علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۲۰، شماره۱، ۱۴۰۱
صفحات ۹۵-۱۰۷
مقاله پژوهشی

تشخیص و شناسایی علائم ترافیکی با استفاده از مدل‌های YOLOv3

مائده نارویی[1*] و فرحناز مهنا[2]

*نویسنده مسئول، دریافت: ۱۴۰۰/۱۰/۰۸، بازنگری: ۱۴۰۱/۰۱/۲۷، پذیرش: ۱۴۰۱/۰۳/۱۷

[1] کارشناس ارشد مهندسی مخابرات سیستم، دانشکده مهندسی برق وکامپیوتر، دانشگاه سیستان و بلوچستان، زاهدان، ایران
[2] دانشیار مهندسی برق-پردازش‌داده، دانشکده مهندسی برق و کامپیوتر، دانشگاه سیستان و بلوچستان، زاهدان، ایران

چکیده

در این در این مقاله از شبکه‌های YOLOv3، YOLOv3-SPP و Tiny YOLOv3 برای شناسایی علائم ترافیکی پایگاه جدیدی شامل ۴۰۰۰ تصویر و ۲۴ کلاس مختلف از علائم ترافیکی شهر زاهدان استفاده شده است. تصویربرداری در حال حرکت و در شرایط مختلف آب و هوایی، زوایای دید متفاوت، همراه با انسداد علائم، و پس‌زمینه شلوغ انجام شده است. برای عملکرد بهتر شبکه، تعداد این تصاویر با استفاده از تکنیک‌های افزایش داده مانند چرخش، تغییر اندازه، افزودن نویز و تولید داده‌های مصنوعی به ۲۶۰۰۰ تصویر افزایش یافته است. آموزش و ارزیابی شبکه‌ها با استفاده از هر دو پایگاه انجام شده است. برای کاهش تعداد گام‌ها و زمان آموزش و جلوگیری از بیش‌برازش از انتقال یادگیری استفاده شده است. مقایسه‌ی نتایج بالاتر بودن میانگین دقت متوسط شناسایی را برای YOLOv3-SPP و بالاتربودن سرعت را برای Tiny YOLOv3 نشان می‌دهد. میانگین دقت متوسط برروی پایگاه تصاویر افزایش‌یافته برای مدل‌های YOLOv3 و YOLOv3-SPP به ۰/۹۸۸ رسیده است.

کلمات کلیدی: شناسایی علائم ترافیکی، مدل YOLOv3، تولید داده مصنوعی، آموزش شبکه

۱- مقدمه

به‌طورکلی شناسایی علائم ترافیکی از دو بخش، تشخیص و طبقه‌بندی تشکیل شده است. هدف از تشخیص، یافتن مکان و ابعاد علائم ترافیکی موجود در تصویر است، و طبقه‌بندی به‌هر یک از تصاویر تشخیص داده شده، برچسبی اختصاص می‌دهد که برای هر دو بخش از تکنیک‌های پردازش تصویر مبتنی بررنگ و شکل علائم ترافیکی استفاده شده است. در سال‌های اخیر با پیشرفت روش‌های یادگیری ماشین[1] به‌خصوص یادگیری عمیق[2]، این روش‌ها در تشخیص و طبقه‌بندی علائم ترافیکی نسبت به‌روش‌های سنتی مبتنی بر رنگ و شکل در برخی از جنبه‌ها برتری‌هایی دارند. روش‌های یادگیری ماشین اغلب به تعداد زیادی از نمونه‌های آموزشی حاوی اطلاعات کافی از شکل و رنگ، نیاز دارند [۱]. یکی از مشهورترین انواع شبکه‌های عصبی عمیق، شبکه عصبی کانولوشن[3] است که در آن شبکه‌ی عصبی با افزایش داده‌ها و اندازه‌ی مدل، توسعه یافته و خود را با شرایط جدید وفق داده و برای دسته‌بندی تصاویر نیازی به‌استخراج ویژگی‌های آن‌ها ندارد که چنین خصوصیاتی باعث شده است که شبکه‌ی عصبی کانولوشن در شناسایی و طبقه‌بندی اشیا بسیار دقیق عمل کند. همچنین روش‌های مبتنی بر شبکه‌ی عصبی کانولوشن، در مقایسه با روش‌های دیگر یادگیری ماشین مانند آدابوست و ماشین بردار پشتیبان، نیاز به-ویژگی‌های طراحی شده به‌صورت دستی ندارد و به‌همین دلیل می‌تواند تعداد بسیار بیشتری از نمونه‌های آموزشی را مدیریت کند و یک مدل تشخیص دهنده با توانایی تعمیم‌دهی بهتری تولید نماید. درکنار امتیازهای شبکه‌ی عصبی کانولوشن، جمع‌آوری و ذخیره‌سازی حجم عظیمی از تصاویر و همچنین نیاز به‌پردازشگرهای گرافیکی قدرتمند و گران قیمت، دو چالش مهم استفاده از این شبکه‌ها هستند. اکثر پایگاه‌های تصاویر در حوزه‌ی شناسایی علائم ترافیکی، دارای تعداد کافی از نمونه‌های آموزشی و آزمایشی نیستند [۲]. ولی با استفاده از روش‌های مبتنی بر یادگیری عمیق در حوزه‌ی شناسایی علائم ترافیکی، امکان بهبود شناسایی اشیا با سایز کوچک، زیاد است. البته مشکلاتی از قبیل نبود اطلاعات ظاهری کافی برای تمایز اشیا کوچک از پس‌زمینه یا طبقه‌های مشابه وجود داشته و نیاز به‌اطلاعات دقیق‌تری برای شناسایی دقیق مکان اشیا کوچک است. از طرفی، تجربیات و دانش در حوزه‌ی تشخیص اشیا کوچک بسیار محدود است زیرا که اکثر کارهای گذشتگان در حوزه‌ی تشخیص اشیا بزرگ انجام شده است. بنابراین باید مدل شبکه را به‌گونه‌ای طراحی

نمود که در زمان بلادرنگ و با دقت کافی شناسایی و طبقه‌بندی علائم ترافیکی را اجرا کند [۳]. البته لازم به‌ذکر است که اگرچه شناسایی علائم ترافیکی، عملی ساده برای چشم انسان است، اما به‌علت وجود عواملی مانند محو بودن بخشی از تابلوی علامت ترافیکی به‌دلیل موانع مختلف، شرایط آب و هوایی، و روشنایی محیط، این موضوع مسئله‌ای چالش برانگیز در بینایی ماشین است. بنابراین برای اینکه روش‌های مبتنی بر شبکه‌ی عصبی کانولوشن، دقت مناسبی داشته باشند، نیاز است که تصاویر فراوانی از تابلوهای علائم ترافیکی با شرایط نورپردازی نامناسب، و یا حاوی دیگر عوامل منفی تاثیرگذار در تشخیص آن‌ها، تهیه شوند. البته تهیه‌ی چنین تصاویری به‌زمان زیادی نیاز دارد و یکی از راهکارهای حل این مشکل، طراحی شبکه‌ای است که بتواند حتی با داده‌های مصنوعی و تعداد کم نمونه‌ها، محاسبات مورد نیاز را در زمان بلادرنگ انجام دهد. در این مقاله برای تشخیص و شناسایی علائم ترافیکی از مدل شبکه‌ی YOLOv3 و دو مدل توسعه‌یافته‌ی آن استفاده شده‌اند که دارای سرعت بلادرنگ مناسبی بوده و سعی‌شده تا در کنار سرعت بلادرنگ، دقت آن‌ها نیز افزایش داده شوند. مزیت اصلی استفاده از مدل شبکه‌ی YOLOv3 این است که این مدل تنها به‌یک راه انتشار روبه جلو در شبکه برای شناسایی چندین شی در یک تصویر ورودی، نیاز دارد. این شبکه به‌صورت هم‌زمان تمام سناریوهای ممکن و اشیا محتمل بر روی تصویر را آنالیز می‌کند و عمل پیش‌بینی چندین ویژگی را انجام می‌دهد که بسیار موثرتر از استفاده از یک پنجره‌ی لغزنده است که به‌چندین تکرار برای پردازش یک تصویر نیاز دارد. مزیت دیگر شبکه‌ی YOLOv3، پیش‌بینی چندمقیاسی است که به‌رفع مشکل شناسایی اشیا کوچک کمک می‌کند. پایگاه تصاویر استفاده شده در این مقاله شامل تصاویر صحنه‌های شهری زاهدان می‌باشد که با دوربین گوشی موبایل Xiaomi A2 Lite و از درون ماشین درحال حرکت برای شبیه‌سازی بهتر در سناریوهای دنیای واقعی تصویربرداری شده‌اند. در ادامه‌ی این مقاله، بررسی کارهای اخیر در شناسایی علائم ترافیکی در بخش دوم مشاهده می‌شود. سپس در بخش سوم شبکه‌های پیشنهادی شامل YOLOv3 ارتقایافته (با اصلاح تابع زیان آن)، YOLOv3-SPP و Tiny YOLOv3 برای شناسایی علائم ترافیکی در پایگاه جدید معرفی می‌شوند. در بخش چهارم، ابتدا پایگاه جدید تصاویر علائم ترافیکی و معیارهای ارزیابی استفاده شده برای تحلیل نتایج شبکه‌های پیشنهادی برروی این پایگاه معرفی و سپس، این نتایج باهم مقایسه می‌شوند. در نهایت در بخش پنجم به‌نتیجه‌گیری می‌پردازیم.

