نشریه علمی علوم و فنون نقشه‌برداری، دوره ۱۳، شماره ۲، صفحه ۱۵ تا ۲۸، آذر ماه ۱۴۰۲ – مقاله پژوهشی

شناسایی خودکار خودرو از تصاویر ماهواره‌ای گوگل ارث بر مبنای شبکه‌های عصبی یادگیری عمیق تک مرحله‌ای آشکارساز شیئ

مصطفی کابلی‌زاده [*۱]، محمد عباسی [۲]

[۱] دانشیار گروه سنجش‌ازدور و سیستم اطلاعات جغرافیایی، دانشکده علوم زمین، دانشگاه شهید چمران اهواز، اهواز، ایران
m.kabolizade@scu.ac.ir

[۲] دانشجوی دکتری گروه سنجش‌ازدور و سیستم اطلاعات جغرافیایی، دانشکده علوم زمین، دانشگاه شهید چمران اهواز، اهواز، ایران
mohammadabbasi@ut.ac.ir

(تاریخ دریافت: آبان ۱۴۰۱، تاریخ تصویب: آذر ۱۴۰۲)

چکیده

شبکه حمل و نقل جاده‌ای و شهری زندگی روزمره ما را برای مسیریابی بهینه تسهیل می‌کنند. در شبکه معابر، مدیریت ترافیک یکی از چالش‌های اصلی مدیران است. در این خصوص، اولین گام برآورد تراکم خودروها در سطح شبکه معابر شهری می‌باشد. برآورد تعداد خودروها یا سطح اشغال خودروها در سطح کل شهر، با درنظر گرفتن زمان و هزینه کمتر، فقط با تصاویر ماهواره‌ای امکان‌پذیر است. در این راستا، در این پژوهش از تصاویر ماهواره‌ای با قدرت تفکیک مکانی بالای در دسترس و قابل دانلود سامانه گوگل ارث استفاده شده‌است. برای شناسایی موقعیت خودروها از روش مبتنی بر یادگیری عمیق تک مرحله‌ای با معماری RetinaNet و بر اساس شبکه‌های عصبی باقیمانده با تعداد لایه ۱۸، ۳۴ و ۵۰ استفاده شده‌است. برای داده‌های آموزشی موقعیت خودروها با جعبه‌های مرزی مشخص شده و سپس تصاویر ماهواره‌ای با ابعاد ۱۲۸ در ۱۲۸ پیکسل و گام ۶۴ پیکسل بریده شده‌است. از کل داده‌های آموزشی ۸۰ درصد برای آموزش و ۲۰ درصد برای اعتبارسنجی به صورت تصادفی استفاده شده‌است. مدل‌ها در ۵۰ دوره تکرار و با میانگین دقت بالای ۰/۷ آموزش داده شده‌اند. برای ارزیابی مدل‌های آموزش دیده از تصاویر ماهواره‌ای حاوی بیش از ۱۵۰۰۰ خودرو استفاده گردید. پارامتر امکان همپوشانی روش سرکوب غیرحداکثری ۲۵ درصد اعمال شده‌است. نتیجه نهایی نشان می‌دهد که استفاده از مدل پیشنهادی در شناسایی خودروها دارای دقت مناسبی می‌باشد. مدل آشکارساز RetinaNet با شبکه یادگیری عمیق باقیمانده دارای ۵۰ لایه از نظر معیار میانگین دقت با ۰/۸۷، معیار دقت با ۰/۷، معیار بازیابی با ۰/۹۹ و معیار F1 با ۰/۸۲ بهترین عملکرد را داشته‌است. چالش اصلی مدل‌های پیشنهادی در مناطق دارای تراکم بالای خودور می‌باشد، که امکان تشخیص دقیق تعداد خودروها را بدلیل اندازه فاصله نمونه‌برداری زمینی تصاویر ماهواره‌ای کاهش می‌دهد اما سطح اشغال را بهتر برآورد می‌کند.

واژگان کلیدی: یادگیری عمیق، تصاویر ماهواره‌ای، RetinaNet، شبکه‌های باقیمانده، گوگل ارث.

