Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال دوم، شماره دوم، پاییز و زمستان ۱۴۰۳، صفحات ۱۲۴ - ۱۱۱

تشخیص نوع خودرو به‌منظور کنترل هوشمند عبور و مرور خودروها توسط یادگیری عمیق

مرتضی احساندوست[۱]، عاطفه سلیمی شهرکی[۲*]، محمد روح ا... یزدانی[۳]، محمد لعلی دستجردی[۱]

[۱]دانشجوی کارشناسی ارشد رشته مهندسی برق، واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران

[۲]استادیار رشته مهندسی برق، واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران

[۳]دانشیار رشته مهندسی برق، واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران

چکیده

با افزایش خودروها در سراسر جهان و توسعه فناوری، سیستم‌های حمل‌ونقل هوشمند به‌عنوان یکی از راه‌حل‌های کلیدی برای کنترل ترافیک و افزایش ایمنی راه‌ها مطرح‌شده‌اند. یکی از بخش‌های مهم این سیستم‌ها، تشخیص نوع خودرو است. در این مقاله از ترکیب مدل‌های EfficientNet-B0 و YOLO-V11 جهت تشخیص نوع خودرو استفاده‌شده است. در این مدل، EfficientNet-B0 به‌عنوان ستون فقرات مدل YOLO-V11 به‌کاررفته است. این مدل، ویژگی‌های تصاویر را استخراج کرده و آن‌ها را به مدل YOLO اعمال می‌کند تا مکان و نوع خودرو را به‌طور دقیق تشخیص دهد. برای ارزیابی عملکرد مدل پیشنهادی از مجموعه داده تصویری BVMMR که شامل بیش از ۵۰۰۰ تصویر از انواع خودروهای ایرانی است، استفاده‌شده است. مدل مورداستفاده با زبان برنامه‌نویسی پایتون نوشته‌شده و در محیط کولی (Colab) اجراشده است. نتایج اجرای کدها نشان می‌دهد که مدل پیشنهادی دارای مقدار میانگین دقت در هم‌پوشانی پنجاه‌درصد (mAP50) برابر با ۹۹/۳٪ و میانگین دقت در هم‌پوشانی بیشتر از پنجاه‌درصد (mAP50-95) برابر با ۹۸/۳٪ است. این مدل، ازنظر سرعت پردازش نیز توانسته است با میانگین زمان ۰/۲ میلی‌ثانیه برای پیش‌پردازش، ۲/۸ میلی‌ثانیه برای استنتاج و ۲/۵ میلی‌ثانیه برای پس‌پردازش در هر تصویر، عملکرد مطلوبی را ارائه دهد.

مقاله پژوهشی

تاریخ دریافت:
۱۴۰۳/۱۲/۲۱
تاریخ پذیرش:
۱۴۰۴/۲/۶

کلیدواژه‌ها:
بینایی رایانه، سیستم حمل‌ونقل هوشمند، یادگیری عمیق، YOLO، Efficient Net

نویسنده‌مسئول:
a.salimi@khuisf.ac.ir

: 10.22034/ABMIR.2025.22898.1112

E-ISSN: 2821-2037

۱- مقدمه

تشخیص نوع وسایل نقلیه بخشی از سیستم‌های حمل‌ونقل هوشمند (ITS)[1] است [۱]. شناسایی دقیق وسایل نقلیه و ردیابی آن‌ها توسط دوربین‌های نصب‌شده در اتوبان‌ها، باعث افزایش نظارت بر ایمنی شهری و مدیریت ترافیک می‌شود [۲]. تشخیص وسایل نقلیه می‌تواند زمان‌بندی چراغ‌های راهنمایی و رانندگی را براساس حجم ترافیک تنظیم کند و باعث کاهش تأخیر ناشی از ازدحام و تخلفات ترافیکی شود [۳]. تشخیص نوع وسیله نقلیه می‌تواند خدماتی را برای نظارت بر وضعیت جاده و سیستم‌های جمع‌آوری عوارض بزرگراه‌ها ارائه دهد [٤]. سیستم‌های رانندگی خودمختار[2] نیز شامل ادراک محیطی، تصمیم‌گیری رفتاری و برنامه‌ریزی و کنترل حرکت هستند. تشخیص دقیق وسایل نقلیه برای درک محیط آن‌ها ضروری است [٥]. استفاده از روش‌های مبتنی بر یادگیری عمیق مانند مدل YOLO[3]، باعث بهبود دقت و سرعت تشخیص خودروها می‌شود.

باکرسی در [٦] از مدل YOLO-V8 برای تشخیص وسایل نقلیه در تصاویر هوایی پهپادهای خودمختار استفاده کرده و با جمع‌آوری داده‌های متنوع از زوایا و ارتفاعات مختلف، تعمیم‌پذیری مدل را بهبود داده است. در آزمایش‌ها، دو نسخه YOLO-V8N و YOLO-V8X در شرایط نوری و محیطی متنوع مقایسه شدند. نتایج نشان داد که YOLO-V8N با دقت[4] ۸۳٪ و فراخوانی[5] ۷۹٪ برای کاربردهای بلادرنگ مناسب‌تر است. درحالی‌که YOLO-V8X با دقت ۹٦٪ و فراخوانی ۸۹٪ عملکرد بهتری دارد.

یوسف و همکاران در [۷] رویکرد جدیدی را برای تشخیص و ردیابی وسایل نقلیه ارائه دادند. در این روش، ابتدا موقعیت مکانی تصاویر تعیین و نویز آن‌ها حذف می‌شود. سپس با تفکیک تصویر و استفاده از الگوریتم‌هایی مانند جنگل تصادفی[6]، تطبیق الگو و شمارش با هیستوگرام گرادیان‌های جهت‌دار (HOG[7]) خودروها شناسایی، شمارش و ردیابی می‌شوند. درنهایت، مسیر حرکت خودروها با مقایسه مکان آن‌ها در فریم‌های مختلف پیش‌بینی می‌شود. از مجموعه داده خودروهای هوایی[8] [۸] و مجموعه داده وسایل نقلیه هوایی بدون سرنشین چندوجهی برای نظارت بر ترافیک در ارتفاع پایین (AU-AIR)[9] [۹] برای ارزیابی رویکرد پیشنهادی استفاده‌شده و به ترتیب دقت‌های ۸۷٪، ۹۲٪، ۸٤٪ و ۸۸٪ به‌دست‌آمده است.

زوهیب و همکاران در [۱۰] رویکردی را برای تشخیص وسایل نقلیه اضطراری با ترکیب داده‌های صوتی و تصویری ارائه دادند. این روش با استفاده از شبکه طیف موقتی مبتنی بر توجه (ATSN)[10] برای تشخیص صدای آژیر و معماری فیوژن فضایی چندسطحی (MLSF-YOLO)[11] برای تشخیص بصری وسایل نقلیه، دقت و زمان پاسخ را بهبود می‌بخشد. نتایج نشان داد که این سیستم با دقت ۹٦/۱۹٪ و میزان خطای ۳/۸۱٪ عملکرد بهتری نسبت به روش‌های سنتی دارد. نتایج نشان می‌دهد که مدل پیشنهادی توانسته است به‌خوبی وسایل نقلیه امدادی را شناسایی و به بهتر شدن سرعت واکنش و کمتر شدن حوادث در شهر کمک کند.

سونگ و همکاران در [۱۱] مدلی جدید برای شناسایی بهتر خودروها و کاهش هزینه‌های محاسباتی ارائه دادند. آن‌ها با بهره‌گیری از معماری Mamba_ViT [۱۲]، ویژگی‌های تصویر را در چند سطح مختلف و به‌صورت جداگانه بررسی کردند تا اطلاعات دقیق‌تری از صحنه به‌دست آورند و عملکرد مدل را بهبود بخشند. جهت ارزیابی مدل پیشنهادی از مجموعه داده شناسایی و ردیابی دانشگاه آلبانی (UA-DETRAC)[12] [۱۳]، استفاده‌شده است. نتایج نشان داد که دقت این مدل ۳/۲٪ بیشتر از مدل

[1] Intelligent Transportation Systems
[2] Autonomous driving systems
[3] You Only Look Once
[4] Precision
[5] Recall
[6] Random Forest Algorithm
[7] Histogram Of Oriented Gradients

[8] Aerial Cars
[9] A Multi-modal Unmanned Aerial Vehicle Dataset for Low Altitude Traffic Surveillance
[10] Attention-Based Temporal Spectrum Network
[11] Multi-Level Spatial Fusion YOLO
[12] University at Albany Detection and Tracking

YOLO-V8-tiny است و تنها از ٦٠٪ پارامترهای خود استفاده می‌کند.