۲- کارهای پیشین

برای شناسایی علائم ترافیکی از مدل YOLOv3 استفاده شده و عملکرد شبکه بر روی پایگاه تصاویر GTSDB شامل ۳۰۰ تصویر آزمایشی ارزیابی و میانگین دقت متوسط (mAP)، ۹۲/۲ ٪ گزارش شده است [۴]. زمان پردازش یک تصویر به‌طور متوسط ۱۰۱ میلی‌ثانیه و نرخ متوسط پردازش هر فریم تقریبا ۹/۸۷ فریم بر ثانیه اعلام شده است. شناسایی علائم ترافیکی کوچک، به‌خوبی انجام شده ولی مکان‌یابی دقیق آن‌ها هنوز مشکل دارد. چند مورد تشخیص مثبت اشتباه نیز در شناسایی علائم ترافیکی در مجاورت علائم تبلیغاتی نیز مشاهده شده که باعث طبقه‌بندی اشتباه آن‌ها شده است. مقایسه‌ی شناسایی علائم ترافیکی مبتنی بر YOLOv*3* با شناسایی مبتنی بر Faster R-CNN روی پایگاه تصاویر آزمایشی GTSDB نشان داده است که میانگین دقت متوسط ۷/۷ درصد افزایش یافته است. با انجام بهبودهایی شامل، هرس‌کردن شبکه، ایجاد شاخه‌های پیش‌بینی با چهار مقیاس (YOLOv3-4SBP) و اصلاح تابع زیان در شبکه‌ی YOLOv*3*، روشی برای شناسایی علائم ترافیکی بر روی پایگاه تصاویر TT100K شامل ۱۰۰۰۰۰ تصویر، ارائه شده است [۵]. دقت شناسایی علائم ترافیکی با مدل شبکه‌ی YOLOv*3*، ۹۰/۴۷ ٪، با YOLOv*3* هرس شده ، ۹۰/۴۵ ٪، با YOLOv3 -4SBP ، ۹۱/۹۱ ٪ و با YOLOv*3* نهایی که در آن تابع زیان اصلاح شده، ۹۴ ٪ گزارش شده‌اند. اندازه‌ی شبکه‌ی YOLOv*3* هرس شده ۱۱/۳ ٪ نسبت به YOLOv*3* اصلی کاهش‌یافته که به‌معنای محاسبات کمتر و زمان پاسخ کوتاه‌تر است. همان‌طور که مشاهده می‌شود مدل نهایی شبکه‌ی YOLOv*3* بعد از تنظیم وزن‌ها در تابع زیان، بهبود قابل توجهی یافته و عملکرد مناسبی در شناسایی و طبقه‌بندی علائم ترافیکی در اندازه‌های متفاوت داشته است. برای شناسایی علائم ترافیکی، هشت شبکه‌ی عصبی کانولوشن با ادغام هرمی فضایی[۴] ترکیب و مدل‌های YOLOv3 SPP، Densenet SPP، Resnet 50 SPP و Tiny YOLOv3 SPP را ارائه داده‌اند [۶]. برای آموزش و تنظیم پارامترهای شبکه‌ها از پایگاه تصاویر علائم ترافیکی تایوان شامل ۱۰۰۰ تصویر و چهار کلاس، استفاده شده است. برای آزمایش، ۲۰ عدد از این تصاویر به‌کار رفته‌اند. میانگین دقت متوسط ۹۸/۷۳٪ برای شبکه‌ی YOLOv3 ، ۹۸/۸۸ ٪ برای YOLOv3 SPP، ۹۸/۳۳ ٪ برای Densent، ۹۸/۵۳ ٪ برای Densent SPP، ۹۷/۰۹ ٪ برای Resnet 50، ۹۷/۷۰ ٪ برای Resnet 50 SPP، ۸۲/۶۹ ٪ برای Tiny YOLOv3، و ۸۹/۷۹ ٪ برای Tiny YOLOv3 SPP گزارش شده‌اند. زمان آزمایش نیز به‌ترتیب برای شبکه‌های مذکور ۱۷، ۱۷/۶، ۳۸/۳، ۴۰، ۱۵/۳، ۱۹/۳، ۵/۴، و ۸ میلی‌ثانیه به‌دست آمده‌اند. همان‌طور که دیده می‌شود، زمان و دقت میانگین متوسط مدل‌های ترکیب شده با معماری SPP افزایش یافته است لذا این مدل‌ها به‌زمان بیشتری برای شناسایی علائم ترافیکی نیاز دارند. روش‌های مبتنی بر ناحیه مورد نظر[۵] با معماری‌های YOLOv3، YOLOv4 و Tiny YOLO ترکیب و برای شناسایی علائم ترافیکی در پایگاه تصاویر DFG اعمال شده‌اند [۷]. این پایگاه شامل ۷۰۰۰ تصویر علائم ترافیکی با کیفیت تمام HD[۶] در ۲۰۰ کلاس است که از محیط جاده‌ای اسلوونی تصویربرداری شده است. در حالت تمام HD، دقت شناسایی یک شیء و زمان پردازش یک تصویر بسیار اهمیت دارد که در این مرجع، برای بررسی دقت شناسایی علائم ترافیکی از معیار mAP برروی تمام کلاس‌های تصاویر استفاده شده است. زمان پردازش نیز، متوسط زمان مورد نیاز برای پردازش یک تصویر از مجموعه تصاویر آزمایشی می‌باشد. معیار mAP و زمان پردازش برای یک تصویر آزمایشی نمونه با اندازه‌ی ۷۰۴×۷۰۴ از پایگاه تصاویر DFG توسط شبکه‌ی Tiny YOLO، ۸۱/۲۸ ٪ و ۱۱/۳۶ میلی‌ثانیه، با YOLOv3، ۸۸/۹۹ ٪ و ۳۵/۷۱ میلی‌ثانیه، و با YOLOv4، ۹۲/۶۵ ٪ و ۳۴/۴۸ میلی‌ثانیه گزارش شده است. از آنجایی‌که در معماری پیچیده‌ی YOLOv4، منابع GPU[۷] بیشتری استفاده شده است، میانگین دقت متوسط افزایش یافته درحالی‌که ویژگی‌های متمایز بیشتری نیز استخراج شده است. یک سیستم شناسایی علائم ترافیکی پیشرفته مبتنی بر تکنیک‌های پیش‌پردازش و بهینه‌سازی در [۸] معرفی شده است که برروی پایگاه تصاویر علائم ترافیکی کشور کره (KTSD) شامل ۳۳۰۰ تصویر با ابعاد کوچک، و با کیفیت و رزولوشن پایین به‌کار رفته‌است. در طی پیش‌پردازش، کنتراست رنگ در تصاویر جاده‌ای ورودی ارتقا یافته و لبه‌ها استخراج شده‌اند تا شناسایی علائم ترافیکی با اندازه کوچک‌تر آسان‌تر شود. همچنین برای طبقه‌بندی علائم ترافیکی از شبکه‌ی YOLOv3 اصلاح شده استفاده شده است. در این اصلاحات اندازه شبکه‌ی سلولی در سطح سوم افزایش یافته و ابعاد سلول شبکه کاهش یافته است تا شبکه چگال‌تر شود. همچنین ۹جعبه‌ی لنگر[۸] با استفاده از تصاویر آموزش، محاسبه و از ۰ تا ۸ عددگذاری شده‌اند. به‌علاوه، برای حذف پیش‌بینی‌های نادرست، از یک آستانه‌ی بهینه استفاده شده که با روش bisection تعیین شده است. نتایج نشان داده که عملکرد YOLOv3 بهینه‌شده به‌طرز قابل توجهی بهتر از عملکرد YOLOv3 اصلی است و دقت متوسط میانگین، ۱۴/۳٪ بهبود پیدا کرده است. همچنین، با استفاده از مدل شبکه‌ی پیشرفته مبتنی بر بهینه‌سازی و تکنیک‌های پیش‌پردازشی (ATSR) ، میانگین دقت متوسط در ازای افزایش اندکی در زمان پردازش، ۲۴/۲۱ ٪ افزایش یافته است. برای مثال، زمان پردازش یک تصویر با اندازه‌ی ۸۰۰×۸۰۰ توسط مدل ATSR، ۰/۰۶۳ ثانیه است درحالی‌که با مدل YOLOv3، این زمان ۰/۰۵۰ ثانیه می‌باشد. در این مرجع همچنین دقت طبقه‌بندی علائم ترافیکی در پایگاه تصاویر علائم ترافیکی آلمان (GTSDB) شامل ۴۳ کلاس با ۹۰۰ تصویر، با استفاده از مدل

YOLOv3، ۹۶/۵۳ ٪ و با مدل ATSR، ۱۰۰ ٪ به‌دست آمده است. در دو بخش از شبکه‌ی استخراج ویژگی YOLOv3 به‌منظور شناسایی اطلاعات جاده‌ای شامل وسایل نقلیه، عابران پیاده، علائم ترافیکی و خطوط این علائم، بهینه‌سازی‌هایی انجام شده است [۹]. در بخش اول، یک شبکه‌ی کمکی استخراج ویژگی با تکثیر شبکه‌ی بک-بون Backbone Darknet-53 موجود در YOLOv3، ایجاد شده که باعث بهبود استخراج ویژگی شده است. در بخش دوم، از مکانیزم توجه[9] برای تلفیق اطلاعات ویژگی‌های استخراجی توسط شبکه‌ی YOLOv3 اصلی با اطلاعات ویژگی‌های استخراجی توسط شبکه‌های کمکی backbone تکثیری استفاده شده که منجر به-بهبود عملکرد تمام شبکه‌ی استخراج ویژگی شده است. پایگاه تصاویر این مرجع، پایگاه منبع باز[10] خودران ApolloScape است که بسیاری از وضعیت‌های پیچیده جاده‌ها را پوشش داده و شامل ۴ کلاس و هر کلاس شامل ۲۰۰۰ تصویر آموزشی و ۵۰۰ تصویر آزمایشی می‌باشد. نتایج آزمایش‌ها نشان داده که میانگین دقت متوسط عملکرد شبکه‌ی YOLOv3 بهینه در مقایسه با شبکه‌ی YOLOv3 اصلی، ۵/۴۳٪ افزایش یافته و زمان متوسط پردازش یک تصویر توسط شبکه‌ی YOLOv3 بهینه، ۲۴/۳۹ میلی‌ثانیه گزارش شده است. در راستای پیاده‌سازی یک سیستم کاربردی به‌منظور تشخیص علائم ترافیکی تایلند از مدل شبکه‌ی YOLOv3 و پایگاه تصاویر علائم ترافیکی آلمان GTSRB استفاده شده است [۱۰]. در تایلند از علائم ترافیکی پایگاه GTSRB المان استفاده می‌شود که ۴۳ کلاس با ۵۰۰۰۰ تصویر دارد. انتخاب شبکه‌ی YOLOv3 به‌دلیل سرعت بالای آن بوده که کاربرد این شبکه را در شناسایی بلادرنگ علائم ترافیکی مناسب کرده است. دقت متوسط شناسایی این علائم ترافیکی به‌ترتیب، توسط مدل‌های YOLOv3-608 و Tiny YOLOv3، ۹۱ ٪ و ۷۳ ٪ گزارش شده است. همچنین زمان شناسایی مدل شبکه‌ی YOLOv3-608، ۵۱ میلی‌ثانیه ارائه شده است. شناسایی اشیاء کوچک نظیر علائم ترافیکی در یک معیار بزرگ[11] از علائم ترافیکی TT100K در [۱۱] ارائه شده است. این پایگاه از ۱۰۰۰۰۰ پانوراما‌ی نمای خیابان Tencent کشور چین شامل ۳۰۰۰۰ نمونه علائم ترافیکی، ایجاد شده است که فراتر از معیارهای قبلی است و همچنین تغییرات زیادی در میزان روشنایی و شرایط آب و هوایی را پوشش داده است. هر علامت ترافیکی در TT100K، با یک برچسب کلاس و یک جعبه‌ی محصورکننده و ماسک پیکسل، حاشیه‌نویسی شده است. سپس نشان داده شده که یک شبکه‌ی عصبی کانولوشن (CNN) می‌تواند به-طور هم‌زمان علائم ترافیکی را شناسایی و طبقه‌بندی کند. یک نرم‌افزار به‌منظور شناسایی علائم ترافیکی مبتنی بر شبکه‌های YOLOv3 و CNN توسعه‌یافته ارائه شده که شبکه‌ی YOLOv3 فقط برای شناسایی پنج شیء شامل ماشین، کامیون، عابرپیاده، علامت ترافیکی و چراغ ترافیکی آموزش دیده است [۱۲]. سپس اشیاء شناسایی شده وارد شبکه‌ی CNN توسعه‌یافته شده و به ۷۵ گروه طبقه‌بندی شده‌اند. در این مرجع، شناسایی بیش از ۹۹/۲ درصد از علائم ترافیکی در شرایط متنوع انجام شده است. چارچوب YOLOv3 برای شناسایی علائم ترافیکی مالزی در محیط واقعی پیاده‌سازی شده که با گنجاندن ادغام هرم فضایی(SPP) در چارچوب YOLOv3، شناسایی علائم ترافیکی دور و کوچک را در محیط واقعی به‌دلیل ادغام شبکه‌های چند اندازه‌ی SPP امکان‌پذیر شده است [۱۳]. نتایج برروی یک پایگاه تصاویر شامل ۲۰۰۰ تصویر علائم ترافیکی مالزی میانگین دقت شناسایی را ۸۲/۵ ٪ نشان داده است. روشی مبتنی برمتن برای شناسایی متن در پنل‌های ترافیکی در مرجع [۱۴] ارائه شده است. دو پایگاه تصاویر پنل‌های ترافیکی مبتنی برمتن فارسی از خیابان‌های تهران جمع‌آوری شده‌اند که اولی، ۹۲۹۴ و دومی ۳۳۰۵ تصویر دارند. از آنجایی‌که پایگاه تصاویر دومی یکنواخت‌تر از اولی می‌باشد، شبکه با استفاده از پایگاه تصاویر اولی، پیش‌آموزش و با پایگاه دوم آموزش داده شده است. علاوه‌براین، از مدل شبکه‌ی Tiny YOLOv3، نیز برای پیش‌آموزش، آموزش و آزمایش پایگاه تصاویر بهره برده شده است که دارای پیچیدگی کم و سرعت بالا می‌باشد. همچنین از روش اعتبارسنجی متقاطع K-fold استفاده شده و نتایج، دقت ۰/۹۷۳ را نشان داده است.

برای شناسایی با دقت بیشتر اهداف کوچک، یک روش مبتنی بر مدل YOLOv3، برای شبکه‌های چندمقیاسی پیشنهاد شده است [۱۵]. این روش از چهار بخش شامل ۱) خوشه‌بندی K-means برای انتخاب فریم‌های لنگر و همگرایی مدل شتاب-دهنده، ۲) همجوشی تطبیقی چندمقیاسی برای استخراج ویژگی‌ها و افزایش اطلاعات پردازشی شبکه، ۳) شناسایی انتها به انتها برای پیش‌بینی شبکه به‌منظور بهبود سرعت شناسایی، و ۴) امتیاز آستانه برای رتبه‌بندی و استفاده از NMS برای فیلترکردن ماکزیمم‌های محلی و تولید جعبه‌ی محصورکننده‌ی پیش‌بینی، تشکیل شده است. آموزش و آزمایش شبکه روی پایگاه تصاویر علائم ترافیکی CCTSDB انجام شده‌اند و آزمایش‌ها نشان داده‌اند که این روش در مقایسه با YOLOv3 اصلی، بهبود قابل توجهی در شناسایی اهداف کوچک داشته است. سیستمی ارائه شده است که بتواند علائم ترافیکی و فاصله‌ی آن‌ها از وسایل نقلیه را در روشنایی غیرایده‌آل و همچنین در شرایط آب و هوایی مختلف تشخیص دهد [۱۶]. شبکه‌ی مورد استفاده YOLOv3 می‌باشد که آموزش و آزمایش آن با پایگاه تصاویر GTSRB انجام و دقت ۹۸/۵ ٪ گزارش شده است. راه‌حلی برای شناسایی مجموعه‌ای خاص از ۱۱ علامت ترافیکی معمولی برای اکثر کشورهای اروپایی معرفی شده است [۱۷] که مدل شبکه‌ی مورد استفاده YOLOv3 می‌باشد. این شبکه برروی پایگاه تصاویر ویدیویی تهیه‌شده توسط نویسندگان آموزش دیده و آزمایش شده است. این تصاویر ویدیویی با نصب یک دوربین نمای جلو در داخل وسیله‌ی نقلیه و در شهر Osijek در شرایط آب و هوایی مختلف ضبط شده‌اند که از ۲۸ توالی ویدیویی مختلف شامل ۵۵۶۷ تصویر از ۶۷۵۱ علائم ترافیکی تشکیل شده‌اند. عملکرد روش آن‌ها دقت خوبی داشته است.