* نویسنده رابط

۱- مقدمه

شناسایی زیرساخت‌های جدید (تجاری، خدماتی، صنعتی یا مسکونی) از تصاویر ماهواره‌ای یک روش اثبات شده برای بررسی و ارزیابی رشد اقتصادی و شهری است. سطح فعالیت یا بهره‌برداری از این مکان‌ها ممکن است به سختی با بازرسی میدانی تعیین شود، اما می‌توان با استفاده از تراکم حضور خودروها در خیابان‌ها و پارکینگ‌های مجاور، میزان رشد را استنباط نمود. دستیابی به این مهم نیازمند امکان شمارش خودروها در سطح معابر است. همچنین با افزایش سریع تعداد وسایل نقلیه، محققان با یک کار چالش برانگیز روبرو هستند، زیرا دوربین‌های مکان ثابت و سنسورهای حرکت به اندازه کافی گسترده نیستند. شمارش این تعداد عظیم وسایل نقلیه برای بهبود مدیریت ترافیک، تشخیص نیازهای سوخت در مکان‌های خاص و تخمین انتشار گازهای گلخانه‌ای در مناطق شلوغ به منظور اطلاع از درصد آلودگی ضروری است. شمارش دوره‌ای خودروها نیز برای تخمین ازدحام احتمالی آینده برای برنامه‌ریزی زیرساخت‌های حمل و نقل مهم است.

ناوگان وسایل نقلیه در جهان به ویژه در شهرها به طور مداوم افزایش می‌یابد. بسیاری از شهرها از تجهیزات میدانی مانند دوربین‌های مکان ثابت یا حسگرهای حرکت در چراغ-های راهنمایی برای نظارت بر وسایل نقلیه استفاده می‌کنند، اخیراً دوربین‌های نصب شده بر روی پهپاد برای ارائه میدان دید وسیع‌تر استفاده می‌شوند [۱و۲].

در سال‌های اخیر تشخیص و شمارش وسایل نقلیه از تصاویر سنجش از دور یک موضوع تحقیقاتی فعال است که برای نظارت بر ترافیک و سیستم‌های حمل و نقل، بررسی زیرساخت‌های تجاری یا صنعتی جدید و همچنین مطالعه سطوح شهرنشینی و غیره به کار می‌رود. در دهه اخیر این وظیفه به لطف توسعه فناوری‌های تصاویر ماهواره‌ای با قدرت تفکیک بالای مکانی همراه با روش‌های مدرن تشخیص اشیا در حوزه بینایی کامپیوتر و یادگیری ماشین، توجه محققان بسیاری را به خود جلب کرده‌است. بسیاری از مطالعات از تصاویر هوایی استفاده کرده و اثر بخشی آن‌ها برای شناسایی وسایل نقلیه از تصاویر با وضوح مکانی بالاتر از ۰/۳ متر ثابت شده است [۳و۴].

برای کاهش محدوده جستجو جهت تشخیص وسیله نقلیه در تصاویر ماهواره‌ای، می‌توان ابتدا معابر را تشخیص و استخراج نمود [۵]، سپس جستجو برای تشخیص وسیله نقلیه را فقط در محدوده معابر استخراج شده انجام داد. شناسایی جاده‌ها و وسایل نقلیه می‌تواند در شهرهای هوشمند کاربرد داشته باشد، جایی که راه‌حل‌های نظارت و زمان‌بندی ترافیک در آن‌ها مورد نیاز است.

اما تشخیص اشیای کوچک (مانند: خودروها) چالشی منحصر به فرد است، زیرا مقدار اطلاعات موجود برای آشکارساز کاهش می‌یابد [۶]. در حالی که با یک جسم بزرگ، پیکسل‌های زیادی برای آشکارساز وجود دارد تا بتواند ویژگی‌ها را استخراج نماید، در اشیاء کوچک تعداد پیکسل‌ها به میزان قابل توجهی کاهش می‌یابد. اندازه اجسام نیز بسته به وضوح مکانی ماهواره‌ای که تصویر را اخذ کرده، می‌تواند بسیار متفاوت باشد. اندازه‌های متفاوت اشیا می‌تواند تعمیم شبکه‌ها از یک مجموعه داده به مجموعه دیگر را دشوارتر کند.