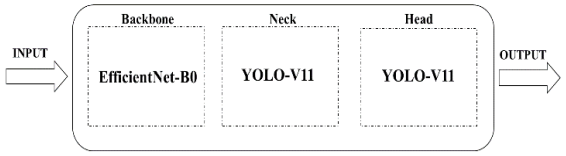
ژانگ و همکاران در [١٤] الگوریتمی به‌نام YOLO با آگاهی از بافت محلی و نمای کلی تصویر (LGA-YOLO[1]) را برای تشخیص وسایل نقلیه معرفی کردند. این مدل شامل دو بلوک افزایش میدان دید و تقویت ویژگی‌های محلی (MLKM)[2] و استخراج اطلاعات زمینه‌ای کلی (DGAM[3]) است که باعث بهبود تمایز وسایل نقلیه از پس‌زمینه‌های پیچیده می‌شود. این مدل با استفاده از شبکه ترکیب ویژگی‌های سطح بالا و پایین، اشیای چندمقیاسی را بهتر شناسایی می‌کند. مجموعه داده‌های یونیکورن ویژه شناسایی اشیای کوچک (USOD)[4] [١٥]، تشخیص وسیله نقلیه در تصاویر هوایی (VEDAI)[5] [١٦] و تشخیص اشیا در تصاویر هوایی (DOTA[6]) [١٧] برای ارزیابی این مدل به‌کاررفته‌اند. نتایج ارزیابی نشان دادند که میانگین دقت متوسط در هم‌پوشانی پنجاه‌درصد (mAP50)[7] به‌ترتیب عبارت از ۹۳٪، ۸۰/۳٪ و ۷۸/۱٪ هستند.

پارک و همکاران در [١٨] از ترکیب مدل‌های YOLO-V8 و MobileNet جهت تشخیص و طبقه‌بندی وسایل نقلیه و پلاک آن‌ها استفاده کرده‌اند. این دو الگوریتم به‌ترتیب برای تشخیص اشیا در زمان واقعی و پیش‌پردازش کارآمد به‌کار می‌روند و می‌توانند در دستگاه‌هایی که منابع محدودی دارند (مانند تلفن همراه) کار کنند. نویسندگان از این کار برای مدیریت یک پارکینگ خصوصی استفاده نموده و به دقت‌های ٩٥/٧٦٪ و ٩٧/١٨٪ رسیده‌اند.

در این مقاله، از یک مدل ترکیبی کارآمد جهت شناسایی و دسته‌بندی انواع خودروها استفاده‌شده است. در این مدل ترکیبی، مدل EfficientNet-B0 [١٩] با مدل YOLO-V11[٢٠] ادغام شده است و عملکرد قابل قبولی را از خود نشان داده است. این مدل به‌صورت یکپارچه عمل می‌کند و هدف اصلی آن تشخیص انواع خودروها با کمترین خطا است. در بخش ۲ این مقاله روش پیشنهادی معرفی‌شده است. در این بخش، ابتدا مدل‌های YOLO-V11 و EfficientNet-B0 و نحوه ادغام آن‌ها ذکرشده است. در بخش ۳ مجموعه داده مورداستفاده در این مقاله معرفی‌شده و در بخش ٤ نیز نتایج ارزیابی مدل نشان داده‌شده است. درنهایت، در بخش ۵ یک نتیجه‌گیری کلی از این مقاله ارائه‌شده و پیشنهادهایی برای کارهای آینده بیان‌شده است.

۲- روش پیشنهادی

در این مقاله، جهت تشخیص و دسته‌بندی نوع خودرو از ترکیب مدل‌های EfficientNet-B0 و YOLO-V11 استفاده‌شده است. مدل EfficientNet-B0 جایگزین ستون فقرات[8] مدل YOLO-V11 شده تا دقت تشخیص را افزایش دهد و عملکرد مدل در شرایط مختلف را بهبود ببخشد. شکل (۱) بلوک دیاگرام روش پیشنهادی را نشان می‌دهد. این مدل ترکیبی [٢١] برمبنای مدل YOLO طراحی‌شده و به‌صورت یکپارچه عمل می‌کند. یعنی هر ۳ بخش ستون فقرات، گردن[9] و سر[10] مدل در یک بلوک قرارگرفته‌اند و با کمک یکدیگر عملیات موردنظر را انجام می‌دهند. قسمت ستون فقرات وظیفه استخراج ویژگی‌های داده‌ها را برعهده دارد. قسمت گردن نیز، ویژگی‌های استخراج شده جهت تشخیص نوع خودرو را پردازش کرده و آن‌ها را بهبود می‌بخشد. درنهایت، قسمت سر خروجی نهایی که شامل مکان و نوع خودرو است را تولید کرده و به خروجی شبکه منتقل می‌کند.



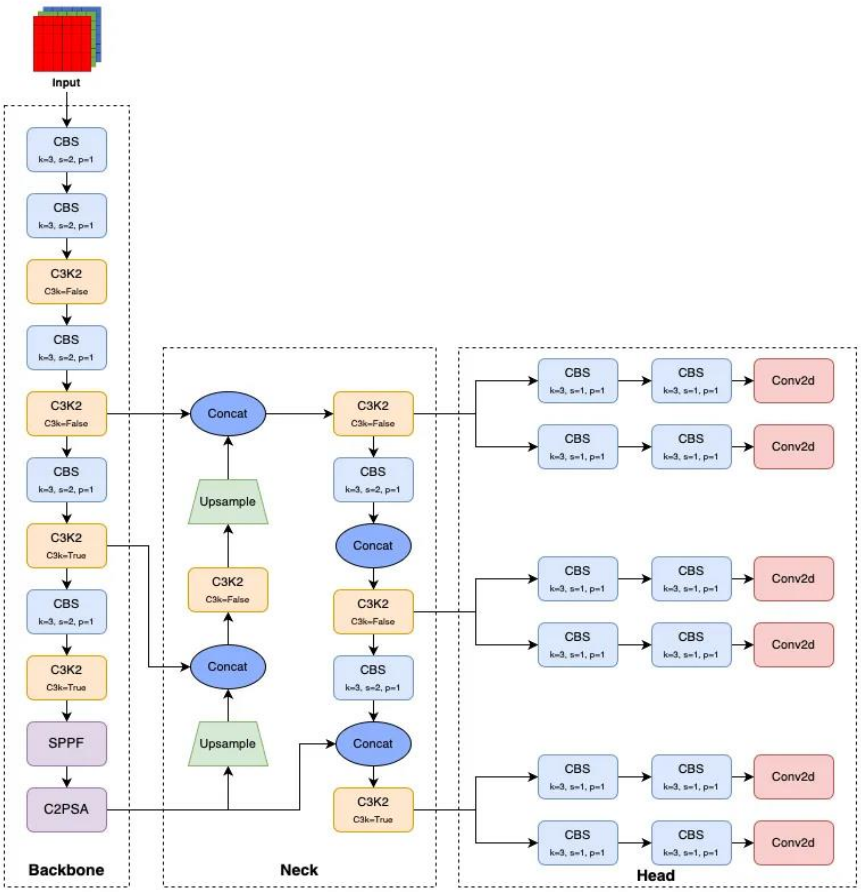
شکل (۱): بلوک دیاگرام روش پیشنهادی

۲-۱ مدل YOLO-V11

مدل YOLO-V11 در سال ۲۰۲٤ معرفی‌شده و جهت کلاس‌بندی، بخش‌بندی و تشخیص اشیا به‌کار می‌رود [٢٢]. در شکل (۲) بلوک