۳- شبکه‌های پیشنهادی برای شناسایی علائم ترافیکی‌ها

با توجه به‌تحقیقات گذشتگان، شبکه‌ی انتخابی برای شناسایی علائم ترافیکی با سرعت بلادرنگ، مدل‌های ارتقایافته‌ی شبکه‌ی YOLOv3 می‌باشند که با کاهش لایه‌ها و پارامترهای آن می‌توان شبکه را بهبود بخشید و محاسبات آن را کاهش داد.

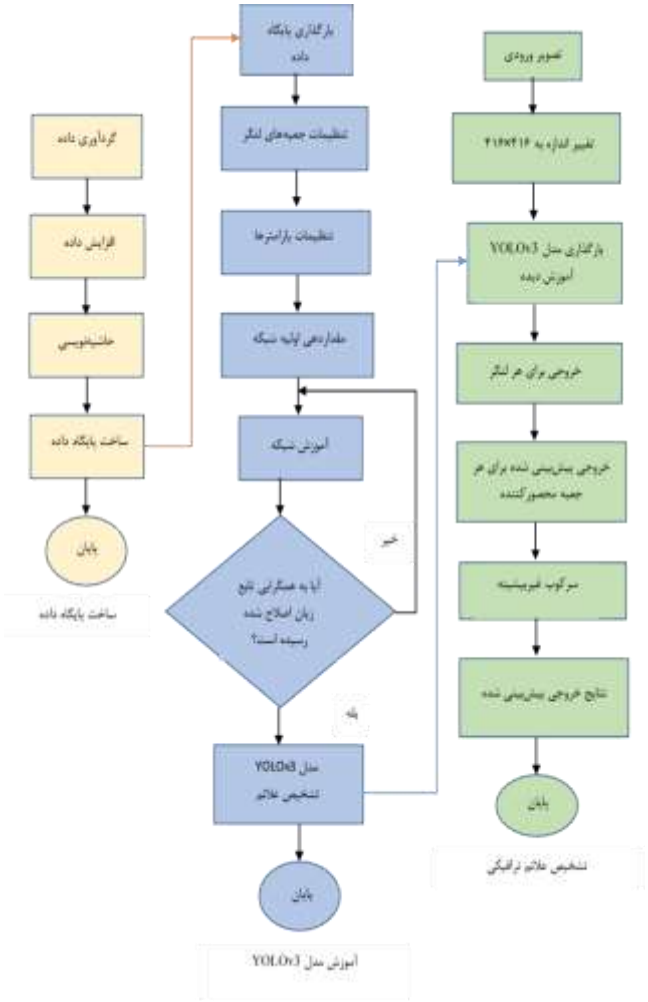
۳-۱- مدل شبکه‌ی YOLOv3

مدل شبکه‌ی YOLOv3 یک تشخیص‌دهنده‌ی سریع و پیشرفته‌ی شی است که نسخه‌ی بهبودیافته‌ی شبکه‌ی YOLO است. الگوریتم این شبکه در شکل ۱ آورده شده است. تشخیص‌دهنده‌ی YOLO از یک شبکه عصبی استفاده می‌کند که در یک تکرار، موقعیت اشیا و نمره کلاس را پیش‌بینی می‌کند [۱۸]. این‌کار با در نظر گرفتن عمل تشخیص به‌عنوان یک مسئله رگراسیون، که مدل را بر روی یک تصویر ورودی در موقعیت‌ها و مقیاس‌های مختلف اعمال می‌کند، قابل دستیابی است. مناطقی با امتیاز بالاتر در تصویر به‌عنوان تشخیص‌ها درنظر گرفته می‌شوند. در واقع این شبکه تصویر ورودی را به‌تعداد زیادی شبکه‌های سلولی S×S تقسیم می‌کند و برای هر سلول جعبه‌های محصورکننده‌ی[12] B که شامل ۴ ویژگی، طول، عرض و مختصات مرکز جعبه است، پیش‌بینی می‌کند. هر کدام از این جعبه‌ها دارای مقدار $Pobject$ بوده و تعداد کلاس‌های داخل هر جعبه با مقدار C بیان می‌شود. احتمال وجود جعبه در هر سلول با احتمال شرطی کلاس، $Pclass$ مشخص می‌شود. پیش‌بینی کلی شبکه برابر با $S \times S \times (B \times 5 + C)$ است که S×S اندازه شبکه سلولی، B تعداد جعبه‌های محصورکننده، و رقم ۵، جمع ۴ به‌عنوان تعداد ویژگی‌های جعبه‌های محصورکننده، و ۱ به‌عنوان احتمال شی می‌باشد. C نیز تعداد کلاس‌ها را مشخص می‌کند (شکل ۲). در طول آموزش شبکه، تعداد کلاس‌های حاضر در هر شبکه سلولی در تصویر با استفاده از معادله (۱) محاسبه می‌شود.

$$p(class_i) = p(class_i|object) \times p(object) \quad (۱)$$

که $p(class_i)$ برابر احتمال i-امین کلاس، $Pobject$ احتمال شبکه سلولی شامل شی و $P(class_i|object)$ احتمال شرطی کلاس i-ام که در آن شی حاضر می‌باشد،

است. $Pmin$ در ابتدا تعریف شده و شبکه تنها اشیائی را تشخیص می‌دهد که $Pclass > Pmin$ در طول پس‌پردازش، تشخیص‌های تکراری از یک شی به وسیله سرکوب غیربیشینه[۱۳] حذف می‌شوند. از آنجایی که شبکه برای هر سلول در تصویر

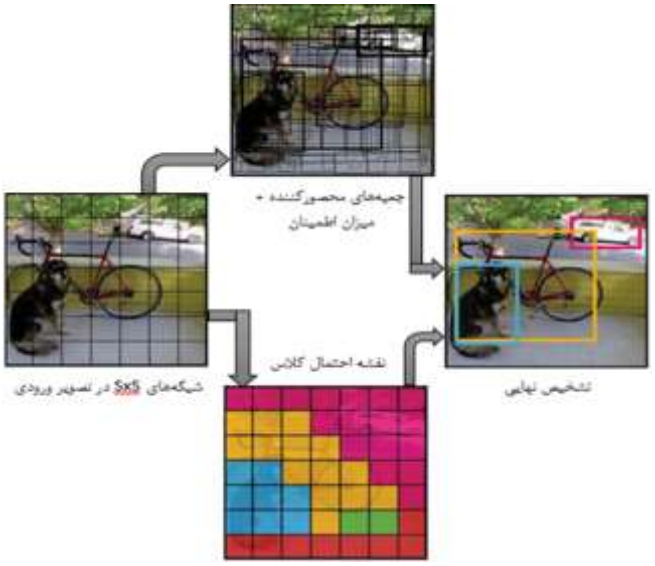


شکل ۱- بلوک دیاگرام الگوریتم مدل شبکه‌ی YOLOv3

چندین جعبه‌ی محصورکننده پیش‌بینی می‌کند، تنها جعبه‌هایی که بالاترین نمره اطوینان را دارند، انتخاب می‌شوند. درنتیجه، شبکه، خروجی برابر با $S \times S \times (B \times 5 + C)$ تولید می‌کند. در YOLOv3 جعبه‌های محصورکننده با لنگر جایگزین شده‌اند که مشکل گرادیان ناپایدار ایجاد شده در طول آموزش شبکه را رفع کنند. گرادیان ناپایدار یک مشکل اساسی است که در شبکه‌های عمیق در لایه‌های اولیه رخ می‌دهد که موجب محوشدگی و انفجار گرادیان می‌شود. البته گرادیان ناپایدار تنها مشکل محوشدگی گرادیان نیست بلکه بیشتر به‌دلیل این مورد است که گرادیان در لایه‌های اولیه حاصل روند تمام لایه‌ها است و لایه‌های بیشتر، شبکه را ذاتی به‌یک راه‌حل ناپایدار سوق می‌دهند. ایجاد توازن بین تمام موارد، تنها راهی است که هر لایه در شبکه‌ی عصبی می‌تواند با سرعت یکسان پیش رود و از محوشدگی گرادیان‌ها جلوگیری کند. شانس ایجاد توازن با بالا رفتن تعداد لایه‌ها کمتر می‌شود. درنتیجه شبکه‌های عصبی لایه‌هایی دارند که برای ایجاد تعادل در سرعت یادگیری، با سرعت‌های متفاوت یاد می‌گیرند. هنگامی که دامنه‌ی گرادیان‌ها جمع می‌شوند، ناپایداری در شبکه‌ها بیشتر اتفاق می‌افتد که باعث نتایج پیش‌بینی ضعیف می‌شود. درنتیجه، YOLOv3 خروجی‌هایی با نمرات اطمینان را به‌وسیله تولیدکردن یک بردار از جعبه‌های محصورکننده، در هر زمانی که ورودی به‌شکل تصویر یا ویدئو به‌شبکه داده شود، پیش‌بینی می‌کند.

۳-۲- معماری شبکه‌ی YOLOv3

این شبکه از سه شبکه‌ی، بک‌بون[۱۴] Darknet-53، شبکه نمونه‌برداری افزایشی و در آخر لایه‌های شناسایی که لایه‌های YOLO نامیده می‌شوند، تشکیل شده است که در مجموع ۳۳۳ لایه دارد. تعداد لایه‌های شناسایی ۱۰۶ لایه است که ۷۵ لایه‌ی آن کانولوشنی هستند. در این شبکه، لایه‌های ادغام، حذف شده و با لایه‌های کانولوشنی با گام[۱۵] ۲ جایگزین شده‌اند تا از تلفات در هنگام فرآیند ادغام جلوگیری کنند. گام در واقع تعداد پیکسل‌هایی است که نسبت به‌ماتریس ورودی تغییر مکان می‌یابند. شبکه‌ی بک‌بون ساختارهای پیشرفته‌ای شامل ۱) بلوک‌های باقی‌مانده[۱۶] ، که لایه‌های میان‌بر برای آسان‌تر کردن آموزش شبکه هستند، ۲) ساختار آغازین[۱۷]، که شامل هسته‌های کانولوشنی ۳×۳ و ۱×۱ برای کاهش هزینه‌های محاسباتی می‌باشند، ۳) لایه نرمال‌سازی دسته‌ای[۱۸] که باعث می‌شود تا یادگیری لایه‌ها در شبکه از یکدیگر مستقل باشند. YOLOv3 تشخیص را در سه مقیاس انجام می‌دهد. هسته‌های تشخیص ۱×۱ با سه اندازه یکتا و در سه موقعیت یکتا به‌نقشه‌های ویژگی در شبکه اعمال می‌شود. شکل هسته‌های تشخیص برابر $1 \times 1 \times (B \times (4 + 1 + C))$ است، که B تعداد جعبه‌های محصورکننده، ۴ تعداد ویژگی‌های جعبه‌های محصورکننده، ۱ نمایانگر اطمینان شی، و C تعداد کلاس‌ها می‌باشد. در YOLOv3 ابعاد تصویر ورودی با ابعاد ۳۲، ۱۶ و ۸ نمونه‌گیری می‌شود تا پیش‌بینی‌ها را به‌ترتیب در ابعاد ۳، ۲ و ۱ انجام دهد.