علاوه بر چالش کاهش اندازه جسم، تشخیص وسیله نقلیه در تصاویر ماهواره‌ای به دلیل چشم انداز غیر معمول و از بالا به پایین دشوارتر است. وسایل نقلیه عموماً به گونه‌ای طراحی شده‌اند که از دید انسان به راحتی قابل شناسایی باشند و بالای اکثر خودروها دارای سطحی صاف با جزئیات بسیار کم هستند. علاوه بر این، شناسایی اشیایی مانند وسایل نقلیه با رنگ‌های تیره‌تر ممکن است در محوطه پارکینگ به دلیل عدم تضاد بین خودرو و زمین دشوارتر باشد. به همین ترتیب با کاهش تعداد ویژگی‌ها، تشخیص اشتباه یک چالش همیشگی است. به عنوان مثال، اگر قدرت تفکیک‌پذیری پایین باشد، یک هواگیر مستطیل شکل یا پنجره تیره روی سقف یک ساختمان بتنی ممکن است از یک وسیله نقلیه روی سنگفرش قابل تشخیص نباشد.

تشخیص اشیا از دیرباز یک هدف در بینایی کامپیوتری بوده‌است. موج یادگیری عمیق، آشکارسازهای بسیاری را به وجود آورده‌است که از روش‌های کلاسیک بهتر عمل می-کنند. آشکارسازهای یادگیری عمیق که توسعه یافته‌اند به دو گروه تقسیم می‌شوند: آشکارسازهای دومرحله‌ای و یک مرحله‌ای. آشکارسازهای دومرحله‌ای، مانند شبکه عصبی کانولوشن مبتنی بر ناحیه (R-CNN[۱]) [۷] و شبکه‌های عصبی پیچشی سریعتر مبتنی بر ناحیه (FRCNN[۲]) [۸]، ابتدا از یک شبکه پیشنهادی ناحیه‌ای استفاده می‌کنند که نشان می‌دهد کجا در یک تصویر ارزش جستجوی اشیا را دارد

۱ Regions with Convolutional Neural Networks

۲ Faster R-CNN

و یک شبکه ثانویه در واقع وظیفه طبقه‌بندی و محلی‌سازی اشیاء را انجام می‌دهد. دسته دوم آشکارسازها را آشکارسازهای یک مرحله‌ای می‌نامند. این آشکارسازها مانند شبکه عصبی یادگیری عمیق تک نگاه (SDD[3]) [۹] و YOLO[4] [۱۰] تنها یک بار از روی یک تصویر عبور می‌کنند و از مرحله پیشنهادی منطقه می‌گذرند. به دلیل سبک‌تر بودن روش‌های یک مرحله‌ای، آن‌ها سریع‌تر از همتایان دو مرحله‌ای خود اجرا می‌شوند، با این حال این افزایش سرعت اغلب به قیمت دقت تمام می‌شود. هنگامی که آشکارساز RetinaNet تک نگاه در سال ۲۰۱۸ معرفی شد، روند تغییر کرد و از آشکارسازهای دو مرحله‌ای بر روی مجموعه داده COCO بهتر عمل کرد [۱۱].

یانگ و همکاران [۱۲] یک شبکه عصبی پیچشی سریعتر مبتنی بر ناحیه را برای تشخیص خودرو در تصاویر هوایی با افزودن اتصالات میانبر از لایه‌های کم عمق به عمیق به منظور یادگیری ویژگی‌هایی با اطلاعات غنی از جزئیات گسترش دادند. نویسندگان از تابع ضرر کانونی برای طبقه‌بندی با توجه به موضوع مثال‌های مثبت آسان و مثال‌های منفی سخت در طول آموزش استفاده کردند. کاربرد روش تشخیص پیشنهادی در یک مجموعه داده با تصاویر ثبت شده در هر دو نمای نادیر و مایل نشان داده شد.

دویلارد[5] در تحقیقات خود از شبکه RetinaNet برای تشخیص وسیله نقلیه در تصاویر هوایی استفاده نمود. با وجود دستیابی به دقت مناسب، مشکل اصلی این بود که مدل شفت‌های تهویه بالای ساختمان‌ها را هم ماشین در نظر می‌گرفت [۱۳].

استوپارو[6] و همکاران یک مدل تشخیص شیئ یک مرحله‌ای را برای یافتن وسایل نقلیه در تصاویر ماهواره‌ای با استفاده از معماری RetinaNet و مجموعه داده‌های Cars Overhead With Context پیشنهاد و ارائه کردند که دارای دقت مناسب بر روی مجموعه داده‌های مورد استفاده بود [۱۴].