[1] Local And Global Aware YOLO
[2] Multi scale Large Kernel Local Aware Module
[3] Directional Global Context Aware Module
[4] Unicorn Small Object Dataset (USOD)
[5] Vehicle detection in aerial imagery (VEDAI)
[6] Dataset for object detection in aerial images (DOTA)
[7] Mean Average Precision at 50% Intersection over Union (mAP50)
[8] Backbone
[9] Neck
[10] Head

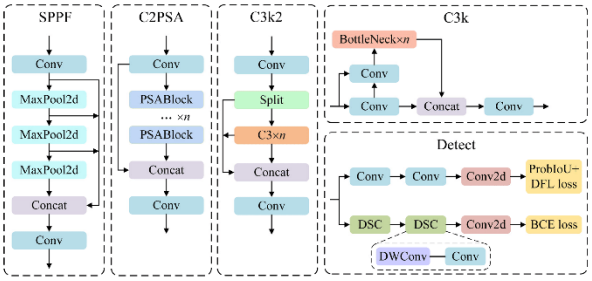
دیاگرام YOLO-V11 نشان داده‌شده است. ستون فقرات مدل ویژگی‌های تصاویر را از طریق بلوک‌های مختلفی استخراج می‌کند. بلوک کانولوشن، نرمال‌سازی دسته‌ای و تابع فعال‌سازی غیرخطی سیگموید (CBS)[1] به شبکه کمک می‌کند تا الگوهای پیچیده موجود در تصاویر را یاد بگیرد. بلوک میان مرحله‌ای جزئی با اندازه هسته ۲×۲ (C3k2)[2] با استفاده از لایه‌های گلوگاه[3]، ابعاد ویژگی‌ها را کاهش داده و آن‌ها را ترکیب می‌کند. این کار باعث می‌شود تا ویژگی‌های بیشتری استخراج شوند. بلوک ادغام هرم فضایی سریع (SPPF)[4] به شبکه کمک می‌کند تا تصاویر با اندازه‌های مختلف را بهتر مدیریت کند و ویژگی‌ها را از مقیاس‌های گوناگون استخراج کند. بلوک مکانیزم توجه فضایی نقطه‌ای با اشتراک‌گذاری جزئی میان مرحله‌ای (C2PSA)[5] با استفاده از مکانیزم توجه[6] ویژگی‌های مکانی تصویر را بهتر درک کرده و دقت تشخیص اشیا را افزایش می‌دهد. پس از استخراج ویژگی‌ها، قسمت گردن این ویژگی‌ها را ترکیب و تقویت می‌کند تا دقت تشخیص بهبود یابد. درنهایت، قسمت سر مختصات جعبه‌های محدودکننده[7]، امتیاز حضور اشیا[8] و کلاس‌های احتمالی آن‌ها را پیش‌بینی می‌کند.



شکل (۲): بلوک دیاگرام مدل YOLO-V11 [۲۲]

[1] Convolution, Batch Normalization, Sigmoid Linear Unit (CBS)
[2] Cross Stage Partial with kernel size 2 (c3k2)
[3] Bottleneck Layers
[4] Spatial Pyramid Pooling – Fast (SPPF)
[5] Cross-Stage Partial Point-wise Spatial Attention (C2PSA)
[6] Attention Mechanism
[7] Bounding Boxes
[8] Object Presence Score

شکل (۳) ساختار داخلی بلوک‌های فوق را نشان می‌دهد. بلوک SPPF از لایه‌های کانولوشن و MaxPool2d برای پردازش ویژگی‌ها در چندین سطح استفاده می‌کند تا مدل بتواند اشیا با اندازه‌های مختلف را تشخیص دهد. بلوک C2PSA با استفاده از مکانیزم توجه، بر بخش‌های مهم تصویر تمرکز کرده و اطلاعات اشیا را پردازش می‌کند. بلوک‌های C3k2 و C3k از لایه‌های کانولوشن و گلوگاه برای پردازش عمیق‌تر ویژگی‌ها استفاده نموده و نتایج حاصل‌شده را با یکدیگر ترکیب می‌کنند. این کار به تقویت ویژگی‌های استخراج شده کمک می‌کند. درنهایت، بلوک Detect مسئول شناسایی اشیا و نام کلاس آن‌ها است و این عمل را طریق دو مسیر پردازش موازی انجام می‌دهد. مسیر اول، دقت تشخیص و مکان‌یابی جعبه‌های محدودکننده اشیا را بهبود بخشیده و مسیر دوم، پیش‌بینی کلاس‌های اشیا را بهینه می‌کند. با استفاده از این بلوک‌ها، دقت نهایی YOLO-V11 افزایش می‌یابد.

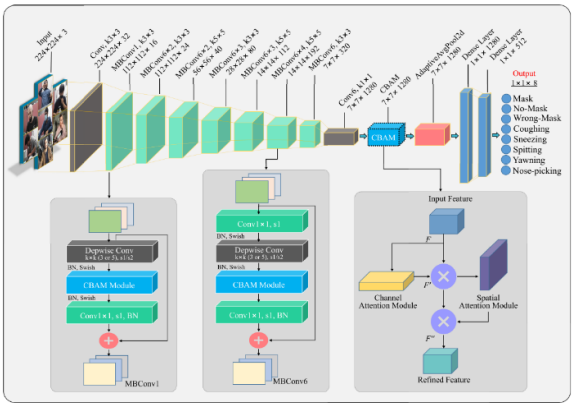


شکل (۳): معماری بخش‌های مختلف مدل YOLOV-11 [۲۳]

لازم به ذکر است که مدل YOLO-V11 نیز همانند مدل‌های قبلی خانواده YOLO، تشخیص اشیا را به‌صورت هم‌زمان و با سرعت بالایی انجام می‌دهد. این امر، باعث کاهش دقت مدل در تصاویر خیلی پیچیده می‌شود. جهت بهبود دقت مدل در مواجهه با تصاویر بسیار پیچیده می‌توان ستون فقرات مدل را برداشته و آن‌را با یک مدل از پیش آموزش‌دیده قوی‌تر جایگزین کرد. این مدل‌ها جهت استخراج ویژگی‌های تصاویر و دسته‌بندی کردن آن‌ها استفاده می‌شوند. برای انجام این کار، از مدل از پیش آموزش‌دیده EfficientNet-B0 استفاده‌شده است.

۲-۲ مدل EfficientNet-B0

مدل EfficientNet یک شبکه عصبی کانولوشنی است که توسطتان و لی [۱۹] طراحی شده و به‌صورت هم‌زمان، عرض و عمق شبکه و وضوح تصاویر ورودی به آن‌را مقیاس‌بندی می‌کند. پایه‌ای‌ترین نسخه این خانواده مدل EfficientNet-B0 است. در شکل (٤) بلوک دیاگرام این مدل رسم شده است. مشاهده می‌شود که در این مدل از تعداد زیادی بلوک MBConV جهت استخراج ویژگی‌ها استفاده‌شده است. ابتدا، لایه Conv1×1، تعداد کانال‌های ورودی را افزایش می‌دهد تا لایه‌های بعدی بتوانند ویژگی‌های بیشتری از داده‌ها را استخراج کنند. سپس، بلوک Depthwise Conv3×3 یک فیلتر کانولوشن ۳×۳ را به‌طور جداگانه روی هر کانال اعمال می‌کند که ازنظر محاسباتی بسیار کارآمد است. بعدازآن، بلوک فشرده‌سازی و تحریک (SE)[1] اهمیت هر کانال را تنظیم می‌کند تا شبکه بتواند روی ویژگی‌های مهم‌تر تمرکز کند. سپس، لایه Conv1×1 آخر، تعداد کانال‌ها را کاهش می‌دهد تا حجم محاسبات کم‌تر شود. درنهایت، اتصال باقی‌مانده[2] نیز با حفظ اطلاعات ورودی و انتقال مستقیم آن به لایه‌های بعدی، از بروز مشکلاتی مانند محو شدن گرادیان[3] (عدم به‌روزرسانی مناسب وزن‌ها) جلوگیری کرده و موجب پایداری بیشتر در فرآیند یادگیری شبکه می‌شود.