شکل ۲- تقسیم تصویر و پیش‌بینی جعبه‌های محصورکننده [۱۹]

تصویر ورودی با اندازه ۶۴×۶۴ به‌وسیله ۸۱ لایه‌ی اول نمونه‌گیری می‌شود. از آنجایی‌که ۸۱ لایه‌ی اول دارای گام ۳۲ هستند، لایه ۸۲-ام اولین تشخیص را با نقشه ویژگی ۱۳×۱۳ انجام می‌دهد که به‌علت استفاده از یک هسته ۱×۱ برای تشخیص، اندازه نقشه ویژگی ۱۳×۱۳×۲۵۵ می‌باشد که مسئول تشخیص اشیا بزرگ یعنی با مقیاس ۳ است. به‌دنبال اولین تشخیص، نقشه ویژگی ۷۹-امین لایه بعد از قرار گرفتن در برابر تعدادی از لایه‌های کانولوشن، به‌میزان دو برابر نمونه‌گیری می‌شود و ابعاد ۲۶×۲۶ حاصل شده، سپس با نقشه ویژگی لایه ۶۱-ام ترکیب می‌شود. ویژگی‌ها با قرار گرفتن نقشه ویژگی ترکیب‌شده در برابر تعدادی از لایه‌های کانولوشن ۱×۱، بهم پیوند می‌خورند. درنتیجه، لایه‌ی ۹۱-ام دومین تشخیص را با یک نقشه ویژگی ۲۶×۲۶×۲۵۵، انجام می‌دهد که مسئول تشخیص اشیا متوسط یعنی با مقیاس ۲ است. به‌دنبال دومین تشخیص، نقشه ویژگی لایه ۹۱-ام بعد از قرار گرفتن در برابر تعدادی از لایه‌های کانولوشن به‌میزان دو برابر، نمونه‌گیری می‌شود و ابعاد ۵۲×۵۲ نتیجه شده و سپس با نقشه ویژگی لایه ۳۶-ام ترکیب

می‌شود. ویژگی‌ها با قرار گرفتن نقشه ویژگی ترکیب‌شده در برابر تعدادی از لایه‌های کانولوشن ۱×۱، بهم پیوند می‌خورند. در نتیجه لایه‌ی ۱۰۶-ام سومین و آخرین تشخیص را با یک نقشه ویژگی ۵۲×۵۲×۲۵۵ انجام می‌دهد که مسئول تشخیص اشیا کوچک یعنی با مقیاس ۱ است. بنابراین عملکرد YOLOv3 نسبت به‌نسخه‌های قبلی خود در تشخیص اشیا کوچک بهتر می‌باشد. مزیت اصلی استفاده از مدل YOLOv3 این است که این الگوریتم تنها به‌یک راه انتشار روبه جلو در شبکه برای شناسایی چندین شی در یک تصویر ورودی، نیاز دارد. شبکه به‌صورت همزمان تمامی سناریوهای ممکن و اشیاء محتمل بر روی تصویر را آنالیز می‌کند و عمل پیش‌بینی چندین ویژگی را انجام می‌دهد که بسیار موثرتر از الگوریتم پنجره‌ی لغزنده است که به‌چندین تکرار برای پردازش یک تصویر نیاز دارد.

۳-۳- تابع زیان اصلاح‌شده در شبکه‌ی YOLOv3

تابع زیان اصلاح‌شده دارای سه بخش، خطای مکان‌یابی، خطای اطمینان، و خطای طبقه‌بندی است. خطای مکان‌یابی برابر با خطای مربع بین مختصات جعبه محصورکننده پیش‌بینی‌شده $(\hat{x}_i,\hat{y}_i,\hat{w}_i,\hat{h}_i)$ و مختصات درستی مبنا[۱۹] (x_i,y_i,w_i,h_i) است. درواقع، این خطا نشانگر میزان اشتباه مختصات پیش‌بینی‌شده در مقایسه با مختصات درستی مبنا است. خطای اطمینان با استفاده از خطای میانگین مربع محاسبه می‌شود و بیان می‌کند که آیا در جعبه‌ی محصورکننده‌ی پیش‌بینی‌شده، شی وجود دارد یا خیر. خطای طبقه‌بندی با استفاده از خطای اختلاف انتروپی[۲۰] محاسبه می‌شود، و نشان می‌دهد که کلاس پیش‌بینی‌شده در داخل جعبه‌ی محصورکننده چه میزان اشتباه است. این تابع زیان از دو عامل مقیاس λ_{coords} و λ_{noobj} برای پایدارسازی مدل استفاده می‌کند که در این مقاله، به‌ترتیب ۵ و ۰/۵ انتخاب شده‌اند [۵] که λ_{coords} برای افزایش زیان مختصات جعبه‌های محصورکننده پیش‌بینی‌شده و λ_{noobj} برای کاهش زیان جعبه‌های محصورکننده پیش‌بینی‌شده بدون شی استفاده می‌شوند. این تابع زیان مطابق معادله (۲) محاسبه می‌شود [۵].

$$L = L_{loc} + L_{cls} = \lambda_{\text{coords}} \sum_{i=0}^{S^2} \sum_{j=0}^{3} 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{\text{coords}} \sum_{i=0}^{S^2} \sum_{j=0}^{3} 1_{ij}^{obj} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \right] + \sum_{i=0}^{S^2} \sum_{j=0}^{3} 1_{ij}^{obj} \left(-\hat{C}_i \log C_i - (1 - \hat{C}_i) \log(1 - C_i)\right) + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{noobj} \left(-\hat{C}_i \log C_i - (1 - \hat{C}_i) \log(1 - C_i)\right) + \sum_{i=0}^{S^2} 1_i^{obj} \sum_{c \in classes} (-\hat{p}_i(c) \log p_i(c) - (1\ \hat{p}_i(c))(\log(1 - p_i(c)))) \quad (۲)$$

که 1_i^{obj} نشان می‌دهد که آیا در شبکه‌ی سلولی i، شی وجود دارد یا نه. 1_{ij}^{obj} نشان می‌دهد که آیا j-امین جعبه‌ی محصورکننده‌ی i-امین شبکه سلولی مسئول پیش‌بینی شی هست یا خیر. C_i میزان اطمینان i-امین سلول است. $\hat{C}_i$ میزان اطمینان پیش‌بینی‌شده است. $p_i(c)$ احتمال شرطی اینکه آیا i-امین سلول، شی متعلق به‌کلاس c دارد یا خیر. $\hat{p}_i(c)$ احتمال شرطی کلاس پیش‌بینی‌شده است. به‌این ترتیب در این مقاله، با این تابع زیان اصلاح‌شده، محاسبات شبکه‌ی YOLOv3 کاهش داده شده و درنتیجه با کاهش پارامترها سرعت آن نیز افزایش یافته است.

۴-۳- آموزش شبکه

اولین قدم، مقداردهی اولیه‌ی وزن‌ها است. ورودی‌ها با استفاده از پایگاه تصاویر و با تکرار، لایه به لایه به‌شبکه وارد می‌شوند. به‌منظور انتشار گرادیان‌ها به‌وزن‌های شبکه، زیان براساس خروجی شبکه محاسبه می‌شود. قدم نهایی به‌روز رسانی وزن‌های شبکه است. این عملیات تا زمانی که زیان از یک آستانه مشخص کمتر شود، تکرار می‌شود. در طول آموزش شبکه، با چالش‌های متفاوتی روبه‌رو می‌شویم. مورد اول بیش‌برازش و کم‌برازش می‌باشد که نشان‌دهنده‌ی خطر پیچیدگی مدل با پارامترهای زیاد است که باعث بیش‌برازش بر روی داده‌ها می‌شود و در نتیجه قابلیت تعمیم‌دهی در سایر صحنه‌ها کاهش می‌یابد. مورد بعدی مشکلات افزایش سرعت آموزش هنگام برخورد با شبکه‌های عصبی عمیق است. یک راه آسان برای مقابله با بیش‌برازش و کم‌برازش، انتخاب ۲۰٪ کل تصاویر پایگاه داده برای آموزش و انتخاب ۱۰٪ آن برای ایجاد یک زیرمجموعه‌ی جدید به‌نام داده‌های اعتبارسنجی[۲۱] می‌باشد [۲۰]. داده‌های اعتبارسنجی برای آموزش مدل استفاده نمی‌شوند، بلکه بعد از هرگام آموزش، برای بررسی عملکرد مدل بر روی داده‌های جدید استفاده می‌شوند و هر زمان که مدل بهتر از گام قبلی عمل کند، ذخیره می‌شوند. این کار منجر می‌شود که مدل بعد از آموزش، کامل بارگیری شود و عملکرد بهتری بر روی داده‌های ورودی که مدل آن‌ها را ندیده است، داشته باشد. برخی دیگر از راه حل‌های مشکل بیش‌برازش شامل، افزایش تصاویر پایگاه تصاویر، رگولاریزاسیون یعنی دادن جریمه به‌بردارهای وزن-دهی، توقف زودهنگام فرآیند آموزش و برون‌اندازی بعضی از نرون‌های شبکه به‌منظور کاهش پیچیدگی مدل، می‌باشد [۲۱].

سرعت آموزش را می‌توان با یک الگوریتم گرادیان نزولی[۲۲] بهبود داد و درنتیجه همگرایی را آسان کرد. گرادیان نزولی وزن‌ها را به‌روز رسانی کرده و وزن بعدی را محاسبه می‌کند. در این مقاله، برای شبکه‌ی YOLOv3، ابتدا پایگاه تصاویر مناسب از محیط شهری زاهدان جمع‌آوری شده است. پایگاه تصاویر حاشیه‌نویسی و برچسب‌گذاری شده و براساس تکنیک‌های معرفی شده توسعه و تقویت شده است تا تصاویر مناسب برای آموزش و آزمایش شبکه فراهم شوند. از تابع زیان اصلاح‌شده در مرجع [۵] برای بهبود مشکل خطای طبقه‌بندی نادرست، استفاده شده است. این خطا به‌طور مستقیم بر روی عمل تشخیص علائم ترافیکی تاثیر می‌گذارد، بنابراین سعی‌شده تا در تابع زیان اصلاح‌شده، این مورد بهبود یابد. از تمامی روش‌های توضیح داده شده برای مقابله با بیش‌برازش و کم‌برازش در راستای بهبود مشکلات در طول آموزش استفاده شده است. از الگوریتم بهینه‌سازی براساس گرادیان ترکیب‌شده با بردار گشتاور به‌دلیل مزایای آن، از جمله کارآمد بودن محاسباتی، نیاز به‌حافظه کم، مناسب بودن برای داده‌ها و پارامترهای زیاد، تفسیر شهودی پارامترهای ترکیبی و در نتیجه نیاز به‌تنظیم کم، استفاده شده که برای آموزش شبکه مناسب است و منجر به‌محاسبه‌ی نرخ یادگیری انطباقی برای هر پارامتر می‌شود. در حین آموزش، گرادیان با شتاب بالا تمایل به‌حرکت و فرار از نقاط محلی مینیمم را دارد. در طول هر تکرار، ربع حاصل‌ضرب گرادیان محلی با گشتاور به‌گرادیان اضافه می‌شود، که به‌آن بردار گشتاور گویند. به‌روزرسانی وزن‌ها به‌معنای کم کردن بردارهای گشتاور است. نرخ یادگیری هم با گذشت زمان تغییر می‌کند و برای پارامترهای مختلف، متفاوت است. پارامترهایی که به‌طور مرتب تغییر می‌کنند دارای نرخ یادگیری کمتری هستند، درحالی‌که بقیه‌ی پارامترها نرخ یادگیری بیشتری دارند. درنتیجه گشتاور را برای گرادیان‌های بزرگ کاهش و برای گرادیان‌های کوچک افزایش می‌دهند.

۵-۳- مدل شبکه‌ی YOLOv3-SPP

در این شبکه، سه مدول SPP تلفیق‌یافته در جلوی هدرهای[۲۳] تشخیص بین لایه‌های کانولوشن ۵ و ۶ در شبکه‌ی YOLOv3 ارتقایافته، پیاده‌سازی شده‌اند و شبکه‌ی حاصل YOLOv3-SPP نام گرفته است. تعداد کل لایه‌های این شبکه، ۳۴۲ لایه است. الگوریتم فرآیند شناسایی علائم ترافیکی با استفاده از این شبکه شامل، تقسیم تصویر ورودی به‌شبکه‌های S×S، ساخت k جعبه‌ی محصورکننده همراه با تخمین لنگر، سپس پیش‌بینی مختصات هر جعبه‌ی محصورکننده و کلاس ترانهاده‌ی آن می‌باشد. در ادامه، اگر $IoU_{pred}^{truth} > IoU_{thres}$، یعنی جعبه‌ی محصورکننده حاوی شی است، درغیر اینصورت شی در این جعبه نیست [۲۲]. همچنین، کلاس با بیشترین احتمال پیش‌بینی به‌عنوان کلاس مربوط به‌شی انتخاب می‌شود. در نهایت سرکوب غیربیشینه اعمال می‌شود تا جست‌وجوی محلی بیشینه به‌سمت فایق آمدن بر جعبه-ها و خروجی‌ها هدایت شود و به‌این ترتیب تصویر شناسایی شده ارائه می‌شود. معیار

IoU [۲۴]، میزان همپوشانی بین ناحیه‌ی تشخیص شبکه برای حضور شی با پاسخ صحیح مبنا[۲۵] را محاسبه کرده بر اجتماع این دو ناحیه تقسیم می‌کند. در نتیجه این معیار کیفیت پیش‌بینی‌های شبکه را نسبت به‌پاسخ‌های صحیح مبنا نشان می‌دهد. مدل YOLOv3-SPP از نمونه‌برداری با گام ۲ در لایه‌های کانولوشن و دریافت بهترین ویژگی‌ها در لایه‌ی ادغام بیشینه استفاده می‌کند.

