسه می‌شوند. نرخ سوئیچ اندازه‌گیری می‌کند که هرچند وقت یک بار به یک وسیله نقلیه، یک شماره مسیر جدید در هنگام عبور از یک صحنه ترافیک اختصاص داده شود. به عبارت ساده، نسبت سوئیچ وسیله نقلیه به تعداد واقعی وسایل نقلیه است.



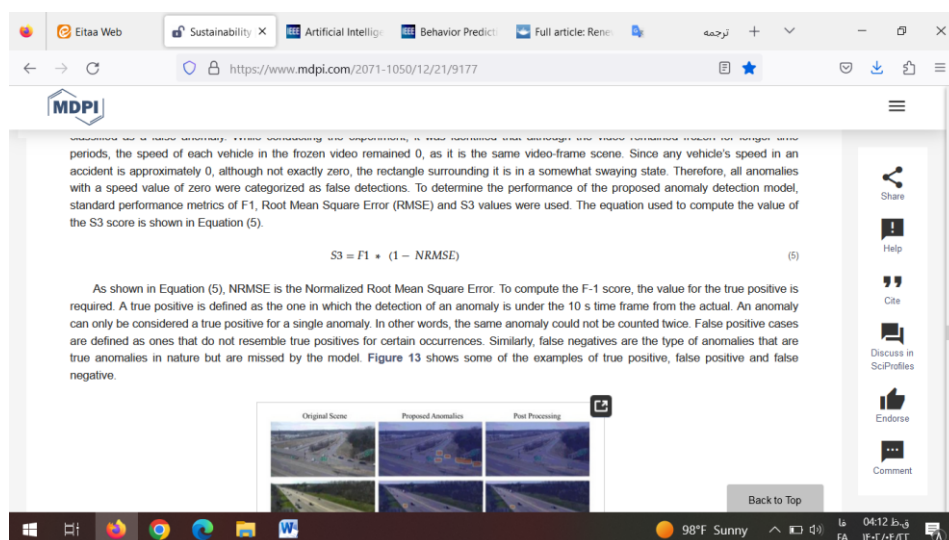
شکل ۱۱. مقایسه نمودارهای خوشه‌ای برای IOU و Feature Tracker



شکل ۱۲. ناهنجاری‌های ترافیکی

سیستم تشخیص ناهنجاری ترافیک پیشنهادی بر روی ۱۰۰ فید ویدیوی ترافیکی با الگوهای ترافیک و آب‌وهوای مختلف ارزیابی شد. وجود فریم‌های منجمد و جلوه‌های پیکسل‌سازی در مجموعه داده‌های ویدیوی ارزیابی، چالش بزرگی را در تشخیص ناهنجاری‌ها ایجاد کرد. ردیاب IOU مورد استفاده در این مطالعه، یک وسیله نقلیه منفرد یا یک جوخه از وسایل نقلیه را یک ناهنجاری احتمالی تصور کرد؛ حتی اگر علامت توقف ترافیک به آن‌ها دستور توقف دهد. اگرچه این وضعیت نمی‌تواند به عنوان یک ناهنجاری طبقه‌بندی شود، IOU آن‌ها را به عنوان چنین موردی برچسب‌گذاری کرد؛ بنابراین، برای تعمیرات اساسی این موضوع، تعیین اینکه یک جاده یک تقاطع یا آزادراه است، مهم است. بر اساس نوع جاده، توقف هر وسیله نقلیه برای بیش از ۳۰ ثانیه در آزادراه به عنوان یک ناهنجاری در نظر گرفته شد و برای یک تقاطع، محدودیت زمانی ۶۰ ثانیه تعیین گردید. به طور مشابه، صلاحیت ردیاب IOU توسط فایل‌های ویدیویی که برای دوره‌های زمانی معینی گیر کرده بودند، بیشتر به چالش کشیده شد. ویدیوها اغلب بیش از

یک دقیقه گیر می‌مانند. در چنین مواردی، ردیاب IOU هر بار که ویدیو برای مدت زمان طولانی تری فریز می‌شد، وسیله نقلیه را به عنوان یک ناهنجاری احتمالی تشخیص می‌داد.