شکل (٤): بلوک دیاگرام مدل EfficientNet-B0 [٢٤]

مدل EfficientNet-B0 یک مدل سبک‌وزن و از پیش آموزش‌دیده است و می‌تواند ویژگی‌های مهم تصاویر را با دقت

[1] Squeeze-and-Excitation
[2] Residual Connection
[3] Vanishing Gradient

بسیار بالایی استخراج کرده و آن‌ها را دسته‌بندی کند. وقتی‌که این مدل در بخش ستون فقرات مدل YOLO-V11 قرار گیرد، سرعت و دقتش متعادل‌تر می‌شود. این ادغام مزایا و معایبی نیز دارد که در ادامه به آن اشاره می‌شود.

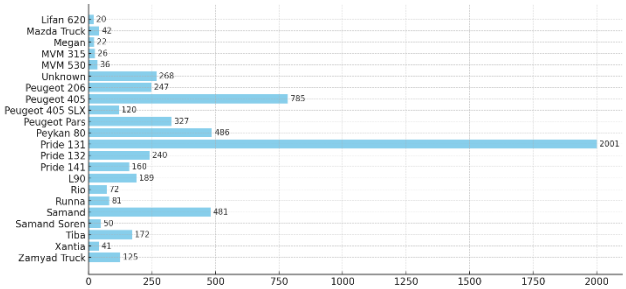
۲-۳ مزایا و معایب مدل ترکیبی

یکی از مهم‌ترین مزایای ادغام مدل EfficientNet-B0 در ستون فقرات مدل YOLO-V11 افزایش دقت تشخیص اشیا بدون افزایش حجم مدل است. EfficientNet-B0 با استفاده از تکنیک مقیاس‌بندی مرکب ویژگی‌های غنی‌تری از تصاویر را استخراج می‌کند و منجر به بهبود توانایی مدل در تشخیص اشیا با جزئیات بیشتر می‌شود. این ترکیب می‌تواند باعث بهبود عملکرد سیستم‌های نظارتی و امنیتی شود. همچنین، استفاده از این مدل ترکیبی باعث می‌شود که مدل در مواجهه با داده‌های مختلف انعطاف‌پذیرتر باشد. این ادغام معایبی نیز دارد که باید موردتوجه قرار گیرد. یکی از چالش‌های اصلی، سازگاری معماری این دو مدل با یکدیگر است. ساختار مدل EfficientNet جهت طبقه‌بندی تصاویر طراحی شده و ممکن است مستقیماً با ساختار مدل YOLO که برای تشخیص اشیا بهینه‌سازی شده همخوانی کامل نداشته باشد و باعث افزایش زمان آموزش مدل شود. چون استخراج ویژگی‌های پیچیده‌تر نیاز به پردازش بیشتری دارند. بنابراین برای بهره‌برداری کامل از این ترکیب باید هزینه‌های محاسباتی، میزان افزایش دقت و نیازهای سخت‌افزاری را به‌دقت بررسی کرد تا بهترین تعادل بین دقت و سرعت حاصل شود.

۳- مجموعه داده

جهت ارزیابی روش پیشنهادی، از مجموعه داده تصویری BVMMR استفاده‌شده است. این مجموعه داده توسط آقای بیگلری [۲۵] تهیه‌شده و شامل بیش از ۵۰۰۰ تصویر از مدل‌های مختلف خودروهای ایرانی است. شکل (۵) چند نمونه از تصاویر موجود در مجموعه داده و شکل (٦) تعداد تصاویر مربوط به هر مدل را نشان می‌دهد.

شکل (۵): تصاویر نمونه موجود در مجوعه داده BVMMR [۲۵]



شکل (٦): تعداد تصویر هر مدل در مجوعه داده BVmmR [۲۵]

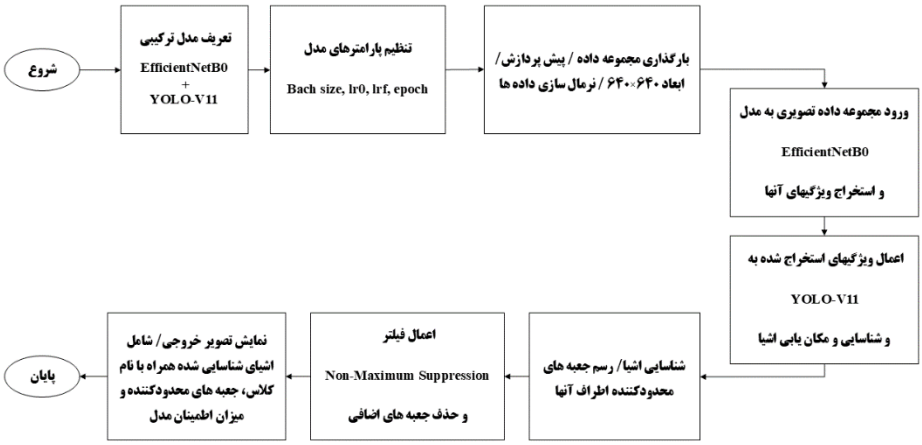
٤- ارزیابی روش پیشنهادی

کد نویسی مدل ترکیبی مورداستفاده در این مقاله با زبان برنامه‌نویسی پایتون و در محیط کولی[۱] انجام گرفته است. این محیط توسط کمپانی گوگل پشتیبانی می‌شود و پژوهشگران را به‌صورت آنلاین به یک سیستم گرافیکی قدرتمند متصل می‌کند.

٤-۱ مراحل اجرای مدل

شکل (۷) نحوه پیاده‌سازی مدل ترکیبی YOLO-V11 و EfficientNetB0 را نشان می‌دهد. ابتدا مدل‌ها معرفی‌شده و پارامترهای آن تنظیم می‌گردند. مهم‌ترین پارامترهای مدل شامل نرخ یادگیری اولیه[۲]، نرخ یادگیری نهایی[۳]، تعداد دوره‌های آموزش[٤] و اندازه بسته‌های داده[٥] هستند. تنظیم دقیق این پارامترها باعث افزایش دقت مدل می‌شود. در این مقاله مقادیر پارامترهای فوق به‌ترتیب عبارت از ۰/۰۰۰۱، ۰/۰۰۱، ۱۰۰ و ۱٦ هستند.

[۱] Google Collaboratory (Colab)
[2] Initial Learning Rate
[3] Final Learning Rate
[4] Number of Epochs
[5] Batch Size



شکل (۷): فلوچارت مربوط به مدل ترکیبی مورداستفاده

پس از تعریف مدل و تنظیم پارامترهای آن، مجموعه داده تصویری موردنظر بارگذاری شده و ابعاد تصاویر آن به ٦٤٠×٦٤٠ (ابعاد موردنظر مدل YOLO) تغییر می‌یابد. در مرحله بعد نوبت به آموزش مدل پیشنهادی می‌رسد. این مدل برپایه مدل YOLO طراحی شده است و به‌صورت یکپارچه عمل می‌کند. یعنی نیازی به آموزش جداگانه مدل‌ها نیست زیرا انتهای یک مدل به ابتدای مدل بعدی متصل شده است. با استفاده از این روش ترکیبی، تصاویر ورودی ابتدا به مدل EfficientNet-B0 واردشده تا ویژگی‌های آن‌ها استخراج شوند. سپس، این ویژگی‌ها به YOLO-V11 ارسال می‌شوند تا اشیای موجود در تصاویر و جعبه‌های محدودکننده اطراف آن‌ها را شناسایی و رسم کند. برای حذف جعبه‌های اضافی و بهبود دقت، از بلوک سرکوب غیر حداکثری[1] استفاده می‌شود. درنهایت، تصویر نهایی به همراه اشیای شناسایی‌شده به کاربر نمایش داده می‌شود و فرآیند اجرای مدل پایان می‌یابد.