۳-۶- مدل شبکه‌ی Tiny YOLOv3

مدل شبکه‌ی Tiny YOLOv3 نسخه‌ی کاهش‌یافته‌ی مدل شبکه‌ی YOLOv3 است که سریع‌تر ولی با دقت کمتر شناسایی را انجام می‌دهد چون در قسمت استخراج ویژگی ضعیف‌تر عمل می‌کند. کل لایه‌های این مدل، ۵۹ لایه است که فقط ۷ لایه‌ی آن کانولوشن هستند. ویژگی‌ها با تعداد کمی لایه‌ی کانولوشن ۱×۱ و ۳×۳ استخراج می‌شوند. به‌علاوه، Tiny YOLOv3 به‌جای لایه‌ی کانولوشن با گام ۲ استفاده شده در YOLOv3، از لایه‌ی ادغام برای کاهش ابعاد استفاده می‌کند. ساختار هر لایه‌ی کانولوشن در مدل شبکه‌ی Tiny YOLOv3 شامل کانولوشن ۲ بعدی، نرمال‌سازی دسته‌ای، و تابع Leaky ReLU می‌باشد که نظیر YOLOv3 است. در ضمن، این مدل مانند مدل YOLOv3 آموزش داده می‌شود و مقدار زیان بااستفاده از یک تابع زیان مشابه محاسبه می‌شود [۲۳].

۳-۷- تولید داده‌ی مصنوعی

در این مقاله، به‌منظور افزایش دقت شبکه در شناسایی اشیاء کوچک از ساده‌ترین تکنیک تولید داده‌ی مصنوعی استفاده شده است. در این تکنیک، یک کپی از هر شی در مکان اصلی آن ایجاد می‌کنیم. سپس، این کپی در مکان‌های مختلفی چسبانده می‌شود. با افزایش اشیاء کوچک در هر تصویر تعداد لنگرهای مطابق با آن افزایش می‌یابد. این موضوع تاثیر اشیا کوچک در محاسبه تابع زیان را در طول آموزش بهبود می‌بخشد. البته، پیش از چسباندن اشیاء کوچک در مکان‌های جدید، تبدیل‌های تصادفی نیز بر روی آنها اعمال می‌کنیم به‌این‌صورت که اشیا را ۲۰٪ تغییر اندازه و ۱۵ درجه می‌چرخانیم. درضمن تنها از اشیایی که مسدود نشده باشند استفاده می‌کنیم زیرا که چسباندن ماسک‌های تقسیم‌بندی جداگانه با قسمت‌های دیده نشده در بین آن‌ها اغلب باعث ایجاد تصاویر غیرواقعی‌تر می‌شود. به‌علاوه اطمینان حاصل شده است که اشیاء چسبانده شده با اشیا موجود در عکس هم‌پوشانی نداشته باشند و حداقل ۵ پیکسل هم از مرزها فاصله داشته باشند.

۳-۸- تکنیک یادگیری انتقالی

تکنیک یادگیری انتقالی[۲۶] روشی است که در آن ابتدا یک شبکه برروی یک پایگاه تصاویر بزرگتر پیش‌آموزش می‌بیند و سپس از این شبکه در پایگاه داده‌ی کوچک‌تر استفاده می‌شود. در این مقاله، شبکه‌های پیشنهادی ابتدا برروی پایگاه تصاویرcoco [۵]، که یکی از بزرگترین پایگاه‌های داده‌ی Open Source برای آموزش مدل‌های یادگیری عمیق است، آموزش داده شده‌اند و سپس از این شبکه‌ها در شناسایی تصاویر علائم ترافیکی در پایگاه جدید استفاده شده‌اند. پایگاه داده coco دارای هزاران تصویر از اشیاء برچسب‌گذاری شده است، لذا پیش‌آموزش بر روی چنین پایگاه داده‌ی عظیم برچسب‌گذاری‌شده باعث همگرایی سریع‌تر شبکه بر روی هر پایگاه داده‌ی آموزشی شده و باعث دستیابی به‌دقت بالاتری در این داده‌ها می‌شود.

۴- نتایج اعمال شبکه‌های پیشنهادی بر روی پایگاه جدید

در این بخش، ابتدا پایگاه جدید شامل علائم ترافیکی شهر زاهدان و معیارهای ارزیابی به‌منظور بررسی نتایج اعمال شبکه‌های پیشنهادی بر روی این پایگاه معرفی می‌شوند. سپس نتایج بر اساس معیارهای ارزیابی توضیح داده شده، تحلیل می‌شوند. در ادامه نیز نتایج مدل‌های متفاوت با یکدیگر مقایسه می‌شوند.

۴-۱- پایگاه جدید تصاویر علائم ترافیکی

پایگاه تصاویر این مقاله، ۴۰۰۰ تصویر از علائم ترافیکی در سطح شهر زاهدان هستند که با دوربین گوشی موبایل Xiaomi A2 Lite از درون ماشین درحال حرکت، برای شبیه‌سازی بهتر سناریوهای دنیای واقعی، تصویربرداری شده‌اند. استفاده از دوربین‌های حرفه‌ای باعث کیفیت بهتر تصاویر می‌شود و درنتیجه تشخیص علائم ترافیکی در عکس راحت‌تر انجام می‌شود ولی در عین‌حال زمان پردازش و تشخیص افزایش می‌یابد. بنابراین از دوربین گوشی معمولی برای گرفتن عکس‌هایی با کیفیت متوسط استفاده شده است. همچنین تصویربرداری از علائم ترافیکی متفاوت مانند علائم مسدودشده به‌وسیله اشیاء مختلف، علائمی که رنگ آن‌ها در طول زمان کمرنگ‌شده و یا ناخوانا هستند، و در روزهایی با آب و هوای متفاوت، شرایط نوری غیرمتوازن و ساعات مختلف شبانه‌روز انجام شده است. سپس، با استفاده از ابزارهای برچسب‌گذاری، این تصاویر، حاشیه‌نویسی شده‌اند. به‌علاوه، این تصاویر به‌وسیله‌ی تکنیک‌های افزایش داده مانند چرخش، انتقال، تغییر مقیاس، برش، تزریق نویز و تولید داده‌های مصنوعی به ۲۶۶۴۰ تصویر رسیده‌اند.

جدول ۱- مشخصات پایگاه جدید اصلی و افزایش‌یافته شامل تصاویر علائم ترافیکی شهر زاهدان-ایران

مشخصه	پایگاه جدید اصلی	پایگاه جدید افزایش‌یافته
تعداد کلاس	۲۴	۲۴
تعداد حاشیه‌نویسی	۴۰۰۰	۲۶۶۴۰
تعداد تصاویر	۴۰۰۰	۲۶۶۴۰
تصاویر حاشیه‌نویسی	تمامی تصاویر	تمامی تصاویر
ابعاد علامت‌ها	۳۰×۳۰ تا ۴۰۰×۴۰۰	۳۰×۳۰ تا ۴۰۰×۴۰۰
ابعاد تصاویر	۲۰۰۰×۲۰۰۰ تا ۴۰۰۰×۳۰۰۰	۴۸۰×۴۸۰ تا ۴۰۰۰×۳۰۰۰
شامل ویدئو هستند	خیر	خیر
کشور	ایران	ایران
اطلاعات اضافی	تصاویر حاشیه نویسی و برچسب‌گذاری شده‌اند. تصاویر از داخل ماشین در حال حرکت و در شرایط مختلف گرفته شده‌اند.	تمامی تصاویر حاشیه نویسی و برچسب‌گذاری شده‌اند. تصاویر از داخل ماشین در حال حرکت در شرایط مختلف گرفته شده‌اند. تعدادی از تصاویر به صورت مصنوعی تولید شده‌اند.

در نهایت پایگاه جدید اصلی و افزایش‌یافته شامل ۲۴ کلاس از علائم ترافیکی پرتکرار در سطح شهر زاهدان هستند که توزیع تصاویر در اکثر کلاس‌ها یکسان می‌باشد ولی بنابر میزان تکرار، در برخی کلاس‌ها تعداد بیشتری تصاویر وجود دارد. برچسب‌ها به‌فرمت فایل txt هستند که هر خط از این فایل از سمت چپ شامل شماره‌ی کلاس شی (۲۳-۰)، مختصات مرکز، عرض و طول جعبه‌ی محصورکننده‌ی شی می‌باشند.

مشخصات این پایگاه جدید اصلی و افزایش‌یافته در جدول ۱ خلاصه شده‌اند. نمونه‌هایی از تصاویر برچسب‌گذاری شده‌ی این پایگاه جدید اصلی در شکل ۳ نمایش داده شده‌اند.

شکل ۳- نمونه‌هایی از تصاویر برچسب‌گذاری شده در پایگاه جدید علائم ترافیکی

۴-۲- معیارهای ارزیابی

زمانی که یک مدل شبکه برای شناسایی اشیاء داریم، تنها مشاهده‌ی موارد شناسایی شده به‌صورت کیفی بدون بیان آن‌ها به‌صورت کمی، کافی نیست و باید معیارهای ارزیابی ارائه شوند تا کیفیت کلی تشخیص‌دهنده‌ها را تعیین کنند. پیش‌بینی‌های شبکه‌ی تشخیص‌دهنده به ۴ دسته تقسیم می‌شوند که عبارتند از: مثبت صحیح (TP)[۲۷] ، یعنی نشانگر شی تشخیص داده شده صحیح است، مثبت اشتباه (FP)[۲۸]، یعنی نشانگر شی تشخیص داده شده اشتباه است، منفی اشتباه (FN) [۲۹]، یعنی نشانگر شی تشخیص داده نشده است، و منفی صحیح (TN) [۳۰]، یعنی نشانگر محل‌هایی است که در آن شی وجود ندارد و این محل‌ها به‌درستی تشخیص داده شده‌اند. اکنون معیارهای دقت (Precision)، حساسیت (Recall) ، و دقت متوسط (AP) را طبق معادله‌های (۳)، (۴)، (۵) تعریف می‌کنیم. دقت نشان می‌دهد که چه مقدار از پیش‌بینی‌ها درست بوده‌اند. حساسیت نشان می‌دهد چه تعداد از موارد صحیح به‌درستی پیش‌بینی شده‌اند [۲۴].

$$\text{Precision}=\frac{TP}{TP+FP} \qquad (۳)$$

$$\text{Recall}=\frac{TP}{TP+FN} \qquad (۴)$$

برای محاسبه‌ی دقت متوسط، همانطوری که معادله‌ی (۵) نشان می‌دهد، مساحت ناحیه‌ی زیر منحنی Precision-Recall را به‌این‌صورت محاسبه می‌کنند که ابتدا این منحنی را با استفاده از درون‌یابی هموار کرده سپس، میانگین مقادیر Precision را به ازای یازده مقدار Recall شامل (0.0, 0.1,… 1.0) محاسبه می‌کنند [۲۵].

$$AP=\frac{1}{11}\sum_{r\in\{0.0,0.1,\ldots 1.0\}} P_r \qquad (۵)$$

در محاسبه‌ی میانگین دقت متوسط (mAP)، مقدار AP را برای هر کلاس حساب و با هم جمع و بر تعداد کلاس‌ها تقسیم می‌کنیم. اگر در این محاسبه فقط کلاس‌هایی را در نظر بگیریم که $0.5 \leq IoU \leq 1$ باشند میانگین دقت متوسط را mAP^{50} و اگر $0.5 \leq IoU \leq 0.95$ باشند میانگین دقت را mAP^{50-95} می‌نامیم. در ارزیابی پیش‌بینی‌های یک شبکه می‌توان از ماتریس اغتشاش نیز استفاده کرد که فرم ساده‌ی آن در جدول ۲ مشاهده می‌شود. در واقع، ماتریس اغتشاش، میزان پیش‌بینی‌های صحیح و نادرست یک مدل را نشان می‌دهد. سطرهای این ماتریس شامل مقادیر پیش‌بینی‌شده و ستون‌های آن مقادیر واقعی هستند و یا برعکس، لذا اعداد روی قطر اصلی این ماتریس، تعداد کلاس‌های درست را نشان می‌دهند. در بررسی عملکرد یک شبکه‌ی تشخیص‌دهنده، اگر در ماتریس اغتشاش، تمام اعداد غیر روی قطر اصلی صفر باشند، آن شبکه دارای دقت شناسایی حداکثر است.

جدول ۲- ماتریس اغتشاش

	کلاس پیش‌بینی		
		بله	خیر
کلاس واقعی	بله	TP	FN
	خیر	FP	TN

۴-۳-نتایج اعمال شبکه‌های پیشنهادی برروی پایگاه جدید اصلی و افزایش‌یافته

در بررسی عملکرد شبکه‌های پیشنهادی برروی دو پایگاه جدید اصلی و افزایش‌یافته، ابتدا ۲۰٪ کل تصاویر پایگاه، برای آموزش، و ۱۰٪ آن برای اعتبارسنجی و ۷۰٪ آن برای آزمایش انتخاب، پیش‌پردازش و نرمالیزه می‌شوند. سپس سه شبکه‌ی پیشنهادی، YOLOv3 اصلاح‌شده، YOLOv3 SPP و Tiny YOLOv3 که از پیش روی پایگاه coco آموزش دیده‌اند اکنون برروی هر دو پایگاه جدید اصلی و افزایش‌یافته، آموزش داده‌شده و آزمایش می‌شوند. نتایج این آموزش‌ها و آزمایش‌ها در جدول‌های ۳، ۴، ۵ و ۶ آورده شده‌اند. همان‌طور که دیده می‌شود هر سه شبکه مبتنی بر معیار ارزیابی دقت بر روی پایگاه جدید افزایش‌یافته شامل ۲۶۶۴۰ تصویر علائم ترافیکی، عملکرد بهتری دارند. از طرفی چون تابع زیان در مدل YOLOv3 از سه قسمت، زیان جعبه‌ی محصورکننده، شی و کلاس تشکیل شده است این موارد جداگانه در جدول‌های ۳ و ۴ بررسی شده‌اند. مدل Tiny YOLOv3 به‌دلیل تعداد لایه‌های کمتر نسبت به‌مدل YOLOv3 و YOLOv3-SPP از دقت پایین‌تری برخوردار است.