بطور کلی استفاده از تصاویر ماهواره‌ای در تشخیص اهداف کوچک دارای مشکلات زیر می‌باشد [۱۵]:

- ✓ ابعاد کوچک اشیا (۱۵ تا ۳۰ پیکسل برای هر خودرو)،
- ✓ ثابت نبودن موقعیت جسم (یعنی می‌توان آن را چرخاند)،
- ✓ حجم زیاد تصویر (صدها مگاپیکسل)،
- ✓ محدودیت تعداد مجموعه داده‌های موجود.

از طرف دیگر، فواصل واقعی و اندازه تصویر همیشه مشخص است، بنابراین محاسبه ابعاد آسان است. علاوه بر این، زاویه مشاهده ثابت است.

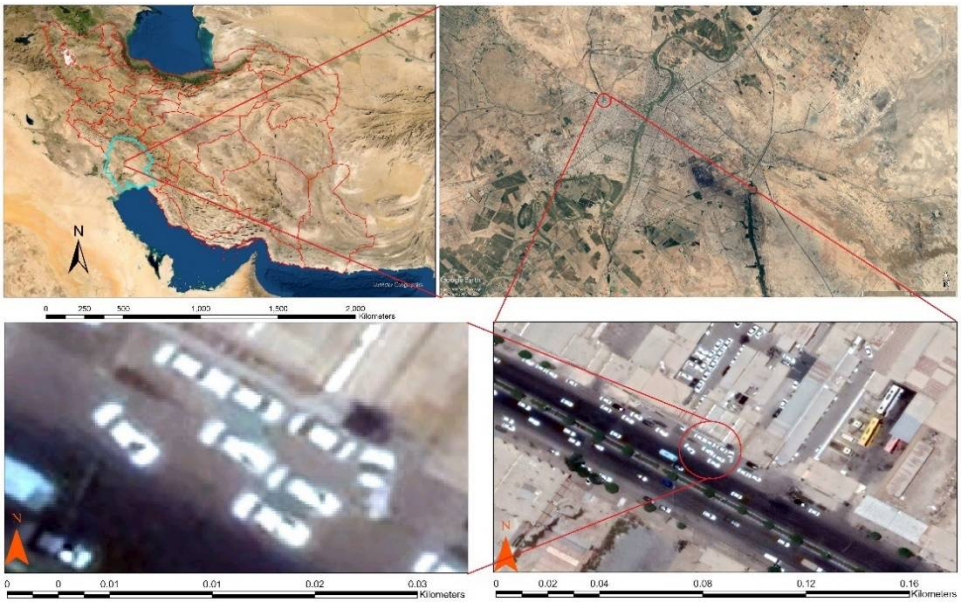
علاوه بر چالش‌های فوق در این پژوهش با توجه به عدم دسترسی و امکان خرید تصاویر ماهواره‌ای با قدرت تفکیک مکانی بالا، از تصاویر ماهواره‌ای سامانه گوگل ارث استفاده شده‌است که هنگام دانلود کیفیت تصویر نسبت به تصویر اصلی کمتر بوده و در نتیجه مشکل تشخیص و استخراج وسیله نقلیه بیشتر می‌شود. در این تحقیق مدلی برمبنای مدل آشکارساز RetinaNet [۱۱] پیشنهاد شده‌است که از یادگیری چند وظیفه‌ای برای بهبود تشخیص اجسام کوچک (در این پژوهش خودروهای سواری) در تصاویر ماهواره‌ای استفاده می‌کند. هدف اصلی تشخیص موقعیت مکانی خودورهای شخصی از تصاویر ماهواره‌ای با دقت بالا و زمان پایین است.