٤-۲ نتایج ارزیابی مدل

شکل (۸) نتایج ارزیابی مدل ترکیبی درازای مجموعه داده BVMMR را نشان می‌دهد. نمودارها به شرح زیر هستند:

۱) نمودار خطای مکان‌یابی جعبه (box_loss) نشان می‌دهد که مدل چقدر در پیدا کردن مکان دقیق اشیا در تصاویر بهتر می‌شود.

۲) نمودار خطای شناسایی کلاس (cls_loss)[2] نشان‌دهنده میزان خطای مدل در تشخیص نام کلاس‌های مختلف اشیا است.

۳) نمودار خطای تمرکزی توزیعی در پیش‌بینی مکان اشیا (dfl_loss)[3] نشان‌دهنده میزان خطای مدل در پیش‌بینی لبه‌های جعبه‌های محدودکننده است.

لازم به‌ذکر است که میزان کم بودن خطا در هر ۳ نمودار فوق نشان‌دهنده عملکرد خوب مدل در مرحله آموزش آن است. به همین ترتیب، مقدار پایین خطای آن‌ها در نمودارهای اعتبارسنجی نیز نشان‌دهنده موفق بودن مدل در تست داده‌های جدید است.

٤) نمودار دقت نشان می‌دهد که چند درصد از اشیایی که مدل شناسایی کرده واقعاً درست هستند. این معیار برای ارزیابی کیفیت مدل در تشخیص اشیا اهمیت زیادی دارد. مقدار دقت از رابطه (۱) به‌دست می‌آید، که در آن (TP)[4] تعداد اشیای درست شناسایی‌شده توسط مدل و (FP)[5] تعداد اشیایی است که مدل به‌اشتباه شناسایی کرده است.

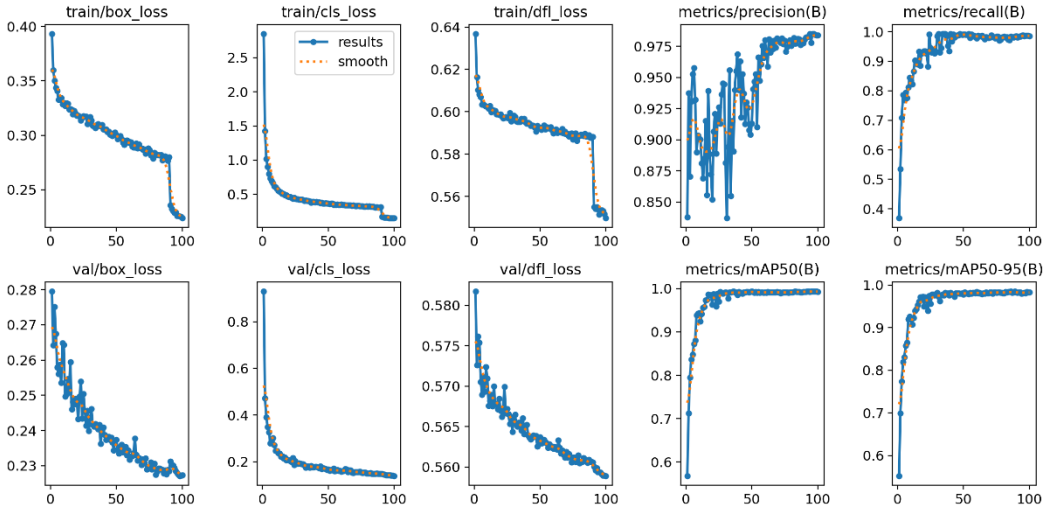
$$P = \frac{TP}{TP + FP} \qquad (۱)$$

[1] Non-Maximum Suppression
[2] Classification Loss
[3] Distribution Focal Loss
[4] True Positives
[5] False Positives

۵) نمودار فراخوانی نشان می‌دهد که مدل چه درصدی از اشیای واقعی موجود در داده‌ها را شناسایی کرده است. مقدار فراخوانی از رابطه (۲) به‌دست می‌آید، که در آن TP تعداد اشیای درست شناسایی‌شده توسط مدل و (FN)[1] تعداد اشیایی است که مدل آن‌ها را تشخیص نداده است.

$$R = \frac{TP}{TP + FN} \qquad (۲)$$

۶) mAP50 میانگین دقت مدل در شناسایی اشیا را محاسبه می‌کند. به شرطی که میزان هم‌پوشانی بین جعبه پیش‌بینی‌شده و جعبه واقعی (IoU)[2] حداقل ۵۰٪ باشد. معیار mAP50 دقت پیش‌بینی را برای تمام کلاس‌ها در نظر می‌گیرد و نشان می‌دهد که مدل چقدر در تشخیص اشیا موفق بوده است. شکل (۹) نحوه محاسبه IoU را نشان می‌دهد. درواقع IoU از تقسیم مساحت مشترک بین دوجعبه بر کل مساحت آن دوجعبه به‌دست می‌آید. مقدار میانگین دقت متوسط mAP از رابطه‌های (۳) و (٤) محاسبه می‌شود که در آن‌ها (AP)[3] میانگین دقت برای هرکلاس، P(R) میزان دقت در یک مقدار مشخص از فراخوانی، n تعداد کل کلاس‌های یک مجموعه داده و i شماره هرکلاس هستند.

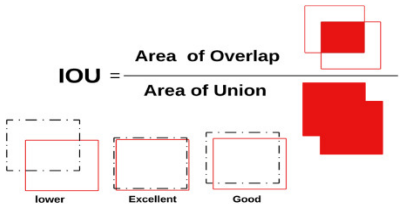


شکل (۸): نتایج ارزیابی مدل ترکیبی مورداستفاده

$$AP = \int_0^1 P(R)\, dR \qquad (۳)$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \qquad (٤)$$

۷) (mAP50-95)[4] میانگین دقت مدل را در یک بازه گسترده از IoU (از ۵۰٪ تا ۹۵٪ با گام‌های ۵٪) محاسبه می‌کند. این معیار سخت‌گیرانه‌تر از mAP50 است، زیرا مدل نه‌تنها اشیا را باید درست تشخیص دهد، بلکه موقعیت و ابعاد آن‌ها را نیز باید با دقت بیشتری پیش‌بینی کند. مقدار mAP50-95 هرچه بالاتر باشد، یعنی مدل هم در شناسایی اشیا و هم در تعیین دقیق موقعیت و اندازه آن‌ها عملکرد بهتری دارد.

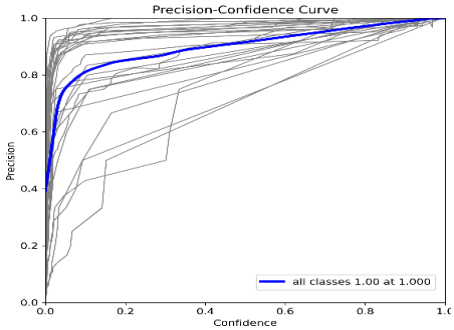


شکل (۹): نحوه محاسبه IoU [۲۶]

شکل (۱۰) رابطه بین دقت مدل و میزان اطمینان آن به پیش‌بینی‌های خود را نمایش می‌دهد. خط آبی ضخیم که میانگین دقت همه

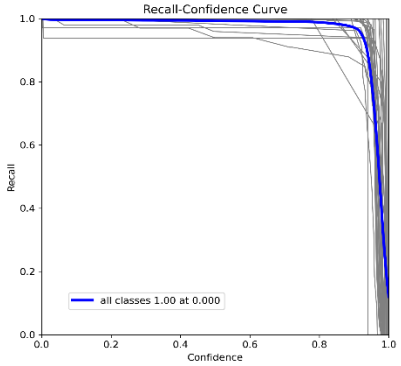
[1] False Negatives
[2] Intersection over Union (IoU)
[3] Average Precision
[4] Mean Average Precision at IoU thresholds from 0.5 to 0.95

کلاس‌ها را نشان می‌دهد، بیانگر این است که با افزایش اطمینان مدل، دقت آن نیز افزایش می‌یابد و در بازه‌ای بالاتر از ۰/٦، دقت به بیش از ۹۵٪ می‌رسد. منحنی‌های خاکستری نیز عملکرد هر کلاس را نشان می‌دهند که باوجود نوسان در سطوح پایین میزان اطمینان مدل، روندی صعودی داشته و با افزایش سطح آن به مقدار قابل قبولی دست‌یافته‌اند.