جدول ۳- نتایج آموزش شبکه‌های پیشنهادی برروی پایگاه جدید با ۴۰۰۰ تصویر علائم ترافیکی شهر زاهدان

	YOLOv3	YOLOv3-SPP	Tiny YOLOv3
زیان جعبه‌ی محصورکننده (٪)	۰/۰۱۶۵۵	۰/۰۱۶۴۸	۰/۰۲۱۶۲
زیان شی (٪)	۰/۰۰۴۰۷	۰/۰۰۴۰۱	۰/۰۰۱۵۱۵
زیان کلاس (٪)	۰/۰۱۶۴۵	۰/۰۱۵۱۴	۰/۰۱۸۴۲
دقت (٪)	۰/۸۶۲۰۷	۰/۸۳۶۲۷	۰/۷۶۳۵۳
حساسیت (٪)	۰/۹۵۵۰۱	۰/۹۵۱۸۷	۰/۸۷۶۰۲

میانگین دقت متوسط (٪) ۰/۵<IOU<۱	۰/۹۴۸۵۸	۰/۹۴۱۸۶	۰/۸۹۶۸۶

جدول ۴- نتایج آموزش شبکه‌های پیشنهادی برروی پایگاه جدید با ۲۶۶۴۳۰ تصویر علائم ترافیکی شهر زاهدان

	YOLOv3	YOLOv3-SPP	Tiny YOLOv3
زیان جعبه‌ی محصورکننده (٪)	۰/۰۰۹۲۳۸	۰/۰۰۹۱۵۱	۰/۰۱۷۷
زیان شی (٪)	۰/۰۰۳۵۳۹	۰/۰۰۳۴۶۵	۰/۰۱۹۲۳
زیان کلاس (٪)	۰/۰۰۹۴۹۳	۰/۰۰۸۵۹۴	۰/۰۱۹۱۳
دقت (٪)	۰/۹۷۳۷	۰/۹۸۳۶	۰/۷۹۹۹
حساسیت (٪)	۰/۹۷۶۲	۰/۹۷۶	۰/۷۵۸۹
میانگین دقت متوسط (٪) ۰/۵<IOU<۱	۰/۹۸۷	۰/۹۸۶۲	۰/۸۴۲۷

جدول ۵- نتایج آزمایش شبکه‌های پیشنهادی برروی پایگاه جدید با ۴۰۰۰ تصویر علائم ترافیکی شهر زاهدان

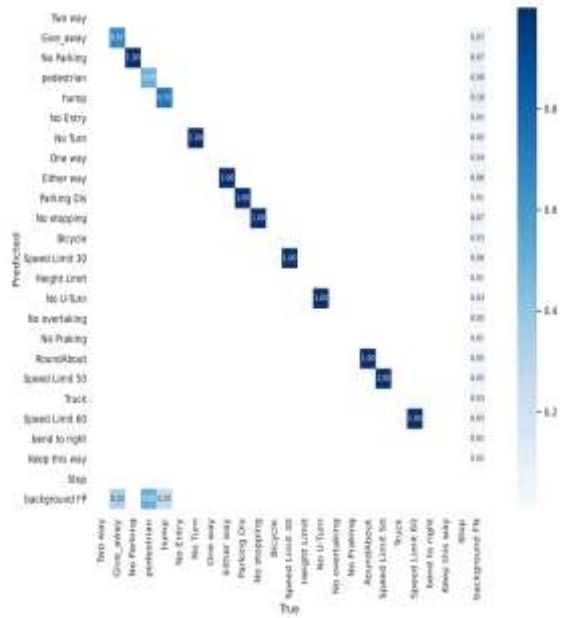
	YOLOv3	YOLOv3-SPP	Tiny YOLOv3
دقت (٪)	۰/۸۲۶	۰/۸۲۵	۰/۷۵۶
حساسیت (٪)	۰/۹۳۱	۰/۹۳۸	۰/۸۵۸
میانگین دقت متوسط (٪) ۰/۵<IOU<۱	۰/۹۲	۰/۹۲۷	۰/۸۸
میانگین دقت متوسط (٪) ۰/۵<IOU<۰/۹۵	۰/۸۲۱	۰/۸۰۷	۰/۶۶۵

جدول ۶- نتایج آزمایش شبکه‌های پیشنهادی برروی پایگاه جدید با ۲۶۶۴۳۰ تصویر علائم ترافیکی شهر زاهدان

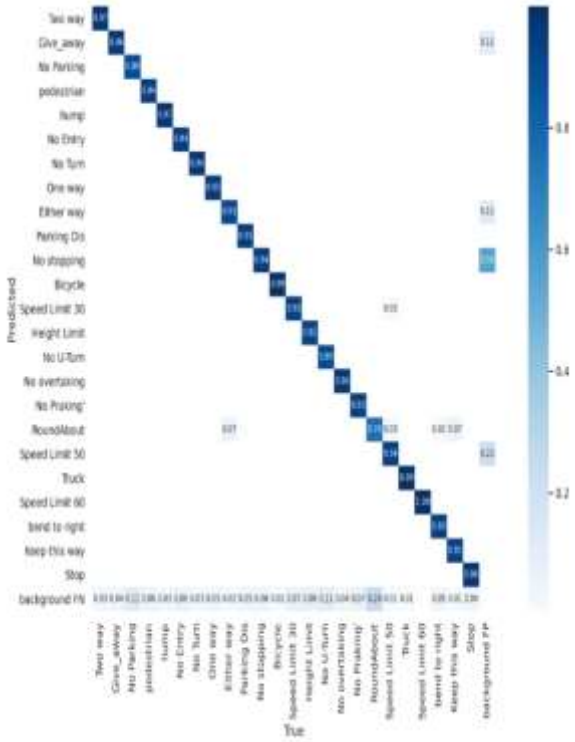
	YOLOv3	YOLOv3-SPP	Tiny YOLOv3
دقت (٪)	۰/۹۷۳	۰/۹۸۲	۰/۷۶
حساسیت (٪)	۰/۹۷۳	۰/۹۷۵	۰/۷۸۷
میانگین دقت متوسط (٪) ۰/۵<IOU<۱	۰/۹۸۸	۰/۹۸۸	۰/۸۴۵
میانگین دقت متوسط (٪) ۰/۵<IOU<۰/۹۵	۰/۸۸۶	۰/۹۰۹	۰/۵۲

در شکل ۴ تعدادی از نتایج پیش‌بینی مدل YOLOv3 برروی تصاویر آزمایشی پایگاه جدید اصلی به‌همراه جعبه‌ی محصورکننده و ضریب اطمینان، مشاهده می‌شوند. ماتریس اغتشاش مدل YOLOv3 به‌ترتیب برروی تصاویر آزمایشی پایگاه جدید اصلی شامل ۴۰۰۰ تصویر و پایگاه جدید افزایش‌یافته شامل ۲۶۶۴۰ تصویر در شکل‌های ۵ و ۶ نشان داده شده‌اند. با توجه به‌شکل ۵، ۵۰ درصد از پیش‌بینی‌های مدل YOLOv3 در کلاس pedestrain، صحیح (TP)، ۵۰ درصد اشتباه (FP)، و ۸ درصد هم به‌صورت منفی اشتباه (FN)، می‌باشند که این شبکه پیش‌بینی‌های مثبت اشتباه را به‌عنوان پس‌زمینه درنظر گرفته است و منفی‌های اشتباه را که باید شناسایی می‌کرده، تشخیص نداده است. البته بعضی از ردیف‌های ماتریس اغتشاش خالی می‌باشند که به‌علت تعداد کم تصاویر در آن کلاس است.

شکل ۴- تعدادی از نتایج پیش‌بینی مدل YOLOv3 برروی تصاویر آزمایشی به همراه جعبه‌ی محصورکننده و ضریب اطمینان



شکل ۵- ماتریس اغتشاش مدل YOLOv3 تصاویر آزمایشی پایگاه جدید اصلی



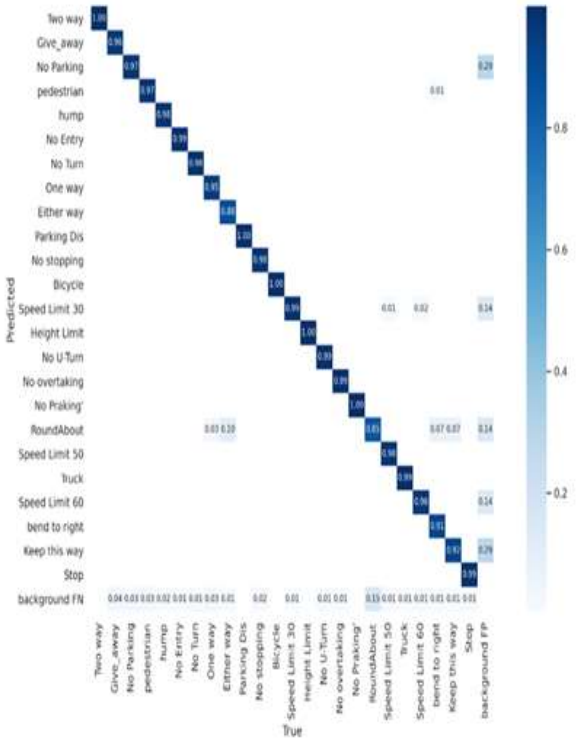
شکل ۶- ماتریس اغتشاش مدل YOLOv3 برروی تصاویر آزمایشی پایگاه جدید افزایش‌یافته

در جدول ۷، سرعت شناسایی مدل‌های پیشنهادی با ۷۳۰ تصویر از تصاویر آزمایشی پایگاه جدید اصلی و ۵۳۴۷ تصویر از تصاویر آزمایشی پایگاه جدید افزایش‌یافته نیز نمایش داده شده‌اند. برای مدل شبکه‌های YOLOv3-SPP و Tiny YOLOv3 نیز ماتریس اغتشاش برروی تصاویر آزمایشی پایگاه جدید افزایش‌یافته به‌ترتیب در شکل‌های ۷ و ۸ مشاهده می‌شوند. با توجه به‌ماتریس‌های اغتشاش هر سه مدل بر روی پایگاه جدید افزایش‌یافته، می‌توان دریافت که دقت‌های پیش‌بینی بالا هستند و مدل‌های پیشنهادی تقریباً تمام کلاس‌ها را درست پیش‌بینی کرده‌اند.

جدول۷- سرعت شناسایی مدل‌های پیشنهادی به ازای هر تصویر آزمایشی الف) از ۷۳۰ تصویر آزمایشی از پایگاه جدید اصلی ب) از ۵۳۴۷ تصویر آزمایشی از پایگاه جدید افزایش‌یافته

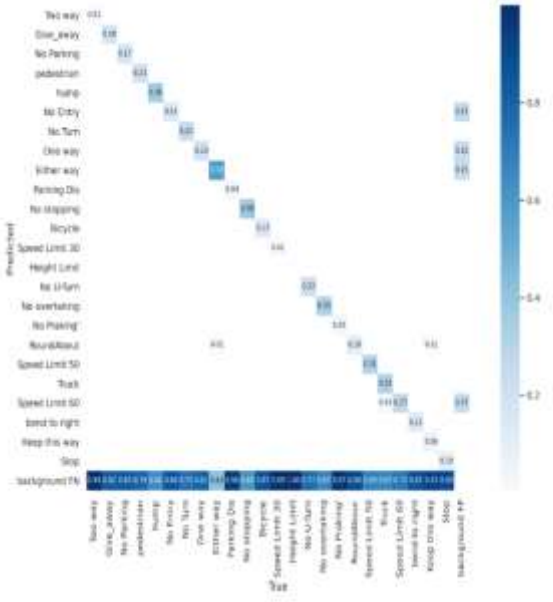
مدل	**سرعت** به ازای هر یک از ۷۳۰ تصویر آزمایشی از پایگاه جدید اصلی (میلی‌ثانیه)	**سرعت** به ازای هر یک از ۵۳۴۷ تصویر آزمایشی از پایگاه جدید افزایش‌یافته (میلی‌ثانیه)
YOLOv3	۸۰/۷	۱۶۷/۰
YOLOv3-SPP	۸۱/۸	۱۶۸/۵
Tiny YOLOv3	۷/۷	۱۵/۴

نمودار Precision-Recall مدل شبکه‌ی YOLOv3، YOLOv3-SPP و YOLOv3 Tiny برای محاسبه‌ی میانگین دقت متوسط (mAP) شناسایی برروی پایگاه جدید اصلی و پایگاه جدید افزایش‌یافته به‌ترتیب در شکل‌های ۹ و ۱۰ ، ۱۱ و ۱۲، ۱۳ و ۱۴ نمایش داده شده‌اند.