۲- منطقه مورد مطالعه و داده‌ها

۲-۱- منطقه مورد مطالعه

منطقه مورد مطالعه در این پژوهش شهر اهواز در استان خوزستان می‌باشد. هر چند مدل آموزش دیده می‌تواند برای تصاویر ماهواره‌ای سایر شهرها و جاده‌های بین‌شهری هم استفاده شوند. چون در ایران تنها تصاویر ماهواره‌ای با قدرت تفکیک مکانی بالای در دسترس، تصاویر ماهواره‌ای در سامانه گوگل ارث می‌باشد، در این تحقیق از تصاویر ماهواره‌ای با قدرت تفکیک بالای مکانی گوگل ارث استفاده شده‌است. تصاویر ماهواره‌ای از منطقه مورد مطالعه از گوگل ارث توسط نرم‌افزار Google Earth Images Downloader با بزرگنمایی ۲۱ دانلود شده‌است. شکل (۱) موقعیت منطقه مورد مطالعه و بخشی از تصاویر ماهواره‌ای دانلود شده از سامانه گوگل ارث را نشان می‌دهد.

۳ Single Shot Multibox Detector
۴ You Only Look Once
۵ Douillard
۶ Stuparu



شکل ۱- منطقه مورد مطالعه و نمونه‌ی تصاویر ماهواره‌ای دانلود شده از سامانه گوگل ارث

۲-۲- داده‌ها

در ابتدا تصاویر ماهواره‌ای با قدرت تفکیک مکانی بالا برای منطقه مورد مطالعه از سامانه گوگل ارث دانلود شده‌است. تصاویر گوگل ارث دانلود شده در باند مرئی (RGB) بوده و فاقد باندهای دیگر در تصاویر اصلی تصاویر ماهواره‌ای کوئیک برد می‌باشد که چالش تشخیص و استخراج خودرو را بدلیل عدم امکان استفاده از سایر باندها و کیفیت پایین‌تر بیشتر می‌کند. برای تولید داده‌های آموزشی و اعتبار سنجی ابتدا موقعیت خودرو از تصاویر ماهواره‌ای دانلود شده، استخراج شده‌است. سپس کل تصویر با یک پنجره متحرک با ابعاد ۱۲۸ در ۱۲۸ پیکسل و گام ۶۴ پیکسل در هر دو جهت افقی و عمودی برش داده شده است، یعنی تصاویر بریده شده دارای هم‌پوشانی ۵۰ درصد هستند. اندازه تصاویر بریده شده با توجه به ابعاد خودروها و ارزیابی دقت اولیه نتایج برای ابعاد مختلف و با تاکید بر زمان پردازش انتخاب شده‌است. شکل (۲) نمونه‌هایی از تصاویر بریده شده را نشان می‌دهد. در هر تصویر بریده شده مختصات خودروها با جعبه-های مرزی مشخص شده‌است. تصاویر بریده شده با ابعاد ۱۲۸ در ۱۲۸ پیکسل از تصاویر ماهواره‌ای گوگل ارث شهر اهواز استخراج شده‌است. بیش از ۲۰۰۰ قطعه تصویر حاوی خودرو برای آموزش و اعتبارسنجی استفاده شده‌است.