همان‌طور که در رابطه (۵) نشان داده شده است، NRMSE خطای میانگین مربع ریشه عادی شده است. برای محاسبه امتیاز F-1، مقدار مثبت واقعی مورد نیاز است. یک مثبت واقعی به عنوان مثبتی تعریف می‌شود که در آن، تشخیص یک ناهنجاری در بازه زمانی ۱۰ ثانیه از زمان واقعی باشد. یک ناهنجاری را فقط برای یک ناهنجاری می‌توان یک مثبت واقعی در نظر گرفت. به عبارت دیگر، همان ناهنجاری را نمی‌توان دو بار شمارش کرد. موارد مثبت کاذب به مواردی گفته می‌شود که برای برخی رویدادها شبیه مثبت واقعی نیستند. به طور مشابه، منفی‌های کاذب نوعی از ناهنجاری‌ها هستند که در طبیعت ناهنجاری‌های واقعی هستند؛ اما توسط مدل نادیده گرفته می‌شوند. شکل ۱۳، برخی از نمونه‌های مثبت واقعی، مثبت کاذب و منفی کاذب را نشان می‌دهد.



شکل ۱۳. طبقه‌بندی ناهنجاری‌های پیش‌بینی‌شده - ردیف اول، (الف): مثبت واقعی. ردیف دوم، (ب): منفی‌های کاذب - ناهنجاری‌هایی که با دایره‌های قرمز نشان داده شده‌اند. ردیف سوم، (ج): مثبت کاذب

خطاها در تشخیص ناهنجاری با ریشه میانگین مربعات خطا (RMSE) نشان داده می‌شوند. مقدار RMSE برای زمان‌های ناهنجاری حقیقت زمینی و زمان‌های ناهنجاری پیش‌بینی شده برای هر تشخیص مثبت واقعی محاسبه می‌شود. مقدار S3 با استفاده از RMSE محاسبه گردیده که با NRMSE نرمال می‌شود. همان‌طور که از رابطه (۵) مشاهده می‌شود، نرمال‌سازی با استفاده از تکنیک نرمال‌سازی حداقل تا حداکثر با بیشترین و کمترین مقادیر به ترتیب ۳۰۰ و ۰ انجام می‌شود. از جدول ۴، امتیاز $F1\ 0.8333$ محاسبه شده است؛ یعنی آشکارساز نزدیک به $83/3\%$ درصد از کل ناهنجاری‌ها را پیش‌بینی می‌کند. با این حال، به دلیل کاستی‌های موجود در مجموعه داده، به‌ویژه برای خودروهایی که دور از دوربین قرار دارند، این مدل نتوانست ناهنجاری‌ها را در آن موقعیت‌ها تشخیص دهد.

اهمیت الگوریتم‌های تشخیص ناهنجاری، کاربرد مورد استفاده را نه تنها از تشخیص حوادث ترافیکی در زمان واقعی فراتر می‌برد بلکه قادر به اندازه‌گیری و محاسبه درست و دقیق مدت زمان آن‌ها و اثرات ثانویه چنین حوادثی است؛ خواه تشکیل صف و یا امکان حوادث پایین‌دست ثانویه صف تشکیل. جای تعجب نیست که حوادث ترافیکی، یک‌چهارم کل ازدحام جاده‌ها در ایالات متحده را تشکیل می‌دهند. میانگین زمان پاک‌سازی برای حوادث گزارش شده از طریق برنامه 42 HELP بین ۴۲ دقیقه تا ۵۰ دقیقه بود. رویکرد معمول برای اندازه‌گیری تأثیر حوادث ترافیکی از نمودارهای صف قطعی همراه با بررسی تغییر ظرفیت شبکه استفاده می‌کند. یک چالش در دستیابی به مدیریت مؤثر حادثه ناشی از فقدان داده‌های دقیق است که تأثیر حوادث را با در نظر گرفتن ویژگی‌های مکانی و زمانی منحصربه‌فرد آن‌ها کمیت می‌کند. پاسخ مدیریت حوادث ترافیکی می‌تواند از بینش‌های ارزشمند برون‌یابی شده از داده‌های مشتق شده از موقعیت‌های حادثه شناسایی شده و همچنین، اثرات اقدامات متقابل اعمال شده، به‌منظور بهبود استقرار پاسخ‌دهنده ثانویه و هماهنگی به نفع و بهبود مدیریت موقعیت آینده بهره‌مند شود.