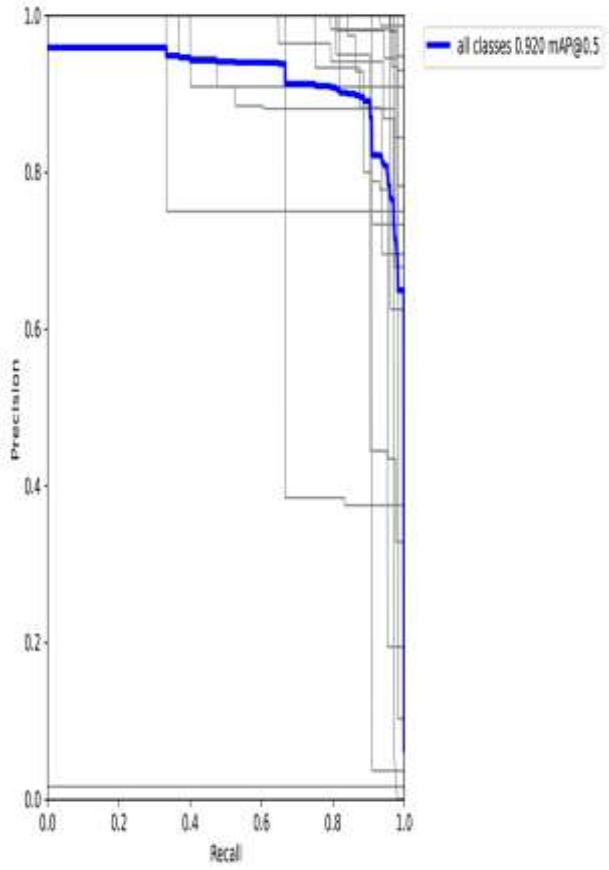


شکل ۷- ماتریس اغتشاش مدل YOLOv3-SPP برروی تصاویر آزمایشی پایگاه جدید افزایش‌یافته

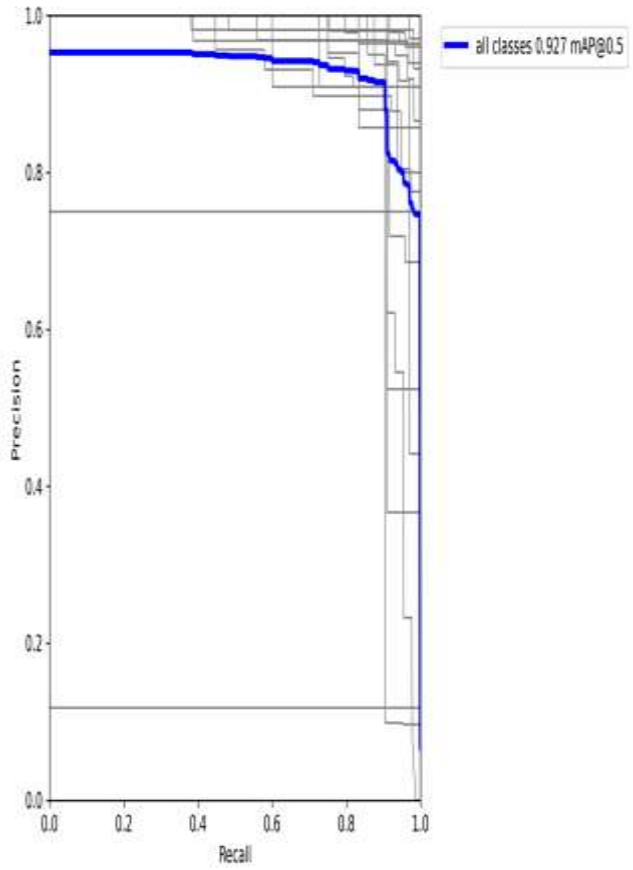
در این نمودارها خطوط خاکستری نمایانگر میانگین دقت متوسط هر کلاس و خط آبی نمایانگر میانگین دقت متوسط کلی است. همان‌طور که مشاهده می‌شود علی‌رغم تعداد کم تصاویر در پایگاه جدید اصلی، میانگین دقت متوسط برای هر سه مدل شبکه، مقدار بالایی به‌دست آمده است و با بررسی نمودار Precision-Recall این شبکه‌ها برروی تصاویر پایگاه جدید افزایش‌یافته، میانگین دقت متوسط برای دو مدل YOLOv3 و YOLOv3-SPP افزایش یافته است ولی برای مدل Tiny YOLOv3 بهبود چندانی نیافته است.



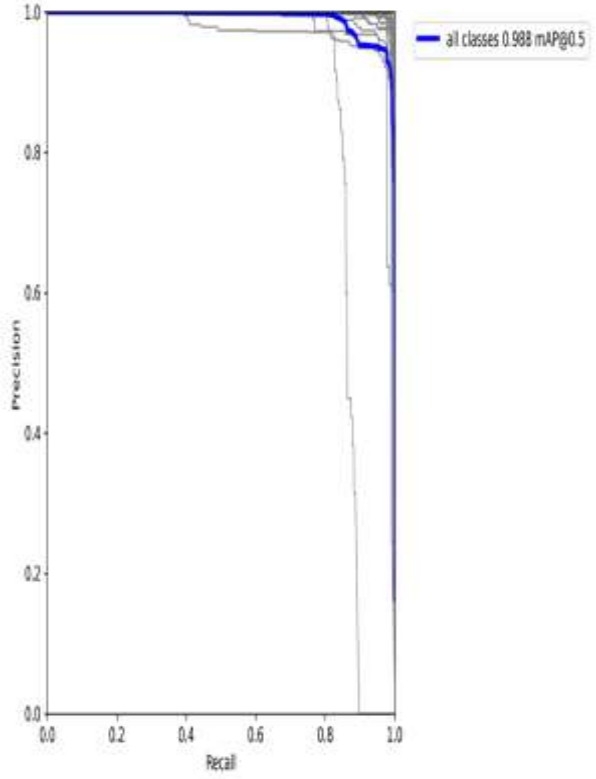
شکل ۸- ماتریس اغتشاش مدل Tiny YOLOv3 برروی تصاویر آزمایشی پایگاه جدید افزایش‌یافته



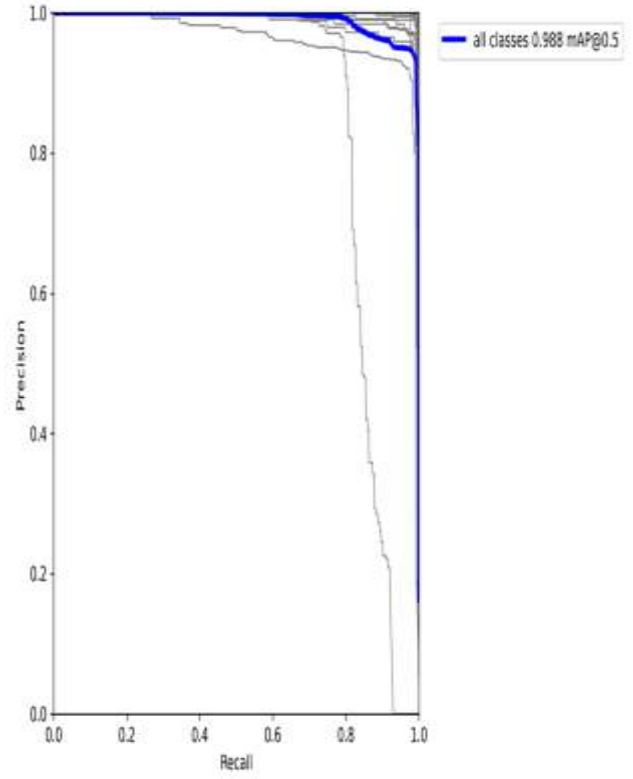
شکل ۹- نمودار Precision-Recall مدل شبکه‌ی YOLOv3 بر روی تصاویر پایگاه جدید اصلی



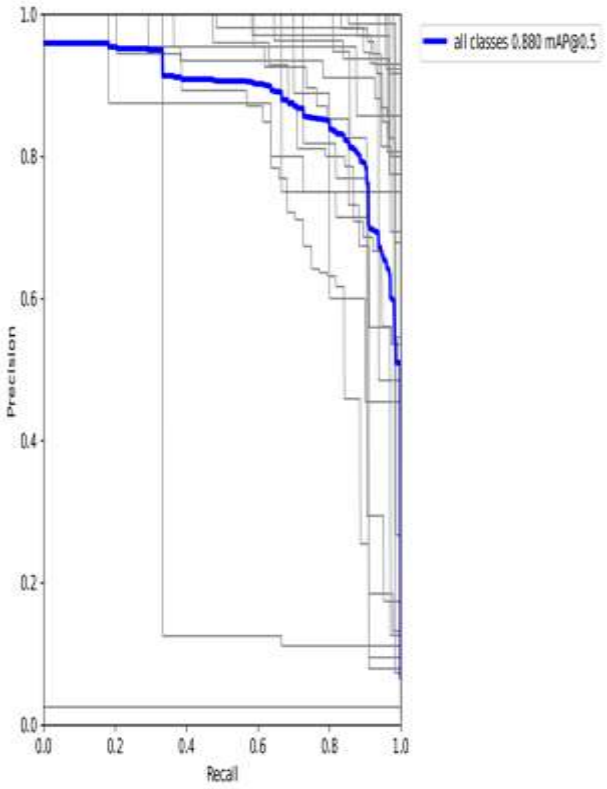
شکل ۱۱- نمودار Precision-Recall مدل شبکه‌ی YOLOv3-SPP بر روی تصاویر پایگاه جدید اصلی



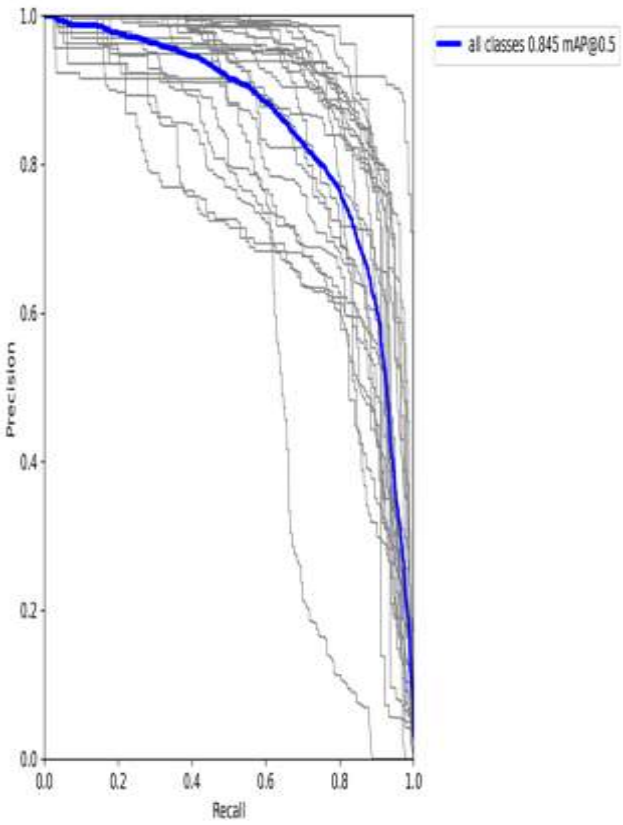
شکل ۱۰- نمودار Precision-Recall مدل شبکه‌ی YOLOv3 بر روی تصاویر پایگاه جدید افزایش‌یافته



شکل ۱۲- نمودار Precision-Recall مدل شبکه‌ی YOLOv3-SPP بر روی تصاویر پایگاه جدید افزایش‌یافته



شکل ۱۳- نمودار Precision-Recall مدل شبکه‌ی Tiny YOLOv3 بر روی تصاویر پایگاه جدید اصلی



شکل ۱۴- نمودار Precision-Recall مدل شبکه‌ی Tiny YOLOv3 بر روی تصاویر پایگاه جدید افزایش‌یافته