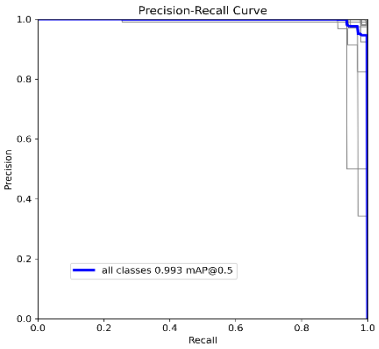
شکل (۱۰): نمودار دقت- اطمینان

شکل (۱۱) رابطه بین فراخوانی و میزان اطمینان مدل به شناسایی نمونه‌های درست را نشان می‌دهد. درواقع این شکل نشان می‌دهد که مدل در شناسایی درست نمونه‌ها چقدر مطمئن است. منحنی آبی که میانگین عملکرد همه دسته‌ها را نشان می‌دهد، بیان می‌کند که مدل در بیشتر مواقع نمونه‌های درست را به‌خوبی شناسایی کرده و فقط در اطمینان‌های خیلی بالا کمی افت دارد. خطوط خاکستری هم مربوط به هر دسته به‌صورت جداگانه هستند که بیشتر آن‌ها روندی مشابه دارند و عملکرد خوبی را نشان می‌دهند.



شکل (۱۱): نمودار فراخوانی- اطمینان

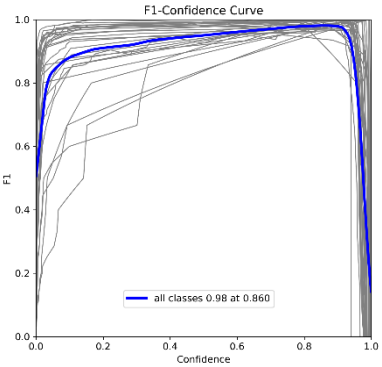
شکل (۱۲) رابطه بین دقت و فراخوانی را نمایش می‌دهد. محور افقی، درصد نمونه‌های مثبت واقعی که به‌درستی شناسایی‌شده‌اند و محور عمودی، نسبت نمونه‌های مثبت شناسایی‌شده را نشان می‌دهند. این نمودار برای ارزیابی تعادل بین مقادیر دقت و فراخوانی استفاده می‌شود.



شکل (۱۲): نمودار دقت - فراخوانی

شکل (۱۳) رابطه بین مقدار F1 و اطمینان مدل به پیش‌بینی‌های خود را نمایش می‌دهد. محور افقی، سطح اطمینان مدل در پیش‌بینی‌ها و محور عمودی، میانگین هماهنگی بین دقت و فراخوانی را نشان می‌دهد. مقدار F1 = ۰/۹۸ در مقدار اطمینان ۰/۸٦ نشان می‌دهد که در این مقدار اطمینان، تعادل بین دقت و فراخوانی برقرار است. مقدار F1 از رابطه (۵) به‌دست می‌آید و در آن P دقت و R فراخوانی مدل است.

$$F1 = 2\times\left(\frac{P\times R}{P+R}\right) \qquad (۵)$$

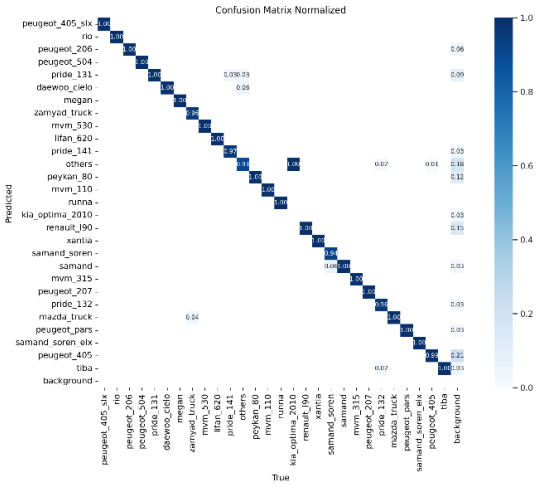


شکل (۱۳): نمودار F1 - اطمینان

شکل (۱۴) ماتریس درهم‌ریختگی[1] مدل را نشان می‌دهد. مقادیر روی قطر اصلی نشان‌دهنده دقت مدل در تشخیص صحیح هر کلاس هستند. هرچه مقدار آن‌ها به یک نزدیک‌تر باشند، مدل

[1] Confusion Matrix

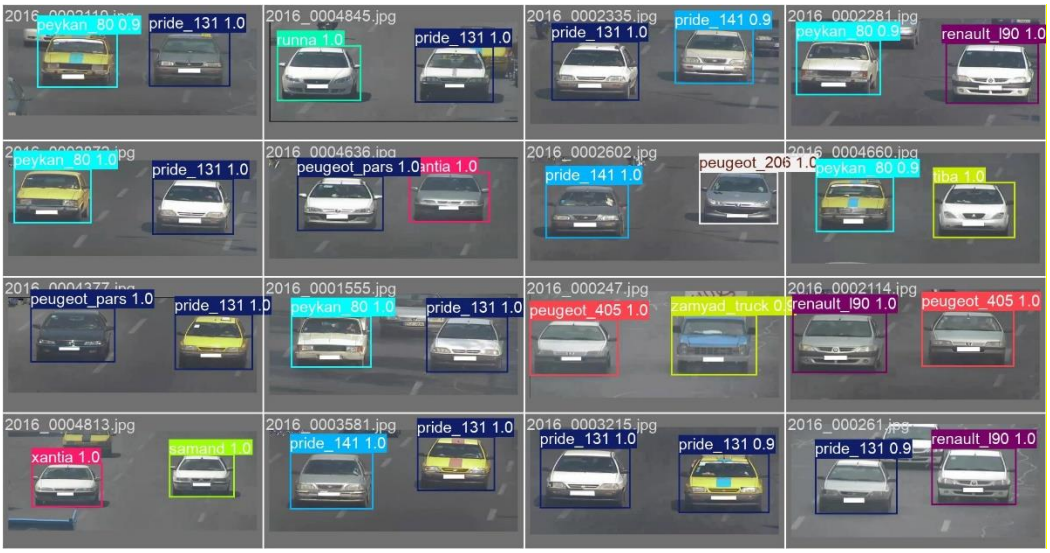
دقیق‌تر عمل می‌کند. مقادیر خارج از قطر اصلی نشان می‌دهند که ممکن است برخی از کلاس‌ها با کلاس‌های دیگر اشتباه گرفته شوند.



شکل (۱٤): ماتریس درهم‌ریختگی برای ارزیابی مدل ترکیبی

شکل (۱۵) تصویر خروجی مدل ترکیبی مورداستفاده در این مقاله را نشان می‌دهد. مشاهده می‌شود که مدل برای هر خودرو یک جعبه محدودکننده رسم کرده و بالای آن، نوع خودرو (پراید، پژو و ...) به‌همراه میزان اطمینان مدل (۱/۰، ۰/۹ و ...) را مشخص نموده است.