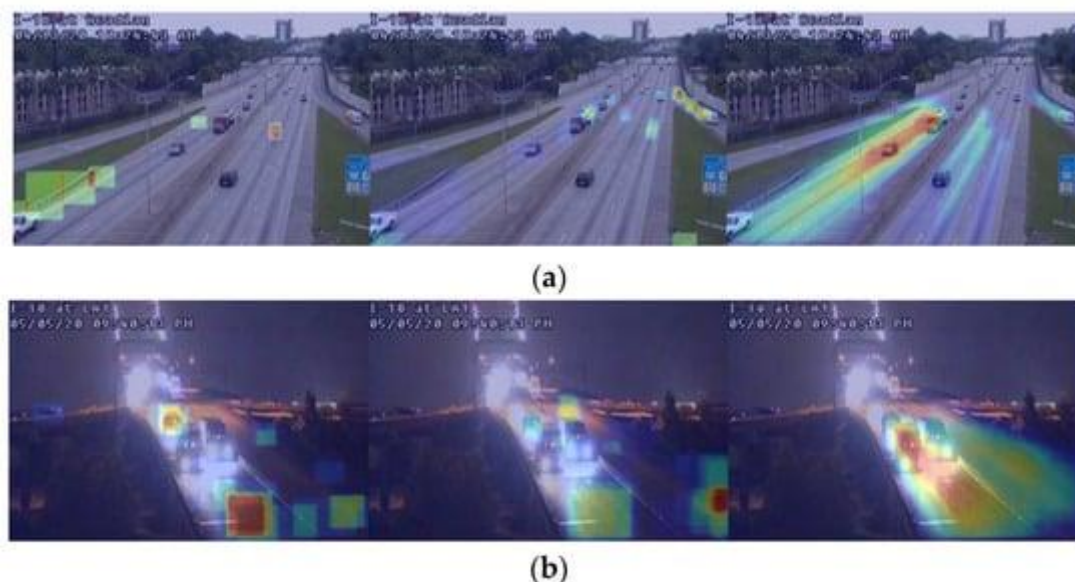
۵-۳- شمارش وسایل نقلیه

با ظهور ITS، شمارش وسایل نقلیه اغلب با استفاده از آشکارسازهای حلقه یا سیستم‌های مبتنی بر دید خودکار می‌شود. اگرچه حلقه‌های القایی شمارش دقیق ترافیک را ارائه می‌دهند؛ اما اغلب در تشخیص نوع وسایل نقلیه (مانند ماشین‌ها، کامیون‌ها، اتوبوس‌ها و غیره) مشکل دارند. فراموش نکنید که این آشکارسازها نفوذی هستند. برعکس، ماهیت غیر نفوذی سیستم‌های مبتنی بر دید، شمارش انواع کلاس‌های خودرو را با امتیازهای اطمینان بالا امکان‌پذیر می‌سازد. از آنجا که شمارش دقیق وسایل نقلیه، این امکان را به TMCها و سایر آژانس‌های حمل‌ونقل می‌دهد تا آن‌ها را در مناطقی کاربرد روزانه خود اعمال کنند، نمی‌توان اهمیت شمارش دقیق وسایل نقلیه را نادیده گرفت. مطالعاتی مانند شمارش حجم روزانه، محاسبه زمان سفر و پیش‌بینی ترافیک، همگی پیش‌آزمون‌های یک سیستم دقیق شمارش وسایل نقلیه هستند. این پارامترها به‌عنوان ابزار مهمی برای بهینه‌سازی ترافیک در جاده‌های مختلف عمل می‌کنند. به‌طور مشابه، شمارش اطلاعات مهندسان را قادر می‌سازد تا پیش‌بینی‌های ترافیکی آینده را به دست آورند که به نوبه خود به شناسایی مسیرهایی که به‌طور گسترده برای تعیین تصمیم‌های برنامه‌ریزی مثبت استفاده می‌شوند، کمک می‌کند.