۵-نتیجه‌گیری

ابتدا شبکه‌ی تک‌سطحی YOLOv3 به‌منظور شناسایی علائم ترافیکی مورد بررسی قرار گرفت. این شبکه شامل یک شبکه‌ی Darknet-53 و به‌دنبال آن یک شبکه‌ی تشخیص می‌باشد. شبکه‌ی Darknet-53 یک شبکه‌ی پیچیده با ۴۰۵۴۹۲۱۶ پارامتر است که با این تعداد پارامتر دقت بهتری را فراهم می‌کند. مدل شبکه‌ی YOLOv3 دارای پیش‌بینی چندمقیاسی است که باعث بهبود مدل در تشخیص اشیا کوچک شده است. ولی از آجایی‌که، درکل، مدل‌های تک‌سطحی به‌نسبت مدل‌های دو سطحی سرعت بیشتری دارند اما دقت آن‌ها کمتر است در این مقاله سعی‌شده است که از شبکه‌ی YOLOv3 ارتقایافته و دو هم‌خانواده‌های آن در شناسایی علائم ترافیکی، استفاده شود. فرآیند آموزش شبکه به‌همراه مشکلات و راه‌حل‌های آن بیان شد. سپس هر سه مدل شبکه‌ی پیشنهادی، شامل YOLOv3 ارتقایافته، YOLOv3-SSP و Tiny YOLOv3 با استفاده از دو پایگاه جدید اصلی و افزایش‌یافته آموزش و آزمایش شدند. پایگاه جدید اصلی شامل ۴۰۰۰ تصویر عکسبرداری‌شده از شهر زاهدان و پایگاه جدید دوم شامل تصاویر پایگاه اصلی و تصاویر حاصل از تکنیک‌های افزایش داده و تولید داده مصنوعی به‌تعداد ۲۶۶۴۰ تصویر است. نزدیکترین پایگاه تصاویر موجود علائم ترافیکی، پایگاه تصاویر GTSDB است که مدل YOLOv3 بر روی این پایگاه به‌میانگین دقت متوسط ۹۲/۲٪ رسیده است درحالی‌که مدل YOLOv3 در این مقاله بر روی پایگاه تصاویر جدید افزایش‌یافته به‌میانگین دقت متوسط ۹۸/۸ ٪ رسیده که افزایش دقت را ۶/۶ درصد نشان داده است. به‌علت وجود لایه‌های زیاد در مدل YOLOv3 عملیات آموزش با پردازنده گرافیکی Nvidia GTX 1660 ti در ۱۷/۲ ساعت انجام شد. این زمان برای مدل YOLOv3-SSP، ۱۷/۴ ساعت و برای مدل Tiny YOLOv3 ، ۲ ساعت به‌دست آمد که چون آموزش برخط انجام نمی‌شود طولانی‌بودن این زمان‌ها اهمیتی ندارد. مهم زمان آزمایش این شبکه‌ها است که بیشترین زمان، ۱۶۸/۵ میلی‌ثانیه برای آزمایش شبکه‌ی YOLOv3-SSP برروی ۵۳۴۷ تصویر آزمایشی از علائم ترافیکی پایگاه جدید افزایش‌یافته حاصل شد که نشانگر یک زمان بلادرنگ مناسب. است. البته زمان آموزش شبکه را می‌توان با حفظ دقت و عمق شبکه با کاهش تعداد پارامترهای قابل آموزش آن، کاهش داد و شبکه‌ی بهبود یافته‌تری ایجاد کرد. کاهش پارامترها هم‌چنین باعث کاهش مصرف منابع شده و شبکه را قابل استفاده در کاربردهای عملی می‌کند. نتایج آموزش و آزمایش شبکه-های پیشنهادی برروی هردو پایگاه جدید اصلی و افزایش‌یافته با استفاده از معیارهای دقت، حساسیت و میانگین دقت متوسط بررسی شدند. همچنین ماتریس‌های اغتشاش برای مدل‌های پیشنهادی محاسبه شد و معیارهای TP، TN، FP و FN برای هر کلاس نمایش داده شدند. همچنین به‌علت عملکرد بهتر مدل‌های آموزش دیده بر روی پایگاه جدید افزایش‌یافته می‌توان از این مدل‌ها برای آزمایش داده‌های کاملا جدید نیز استفاده کرد. در مجموع، از سه مدل پیشنهادی در شناسایی علائم ترافیکی در دو پایگاه جدید اصلی و افزایش‌یافته، مدل‌های YOLOv3 و YOLOv3-SPP دقت بالاتر و مدل Tiny YOLOv3 سرعت بالاتری دارند. اگر بخواهیم دقت شبکه‌های پیشنهادی را در شناسایی علائم ترافیکی پایگاه جدید اصلی و افزایش‌یافته باز هم افزایش دهیم پیشنهاد می‌شود بااستفاده از تکنیک تطبیق حوزه[۳۱]، داده‌های آموزشی را به‌داده‌های آزمایشی نزدیکتر کنیم. تطبیق حوزه زیرشاخه‌ی یادگیری ماشین است که با سناریوهایی سر و کار دارد که یک شبکه‌ی آموزش دیده بر روی توزیع یک منبع، در زمینه‌ای با توزیع هدف متفاوت اما مرتبط

استفاده می‌شود. در واقع تطبیق حوزه باعث می‌شود تا میزان داده‌های آموزشی مورد نیاز برای انجام یک‌کار جدید بینایی ماشینی، به‌عملکرد سطح انسانی نزدیک‌تر شود. این کاری است که ما در ادامه‌ی این مقاله درحال انجام آن هستیم.

۶- مراجع

[1] H. Luo, Y. Yang, B. Tong, F. Wu, and B. Fan, "Traffic Sign Recognition Using a Multi-Task," *IEEE Trans. on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1100 - 1111, 2018.

[2] C. Liu, S. Li, F. Chang, and Y. Wang, "Machine Vision Based Traffic Sign Detection Methods: Review, Analyses and Perspectives," *IEEE Access,* vol. 7, pp. 2169-3536, 2019.

[3] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, "Augmentation for small object detection," *Computer Vision and Pattern Recognition,* arXiv.1902.07296v1, 2019.

[4] S. P. Rajendran, L. Shine, R. Pradeep, S. Vijayaraghavan, "Real-Time Traffic Sign Recognition using YOLOv3 based Detector," in *10th International Conference on Computin, Communication and Networking Technologies*, Kanpur, India, ID. 209696024, 2019.

[5] J. Wan, J., W. Ding, H. Zhu, et al., "An Efficient Small Traffic Sign Detection Method Based on YOLOv3," *Journal of Signal Processing Systems,* Vol. 93, pp. 899-911, 2021.

[6] C. Dewi, R. C. Chen, and S. K. Tai, "Evaluation of Robust Spatial Pyramid Pooling Based on Convolutional Neural Network for Traffic Sign Recognition System," *Electronics,* Vol. 9, No. 6, p. 889, 2020.

[7] A. Avramović, D. Sluga, D. Tabernik, D. Skočaj, V. stojnić, and N. llc, "Neural-Network-Based Traffic Sign Detection and Recognition in High-Definition Images Using Region Focusing and Parallelization," *IEEE Access,* Vol. 8, pp. 189855 - 189868, 2020.

[8] J. A. Khan, Y. Chen, Y. Rehman, and H. Shin, "Performance enhancement techniques for traffic sign recognition using a deep neural network," *Multimedia Tools and Applications,* Vol. 79, pp. 20545–20560, 2020.

[9] Q. Xu, R. Lin, H. Yue, H. Huang, Y. Yang and Z. Yao, "Research on Small Target Detection in Driving Scenarios Based on Improved Yolo Network," *IEEE Access,* vol. 8, pp. 27574-27583, 2020.

[10] M. Shahud, J. Bajracharya, P. Praneetpolgrang and S. Petcharee, "Thai Traffic Sign Detection and Recognition Using Convolutional Neural Networks," in *22nd International Computer Science and Engineering Conference (ICSEC)*, Chiang Mai, Thailand, 2018.

[11] Zhe Zhu; Dun Liang; Songhai Zhang; Xiaolei Huang; Baoli Li; Shimin Hu, "Traffic-Sign Detection and Classification in the Wild," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA , 2016.

[12] Branislav Novak, Velibor Ilić, Bogdan Pavković, "YOLOv3 Algorithm with additional convolutional neural network trained for traffic sign recognition," in *in Zooming Innovation in Consumer Technologies Conference (ZINC)*, Novi Sad, 2020.

[13] W. N. Mohd-Isa, M. S. Abdullah, M. Sarzil, J. Abdullah, A. Ali, and N. Hashim, "Detection of mlaysian traffic signs via modiffied YOLOv3 algorithm," in *Int. Conf. on Data Analytics for Business and Industry: Way Towards a Sustainable Economy*, Sakheer, Bahrain, 2020.

[14] S. Kheirinejad, N. Riaihi, and R. Azmi, "Persian text based traffic sign detection with convolutional neural network: a new dataset," in *10th Int. Conf. on Computer and Knowledge Engineering*, Mashhad, Iran, 2020,

[15] Z. Lin, T. Li, and T. Ding, "A multi-scale network-based for the YOLOv3 small target detection," in *2nd Int. Symposium on Computer Engineering and Intelligent Communications*, Nanjing, China, 2021.

[16] G. S. R. Nath, J. Acharjee, and S. Deb, "Traffic sign recognition and distance estimation with YOLOv3 model," in *Int. Conf. on Innovation and Intelligence for Informatics, Computing, and Technologies*, Zallaq, Bahrain, 2021.

[17] D. Mijic, M. Brisinello, M. Vranjes, and R. Grbic, "Traffic sign detection using YOLOv3," in *IEEE Int. Conf. on Consumer Electronics*, Berlin, Germany, 2020.

[18] P. Tumas, and A. Serackis, "Automated image annotation based on YOLOv3," in *IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering*, 2018.

[19] Sairaj Neelam, "Understanding YOLO and implementing YOLOv3 for object detection," https://medium.com/@sairajneelam/understanding-yolo-and-implementing-yolov3-for-object-detection-5f1f748cc63a

[20] C. Shorten, and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of *Big Data*, vol. 6, no. 60, pp. 1-48, 2019.

[21] S. Tripathi, S. Chandra, A. Agrawal, A. Tyagi, J. M. Rehg and V. Chari, "Learning to generate synthetic data via compositing," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, arXiv: 1904.05475, 2019.

[22] P. Shaotong, L. Dewang, M. Ziru, L. Yunpeng, and L. Yonglin, "Location and identification of insulator and bushing based on YOLOv3-SPP algorithm," IEEE Int. Conf. on Electrical Engineering and Mechatronics Technology, 2021.

[23] C. B. Murthy, and M. F. Hashmi, "Real time pedestrain detection using robust enhanced Tiny-YOLOv3," IEEE 17th India Council Int. Conf., 2020.

[24] C. Dewi, R. C. Chen, Y. T. Liu, X. Jiang, and K. D. Hartomo, "YOLO V4 for advanced traffic sign recognition with synthetic training data generated by various GAN," *IEEE Access*, vol.9, 2021.

[25] Jonathan Hui, "mAP for object detection," 2018. Jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173.

مائده نارویی در سال ۱۳۹۵ مدرک کارشناسی خود را در رشته مهندسی برق-الکترونیک از دانشگاه سیستان و بلوچستان اخذ کرد. سپس در سال ۱۴۰۰ با مدرک کارشناسی‌ارشد مهندسی مخابرات سیستم از همین دانشگاه فارغ‌التحصیل شد. ایشان تاکنون با تیم تحقیقاتی دکتر فرحناز مهنا در دانشگاه سیستان و بلوچستان مشغول می‌باشد.

فرحناز مهنا در سال ۱۳۶۷ مدرک کارشناسی مهندسی برق-الکترونیک را از دانشگاه سیستان و بلوچستان اخذ کرد. در سال ۱۳۷۱ با مدرک کارشناسی‌ارشد مهندسی برق-الکترونیک از دانشگاه تهران فارغ‌التحصیل شد. در سال ۱۳۸۱ مدرک دکتری مهندسی برق-پردازش‌داده را از دانشگاه ساری انگلستان دریافت کرد. سپس، دوره پسادکتری را در مرکز پردازش بینایی، گفتار و سیگنال در دانشگاه ساری انگلستان تا ۱۳۸۳ ادامه داد. ایشان هم‌اکنون دانشیار گروه مهندسی مخابرات دانشگاه سیستان و بلوچستان است و تحقیقاتش در زمینه‌های بینایی ماشین، شبکه‌های عصبی عمیق و هوش مصنوعی می‌باشند.

[1] Machine Learning
[2] Deep Learning
[3] Convolutional Neural Network
[4] Spatial Pyramid Pooling (SPP)
[5] Region of Interest (ROI)
[6] High Definition
[7] Graphics Processing Unit
[8] Anchor Box
[9] Attention Mechanism
[10] Open Source
[11] Benchmark
[12] Bounding Box
[13] Non-maximum Suppression
[14] Backbone
[15] Stride
[16] Residual Blocks
[17] Inception Structure
[18] Batch Normalization
[19] Ground Truth
[20] Cross Entropy

[21] Validation Data
[22] Gradient Descent
[23] Headers
[24] Intersection over Union
[25] Ground Truth
[26] Transfer Learning
[27] True Positive
[28] False Positive
[29] False Negative
[30] True Negative
[31] Domain Adaption