شکل۲- الف: نمونه تصاویر آموزشی حاوی خودرو

شکل۲- ب: نمونه تصاویر آموزشی پس‌زمینه

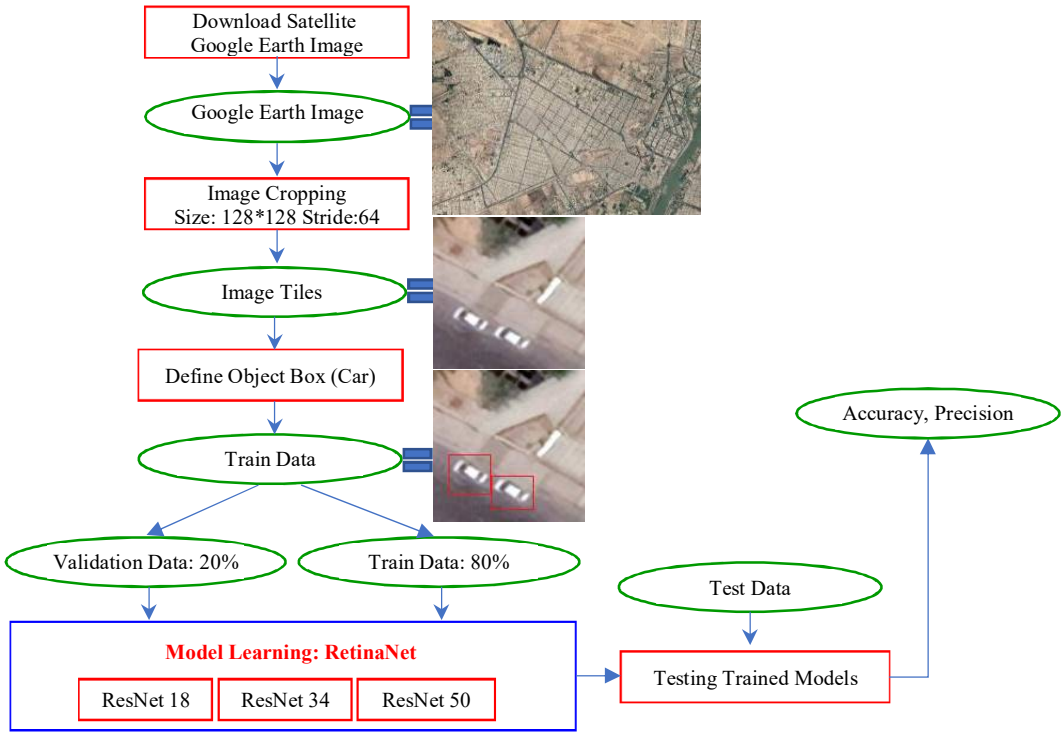
در این پژوهش به‌صورت تصادفی ۲۰ درصد داده‌ها برای اعتبارسنجی و ۸۰ درصد برای آموزش مدل‌های پیشنهادی

استفاده شده‌است. نمونه‌های آموزشی در دوکلاس شامل تصاویر حاوی خودرو و پس‌زمینه تهیه گردید، البته در هر تصویر بریده شده، موقعیت هر خودرو با یک مستطیل (جعبه مرزی) مشخص و در کلاس خودرو قرار داده شده‌است و خارج از این جعبه مرزی کلاس پس‌زمینه منظور شده‌است. در نمونه‌های استخراج شده کیفیت تصویر، رنگ، کنتراست یا حذف سایه تغییری نکرده‌است. برای داده‌های تست مدل هم از تصاویر ماهواره‌ای در شرایط مختلف مانند کوچه‌ها و خیابان‌های با ترافیک بالا و کم تا پارکینگ‌ها استفاده شده‌است. کلیه تصاویر ماهواره‌ای زمین‌مرجع بوده و دارای سیستم تصویر UTM هستند که امکان استخراج موقعیت مختصاتی خودروها وجود داشته باشد.

۳- روش تحقیق

هدف این پژوهش شناسایی خودکار موقعیت خودروهای سواری از تصاویر ماهواره‌ای گوگل ارث می‌باشد. در این پژوهش از روش‌های یادگیری عمیق آشکارساز مبتنی بر معماری RetinaNet استفاده شده‌است. شکل (۳) روند کلی انجام پژوهش را نشان می‌دهد.

در این پژوهش ابتدا تصاویر ماهواره‌ای با قدرت تفکیک مکانی گوگل ارث برای منطقه مورد مطالعه دانلود شده‌است. سپس این تصاویر با ابعاد ۱۲۸ در ۱۲۸ پیکسل و گام هم‌پوشانی ۶۴ پیکسل بریده شده‌اند. موقعیت خودروها بر روی این تصاویر با جعبه‌های مرزی مشخص شده‌است.