برای ارزیابی عملکرد مدل‌های پیشنهادی، تعداد خودروهایی که از جهت شمال و جنوب عبور می‌کنند، به‌صورت دستی شمارش و با شمارش‌های خودکار به‌دست آمده از ترکیب دو آشکارساز و ردیاب شیء متفاوت مقایسه شد. YOLOv4 و CenterNet، دو آشکارساز شیء متفاوت بودند که در ترکیب با IOU و Feature Tracker مورد استفاده قرار گرفتند. برای مقایسه، این چارچوب‌ها در مجموع ۵۴۶ کلیپ ویدیویی، هر یک به مدت ۱ دقیقه شامل بیش از ۹ ساعت طول کلی ویدیو آزمایش شدند.

جدول ۵، مقایسه عملکرد مدل‌های YOLOv4 و CenterNet را در شرایط مختلف نشان می‌دهد. عملکرد این چارچوب‌های آشکارساز-ردیاب با تقسیم مقادیر به دست آمده از آن‌ها با حقایق زمینی شمارش شده دستی که بر حسب صدم یا درصد بیان می‌شوند، ارزیابی می‌گردد. همان‌طور که از جدول ۵ مشاهده می‌شود، ترکیب YOLOv4 و ردیاب ویژگی عملکرد شمارش معقولی را برای هر سه شرایط محیطی مختلف مشخص شده به دست آورد. برای ترکیبات مدل که در آن درصد شمارش بیش از صد به دست آمد، به وضوح در هر دو آشکارساز و ردیاب خطا وجود داشت. استدلال ترکیب آشکارساز و ردیاب برای دستیابی به دقت بیش از ۱۰۰ درصد، عمدتاً مربوط به آشکارساز شیء است که جعبه‌های محدودکننده متعددی را برای یک وسیله نقلیه تولید می‌کند. این امر منجر به شمارش بیش از حد وسایل نقلیه شد. به طور مشابه، IOU در مواقعی در پیش‌بینی مسیر وسایل نقلیه خیلی خوب عمل نمی‌کرد و آن‌ها را به عنوان وسایل نقلیه متفاوت شناسایی می‌کرد.

برای مطالعه عملکرد آشکارسازهای شیء، نقشه‌های حرارتی که منفی‌های کاذب (FN)، مثبت‌های کاذب (FP) و مثبت‌های واقعی (TP) را از چپ به راست نشان می‌دهد، در شکل ۱۴ برای مدل‌های YOLOv4 و CenterNet نشان داده شده است. YOLOv4 در تشخیص FN به خوبی عمل کرد، با این حال، CenterNet چندین وسیله نقلیه را همان‌طور که از نقشه‌های حرارتی تولید شده در جهت جنوب آن مشاهده می‌شود، شناسایی کرد. این امر تا حد زیادی به دلیل تعداد ناکافی تصاویر ترافیکی مورد استفاده برای آموزش بود. احتمال دیگر این است که این مدل به دلیل وجود وسایل نقلیه سنگین وزن ناخالص مانند اتوبوس و کامیون، ازدحام شدید در این مکان‌ها را تجربه کرده است. FP برای آشکارسازهای شیء به طور کلی برای هر دو آشکارساز شیء تمیز بوده که برای این وضعیت ایده‌آل است. برخی از موارد FP را می‌توان از YOLOv4 مشاهده کرد که می‌تواند ناشی از دید کمتر یا شرایط شبانه باشد. برای TP، هر دو مدل YOLOv4 و CenterNet پیش‌بینی‌های دقیقی را با حذف موقعیت خاصی که خودروها از دورترین خیلی دور بودند، ایجاد کردند.