جدول (۱) نتایج روش پیشنهادی این مقاله را با جدیدترین روش‌های موجود مقایسه می‌کند. روش پیشنهادی، ترکیبی از مدل‌های EfficientNet-B0 و YOLO-V11 است. این مدل در معیار ارزیابی mAP50 دارای مقدار ۹۹/۳٪ و در معیار ارزیابی mAP50-95 دارای مقدار ۹۸/۳٪ است. این مقادیر نشان‌دهنده بهبود ٤/۱٪ در mAP50 و ۲٤/۱٪ در mAP50-95 نسبت به نسخه‌های استاندارد YOLO است. همچنین، دقت و فراخوانی این مدل به ترتیب ۹۸/٤٪ و ۹۸/۵٪ است، که بیانگر خطای کمتر و دقت بالاتر آن در مقایسه با سایر روش‌ها است. روش پیشنهادی با استفاده از مدل EfficientNet-B0 به‌عنوان ستون فقرات، ویژگی‌های تصویری را با جزئیات بیشتری استخراج کرده و دقت کلی تشخیص خودروها را بین ۳٪ تا ۷٪ و دقت تشخیص جزئیات آن‌را بیش از ۲۵٪ افزایش داده است. علاوه‌براین، زمان پردازش این روش حدود ۵/۵ میلی‌ثانیه برای هر تصویر است و نشان می‌دهد که این مدل حداقل ۳۰٪ سریع‌تر از مدل‌های قبلی مانند YOLO-V9 و YOLO-V8 عمل می‌کند.



شکل (۱۵): تصویر خروجی مدل ترکیبی مورداستفاده

جدول (۱) نشان می‌دهد که روش پیشنهادی در مقایسه با مدل‌های پیچیده‌تر، در معیار mAP50-95 که سخت‌گیرانه‌ترین معیار برای ارزیابی میانگین دقت متوسط شبکه YOLO است، حداقل ۹٪ پیشرفت داشته است. این معیار نشان می‌دهد که روش پیشنهادی نسبت به سایر روش‌های ذکرشده در جدول، در شناسایی جزئیات تصاویر عملکرد بهتری داشته است. به‌طورکلی، این مدل با دقت بالا، سرعت پردازش بهینه و تعادل مناسب بین دقت و کارایی، گزینه‌ای ایده‌آل برای سیستم‌های حمل‌ونقل هوشمند و نظارت شهری است.

جدول (۱): مقایسه مدل پیشنهادی با برخی از مدل‌های ارائه‌شده

منبع	مدل پیشنهادی	mAP50 (%)	mAP50-95 (%)	Precision (%)	Recall (%)
*	مدل پیشنهادی این مقاله	۹۹/۳	۹۸/۳	۹۸/٤	۹۸/٥
[۲۷]	YOLOv11	۹٥/۲	٦٥/۷	۹٦/۹	۸۹/۸
[۲۸]	YOLOv9	۹۳/٤	۷۳/۲	۹۲/۱	۸٦/۲
[۲۹]	DGNN-YOLO	۷۸/۳۰	٦٤/۷٦	۸۳/۸۲	٦۸/۷٥
[۳]	YOLOv11	۹٥/۱	۸٥/٤	۹٥/۱	۸۹/٤
[۳۰]	YOLOv9+CBAM	۹۸/٤	۸۹/۲	۹۸/۷	۹۷/۳
	Co-DETR	۹۸/۸	۹۰/۱	۹۹/۳	۹۸/٥
[۳۱]	YOLOv8s	۹۷	۸۹/۹	۹۷/۳	۹۲/۱
[۳۲]	YOLOv7	۸۲/۲	٥۹/٦	۷۸/۲	۷٦/۹
[۳۳]	YOLOv7-RAR	۹۸/۳	۸٦/۱	–	–
[۳٤]	YOLOv8-FDD	۹۷/۸	۸٤/۹	–	–

۵- نتیجه‌گیری

در این مقاله، برای تشخیص نوع خودرو از یک مدل ترکیبی جدید استفاده‌شده است که عملکرد بهتری نسبت به مدل‌های دیگر دارد. این مدل ترکیبی شامل مدل‌های EfficientNet-B0 و YOLO-V11 است. از مدل EfficientNet-B0 به‌عنوان بخش اصلی مدل (ستون فقرات) استفاده شد و به مدل ترکیبی کمک کرد تا ویژگی‌های تصویری را با دقت بیشتری استخراج کند. این کار باعث شد تا مدل ترکیبی در شناسایی خودروها دقیق‌تر عمل کند. نتایج آزمایش‌ها نشان داد که مدل پیشنهادی در معیار ارزیابی mAP50 به مقدار ۹۹/۳٪ و در معیار ارزیابی mAP50-95 به مقدار ۹۸/۳٪ رسیده که این مقادیر از مقادیر مشابه مدل‌های مورد مقایسه، بالاتر بوده است. همچنین، مقادیر دقت و فراخوانی مدل به ترتیب برابر با ۹۸/٤٪ و ۹۸/٥٪ بودند. این مقادیر نشان می‌دهند که مدل پیشنهادی نه تنها خودروها را به‌خوبی تشخیص می‌دهد، بلکه خطای کمتری نیز دارد. یکی از مزیت‌های مهم این مدل تعادل بین دقت و سرعت پردازش است. زمان پردازش هر تصویر در محیط کولی، تنها ٥/٥ میلی‌ثانیه است که شامل ۰/۲ میلی‌ثانیه برای پیش‌پردازش، ۲/۸ میلی‌ثانیه برای استنتاج و ۲/٥ میلی‌ثانیه برای پس‌پردازش می‌شود. این سرعت حداقل ۳۰٪ سریع‌تر از مدل‌های قبلی مانند YOLO-V9 و YOLO-V8 است. زمان پردازش مدل‌های قبلی حدود ۷ تا ۸ میلی‌ثانیه محاسبه‌شده است. علاوه‌براین، استفاده از EfficientNet-B0 باعث شد که مدل پیچیدگی کمتری نسبت به مدل‌های مبتنی بر ترانسفورمرها و شبکه‌های عصبی گراف (GNN) داشته باشد و در شرایط مختلف محیطی عملکرد بهتری از خود نشان دهد. مدل پیشنهادی در مقایسه با مدل‌های دیگر مانند YOLO-V9، YOLO-V11، DGNN-YOLO و YOLO-V9 + CBAM، نه تنها در شناسایی کلی خودروها بهتر عمل کرده بلکه در تشخیص جزئیات و کلاس‌بندی آن‌ها نیز برتری قابل‌توجهی داشته است. به‌طورکلی، این مدل با دقت بالا، سرعت پردازش سریع و توانایی کار در شرایط متنوع، گزینه ایده‌آلی برای استفاده در سیستم‌های حمل‌ونقل هوشمند و نظارت شهری است.

سپاس‌گزاری

نویسندگان بر خود لازم می‌دانند تا مراتب تشکر و قدردانی خود را از عوامل محترم این نشریه اعلام نمایند

.

References

[1] M. A. R. Alif, "Yolov11 for vehicle detection: Advancements, performance, and applications in intelligent transportation systems," *arXiv preprint arXiv:2410.22898*, 2024, doi: 10.48550/arXiv.2410.22898.

[2] X. Yi, Q. Wang, Q. Liu, Y. Rui, and B. Ran, "Advances in vehicle re-identification techniques: A survey," *Neurocomputing*, vol. 614, p. 128745, 2025, doi: 10.1016/j.neucom.2024.128745.

[3] M. Ashkanani, A. AlAjmi, A. Alhayyan, Z. Esmael, M. AlBedaiwi, and M. Nadeem, "A Self-Adaptive Traffic Signal System Integrating Real-Time Vehicle Detection and License Plate Recognition for Enhanced Traffic Management," *Inventions*, vol. 10, no. 1, p. 14, 2025, doi: 10.3390/inventions10010014.

[4] J. Zheng and J. Ren, "Multi Self-Supervised Pre-Finetuned Transformer Fusion for Better Vehicle Detection," *IEEE Transactions on Automation Science and Engineering*, 2024, doi: 10.1109/TASE.2024.3374759.

[5] J. Karangwa, J. Liu, and Z. Zeng, "Vehicle detection for autonomous driving: A review of algorithms and datasets," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 11, pp. 11568-11594, 2023, doi: 10.1109/TITS.2023.3292278.

[6] M. Bakirci, "Enhancing vehicle detection in intelligent transportation systems via autonomous UAV platform and YOLOv8 integration," *Applied Soft Computing*, vol. 164, p. 112015, 2024, doi: 10.1016/j.asoc.2024.112015.