شکل ۱۴. نقشه‌های حرارتی تولید شده برای شمارش وسایل نقلیه با استفاده از (الف) CenterNet و (ب) YOLOv4

۶- نتیجه گیری

پیشرفت سریع در زمینه یادگیری عمیق و محاسبات با کارایی بالا، دامنه سیستم‌های نظارت بر ترافیک مبتنی بر ویدیو را به شدت افزایش داده است. در این مطالعه، یک سیستم نظارت خودکار ترافیک ایجاد شد که بر روی مدل‌های یادگیری عمیق قوی ساخته شده و نظارت بر ترافیک را با استفاده از یک رابط کاربر گرافیکی تسهیل می‌کند. الگوریتم‌های یادگیری عمیق مانند Mask R-CNN، Faster R-CNN، YOLO و CenterNet، در کنار دو سیستم ردیابی شیء متفاوت IOU و Feature Tracker پیاده‌سازی شدند. ماسک R-CNN برای تشخیص صف‌های ترافیکی از دوربین‌های مداربسته ترافیکی بلادرنگ استفاده شد؛ درحالی‌که YOLO و R-CNN سریع‌تر برای پیش‌بینی اشیاء به کار گرفته شدند که بعداً با ردیاب اشیاء جهت شناسایی وسایل نقلیه ثابت استفاده گردیدند. Mask R-CNN، صفوف ترافیکی را با دقت ۹۰/۵ درصد پیش‌بینی کرد؛ درحالی‌که بالاترین دقت به دست آمده توسط YOLO 93.7 درصد بود. اختلاف در تشخیص صحیح صف‌ها عمدتاً به دلیل کیفیت پایین تصویر، دور بودن صف‌ها از دوربین و جلوه‌های خیره‌کننده بود. این مسائل به طور قابل توجهی بر دقت مدل‌های پیشنهادی تأثیر گذاشت. به طور مشابه، امتیازات F1، RMSE و S3 برای تشخیص وسایل نقلیه ساکن، به ترتیب ۰/۸۳۳۳، ۱۵۴/۷۷۴۱ و ۰/۴۰۳۴ بود. مشاهده شد که مدل به درستی وسایل نقلیه سرگردان را که نزدیک‌تر به دوربین باقی می‌مانند؛ اما هنگام تشخیص وسایل نقلیه ثابت دور با مشکلاتی مواجه می‌شوند، تشخیص می‌دهد. بخشی از مشکل برای امتیازات پایین‌تر S3 نیز به دلیل مسائلی مانند پیکسل‌سازی ویدیو و وجود تقاطع‌های ترافیکی بود. صرف نظر از این، روش‌هایی مانند سرکوب ناهنجاری و اصلاحات پیکسلی ویدیو در بهبود کارایی مدل پیشنهادی مفید بودند. شایان ذکر است که این اصلاحات منجر به یک سیستم مؤثر پیش‌بینی خودرو ثابت شد. در نهایت، عملکرد چارچوب شمارش خودرو برای هر دو ترکیب CenterNet و YOLO با Feature Tracker رضایت‌بخش بود. با این حال، چارچوب شمارش وسایل نقلیه را می‌توان بیشتر مورد بررسی قرار داد و مدل‌های موجود را بیشتر تنظیم کرد تا یک چارچوب شمارش تقریباً کامل ایجاد شود. درواقع، این مدل می‌تواند برای اکثر آژانس‌های حمل و نقل ایده‌آل باشد؛ زیرا آن‌ها به شدت به تعداد حرکت چرخشی برای بهینه‌سازی علائم ترافیکی در تقاطع‌ها متکی هستند.