[7] M. O. Yusuf *et al.*, "Enhancing vehicle detection and tracking in UAV imagery: a pixel Labeling and particle filter approach," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3401253.

[8] "Dataset for car detection on aerial photos applications." https://github.com/jekhor/aerial-cars-dataset (accessed 2018).

[9] I. Bozcan and E. Kayacan, "Au-air: A multi-modal unmanned aerial vehicle dataset for low altitude traffic surveillance," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020: IEEE, pp. 8504-8510, doi: 10.1109/ICRA40945.2020.9196845.

[10] M. Zohaib, M. Asim, and M. ELAffendi, "Enhancing emergency vehicle detection: a deep learning approach with multimodal fusion," *Mathematics*, vol. 12, no. 10, p. 1514, 2024, doi: 10.3390/math12101514.

[11] Z. Song, Y. Wang, S. Xu, P. Wang, and L. Liu, "Lightweight Vehicle Detection Based on Mamba_ViT," *Sensors*, vol. 24, no. 22, p. 7138, 2024, doi: 10.3390/s24227138.

[12] Z. Wang and C. Ma, "Weak-mamba-unet: Visual mamba makes cnn and vit work better for scribble-based medical image segmentation," *arXiv preprint arXiv:2402.10887*, 2024, doi: 10.48550/arXiv.2402.10887.

[13] L. Wen *et al.*, "UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking," *Computer Vision and Image Understanding*, vol. 193, p. 102907, 2020, doi: 10.1016/j.cviu.2020.102907.

[14] Y. Zhang, W. Wang, M. Ye, J. Yan, and R. Yang, "LGA-YOLO for Vehicle Detection in Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025, doi: 10.1109/JSTARS.2025.3535090.

[15] "data-unicorn-2008." https://github.com/AFRL-RY/data-unicorn-2008 (accessed 2019).

[16] S. Razakarivony and F. Jurie, "Vehicle detection in aerial imagery: A small target detection benchmark," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 187-203, 2016, doi: 10.1016/j.jvcir.2015.11.002.

[17] G.-S. Xia *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3974-3983, doi: 10.1109/cvpr.2018.00418

[18] H. Park, K. Kim, I. Jeong, J. Jung, and J. Cho, "Special Vehicle Classification Algorithm-Based System for Dedicated Parking Zone Violation Detection in South Korea," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3526862.

[19] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019: PMLR, pp. 6105-6114, doi: 10.48550/arXiv.1905.11946.

[20] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," *arXiv preprint arXiv:2410.17725*, 2024, doi: 10.48550/arXiv.2410.17725.

[21] "YOLO11_EfficientNet." https://github.com/JYe9/YOLO11_EfficientNet/tree/main (accessed 2025).

[22] N. Jegham, C. Y. Koh, M. Abdelatti, and A. Hendawi, "Evaluating the evolution of yolo (you only look once) models: A comprehensive benchmark study of yolo11 and its predecessors," *arXiv preprint arXiv:2411.00201*, 2024, doi: 10.48550/arXiv.2411.00201.

[23] J. Huang, K. Wang, Y. Hou, and J. Wang, "LW-YOLO11: A Lightweight Arbitrary-Oriented Ship Detection Method Based on Improved YOLO11," *Sensors*, vol. 25, no. 1, p. 65, 2024, doi: 10.3390/s25010065.

[24] S. U. Amin, M. S. Abbas, B. Kim, Y. Jung, and S. Seo, "Enhanced anomaly detection in pandemic surveillance videos: An attention approach with

EfficientNet-B0 and CBAM integration," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3488797.
[25] M. Biglari, A. Soleimani, and H. Hassanpour, "Part-based recognition of vehicle make and model," *IET image processing*, vol. 11, no. 7, pp. 483-491, 2017, doi: 10.1049/iet-ipr.2016.0969.
[26] M. Y. Shams *et al.*, "Automated On-site Broiler Live Weight Estimation Through YOLO-Based Segmentation," *Smart Agricultural Technology*, p. 100828, 2025, doi: 10.1016/j.atech.2025.100828.
[27] Y. Gao, Y. Xin, H. Yang, and Y. Wang, "A Lightweight Anti-Unmanned Aerial Vehicle Detection Method Based on Improved YOLOv11," *Drones*, vol. 9, no. 1, p. 11, 2024, doi: 10.3390/drones9010011.
[28] S. A. Fahim, "Finetuning YOLOv9 for Vehicle Detection: Deep Learning for Intelligent Transportation Systems in Dhaka, Bangladesh," *arXiv preprint arXiv:2410.08230*, 2024, doi: 10.48550/arXiv.2410.08230.
[29] S. Soudeep, M. Mridha, M. A. Jahin, and N. Dey, "DGNN-YOLO: Dynamic Graph Neural Networks with YOLO11 for Small Object Detection and Tracking in Traffic Surveillance," *arXiv preprint arXiv:2411.17251*, 2024, doi: 10.48550/arXiv.2411.17251.
[30] E. Arthur, A. Aboah, and Y. Huang, "A Novel FHWA-Compliant Dataset for Granular Vehicle Detection and Classification," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3486603.
[31] Q.-A. N. Duc *et al.*, "Optimizing Traffic Light Control using YOLOv8 for Real-Time Vehicle Detection and Traffic Density," in *2024 9th International Conference on Integrated Circuits, Design, and Verification (ICDV)*, 2024: IEEE, pp. 119-124, doi: 10.1109/ICDV61346.2024.10616901.
[32] J. Su, F. Wang, and W. Zhuang, "An Improved YOLOv7 Tiny Algorithm for Vehicle and Pedestrian Detection with Occlusion in Autonomous Driving," *Chinese Journal of Electronics*, vol. 34, no. 1, pp. 282-294, 2025, doi: 10.23919/cje.2023.00.256.
[33] Y. Zhang, Y. Sun, Z. Wang, and Y. Jiang, "YOLOv7-RAR for urban vehicle detection," *Sensors*, vol. 23, no. 4, p. 1801, 2023, doi: 10.3390/s23041801.
[34] X. Liu, Y. Wang, D. Yu, and Z. Yuan, "YOLOv8-FDD: A real-time vehicle detection method based on improved YOLOv8," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3453298

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 2, No. 2, Autumn & Winter 2024, pp. 111-124

Vehicle Models Recognition for Intelligent Control of Vehicle Traffic by Deep Learning

Morteza Ehsandoust[1], Atefeh Salimi Shahraki [2*], Mohammad Rohollah Yazdani[3], Mohammad Lali Dastjerdi[1]

[1]MSc. student Department of Electrical Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

[2]Assistant Professor Department of Electrical Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

[3]Associate Professor Department of Electrical Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

Article Information

Original Research Paper

Received:
2025 February 9

Accepted:
2025 April 26.

Keywords:
Computer Vision, Intelligent Transportation System, Deep Learning, YOLO, Efficient Net

Corresponding Author*:
a.salimi@khuisf.ac.ir

Abstract

With the increasing number of vehicles worldwide and the advancement of technology, intelligent transportation systems have emerged as a key solution for traffic control and road safety enhancement. One of the crucial components of these systems is vehicle type recognition. In this paper, a combination of EfficientNet-B0 and YOLO-V11 models is used for vehicle type classification. In this model, EfficientNet-B0 serves as the backbone of YOLO-V11, extracting image features and feeding them into the YOLO model to accurately determine the location and type of vehicles. To evaluate the performance of the proposed model, the BVMMR image dataset—containing over 5,000 of various Iranian vehicles—was utilized. The model was implemented using Python and executed in the Colab online environment. The results demonstrate that the proposed model achieves a mean Average Precision at 50% IoU (mAP50) of 99.3% and a mean Average Precision across IoU thresholds from 50% to 95% (mAP50-95) of 98.3%. Additionally, in terms of processing speed, the model delivers an efficient performance with an images average preprocessing time of 0.2 milliseconds, inference time of 2.8 milliseconds, and post-processing time of 2.5 milliseconds per image.

: 10.22034/ABMIR.2025.22898.1112

E-ISSN: 2821-2037