در نتیجه، مدل‌های پیشنهادی که یک سیستم نظارت بر ترافیک یکپارچه مبتنی بر هوش مصنوعی را تشکیل می‌دهند، نتایج برتری به دست آوردند و می‌توانند در دستیابی به سطحی از اتوماسیون در مراکز مدیریت ترافیک مفید باشند. از آنجاکه اکثر نرم‌افزارهای فروخته شده توسط شرکت‌های فروشنده حمل و نقل بیش از صدها هزار دلار هزینه دارند، عملکرد آن‌ها محدود است و فقط چند قابلیت نظارت بر ترافیک بیشتر از چارچوب پیشنهادی ما را ارائه می‌دهند. بدین ترتیب، سیستم پیشنهادی در این مقاله می‌تواند یک جایگزین ارزان‌تر و قابل اعتماد برای وارد کردن سطحی از اتوماسیون ترافیک با تکمیل آن با مجموعه‌های نرم‌افزاری اضافی کم‌هزینه باشد.

منابع

- Carli, R., Dotoli, M., Epicoco, N., Angelico, B., & Vinciullo, A. (2015, August). Automated evaluation of urban traffic congestion using bus as a probe. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)* (pp. 967-972). IEEE.
- Chakraborty, P., Adu-Gyamfi, Y. O., Poddar, S., Ahsani, V., Sharma, A., & Sarkar, S. (2018). Traffic congestion detection from camera images using deep convolution neural networks. *Transportation Research Record*, 2672(45), 222-231.
- Fouladgar, M., Parchami, M., Elmasri, R., & Ghaderi, A. (2017). Scalable deep traffic flow neural networks for urban traffic congestion prediction. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 2251-2258). IEEE.

- GitHub. (2020). *Video demonstration of a GUI based AI enabled traffic monitoring system*. <https://github.com/titanmu/aienabled>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2961-2969).
- He, K., Sun, J., & Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12), 2341-2353.
- Land, E. H. (1986). An alternative technique for the computation of the designator in the retinex theory of color vision. *Proceedings of the national academy of sciences*, 83(10), 3078-3080.
- Litman, T. (2007). Developing indicators for comprehensive and sustainable transport planning. *Transportation Research Record*, 2017(1), 10-15.
- Ma, X., Yu, H., Wang, Y., & Wang, Y. (2015). Large-scale transportation network congestion evolution prediction using deep learning theory. *PloS One*, 10(3), e0119044.
- Morris, T., Schwach, J. A., & Michalopoulos, P. G. (2011). *Low-cost portable video-based queue detection for work-zone safety* (No. CTS 11-02). University of Minnesota. Center for Transportation Studies. Intelligent Transportation Systems Institute.
- Rahman, Z. U., Jobson, D. J., & Woodell, G. A. (1996). Multi-scale retinex for color image enhancement. In *Proceedings of 3rd IEEE International Conference on Image Processing* (Vol. 3, pp. 1003-1006). IEEE.]
- Wang, J., Gu, Q., Wu, J., Liu, G., & Xiong, Z. (2016, December). Traffic speed prediction and congestion source exploration: A deep learning method. In *2016 IEEE 16th International Conference on Data Mining (ICDM)* (pp. 499-508). IEEE.
- Willis, C., Harborne, D., Tomsett, R., & Alzantot, M. (2017). A deep convolutional network for traffic congestion classification. In *Proceedings of the NATO IST-158/RSM-010 Specialists' Meeting on Content Based Real-Time Analytics of Multi-Media Streams* (pp. 1-11).