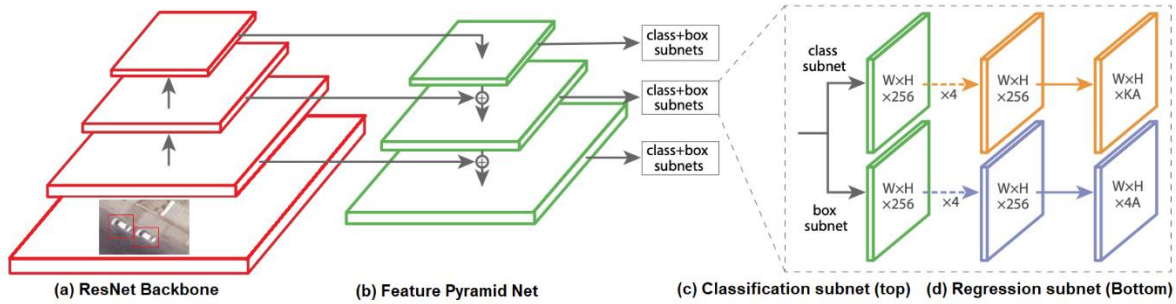
شکل۳- روند نمای کلی پژوهش

در نهایت این داده‌ها با اعمال ۸۰ درصد جهت یادگیری و ۲۰ درصد جهت اعتبارسنجی برای آموزش مدل پیشنهادی بر مبنای معماری آشکارساز RetinaNet استفاده شده‌است. برای ستون اصلی شبکه یادگیری عمیق مبتنی بر معماری آشکارساز RetinaNet از شبکه‌های یادگیری عمیق باقیمانده[۷] دارای ۱۸ لایه (ResNet 18)، ۳۴ لایه (ResNet 34) و ۵۰ لایه (ResNet 50) استفاده شده‌است. برای آموزش مدل‌های پیشنهادی از کامپیوتر شخصی با پردازش‌گر Core i7 نسل ۸ و حافظه با دستیابی تصادفی ۱۲ گیگابایت استفاده شده‌است. آماده‌سازی داده‌ها در نرم‌افزار ArcGIS Pro انجام شده‌است. برای اجرای مدل از کتابخانه‌های آماده یادگیری عمیق (تنسورفلو[۸] و پای‌تورچ[۹]) و زبان برنامه‌نویسی پایتون استفاده شده‌است. نتایج مدل‌ها با داده‌های تست (تصاویر ماهواره‌ای جدید) مورد ارزیابی قرار گرفته و دقت و سرعت مدل‌ها برآورد شده‌است. در ادامه ساختار مدل پیشنهادی شرح داده شده‌است.

۷ Deep Residual Network
۸ TensorFlow
۹ PyTorch

۳-۱- معماری RetinaNet

معماری RetinaNet برای اولین بار [۱۱] توسط فیس بوک در سال ۲۰۱۸ پیشنهاد شد که یک آشکارساز شیئ تک نگاه است که به عملکردی پیشرفته دست یافته‌است. معماری این شبکه از سه جزء مجزا تشکیل شده است: شبکه هرمی ویژگی (FPN[۱۰])، شبکه‌های عصبی یادگیری عمیق باقیمانده (ResNet) و شبکه‌های فرعی برای طبقه‌بندی و رگرسیون. شکل (۴) ساختار کلی معماری آشکارساز RetinaNet را نشان می‌دهد.



شکل۴- ساختار کلی معماری آشکاساز RetinaNet [۱۱]

شکل۵- ساختار مدل‌های یادگیری عمیق باقیمانده با ۱۸، ۳۴ و ۵۰ لایه

۳-۱-۱- شبکه‌های باقیمانده (ResNet)

شبکه‌های باقیمانده اولین بار توسط هی[۱۱] و همکاران [۱۶] ارائه گردید. شبکه‌های باقیمانده خانواده‌ای از شبکه‌های عصبی هستند که با شبکه‌های عصبی کانولوشن مرسوم، متفاوت هستند، این تفاوت در داشتن اتصالات میانبر بین لایه‌ها در شبکه‌های باقیمانده است. افزودن این اتصالات میانبر می‌تواند امکان ایجاد شبکه‌های عصبی عمیق‌تر را فراهم کند که با دقت بالاتری نسبت به پیشینیان خود آموزش داده شوند. در این پژوهش از شبکه‌های عمیق ۱۸، ۳۴ و ۵۰ لایه استفاده شده‌است. ابتدا، یک کانولوشن با پرش دو گام بر روی تصویر ورودی، و به دنبال آن تابع ادغام ماکزیمم، که منجر به کاهش شدید نمونه‌گیری ورودی

۱۰ Feature Pyramid Net

۱۱ He

می‌شود، اعمال می‌شود. سپس یک واحد با دو بلوک پایه اعمال می‌شود. واحد دوم نیز شامل دو بلوک است. اما در کانولوشن اول مقدار پرش برابر دو گام است، بنابراین نقشه‌های ویژگی را کم‌نمونه می‌کند. شکل (۵) ساختار شبکه‌های یادگیری عمیق باقیمانده ۱۸، ۳۴ و ۵۰ لایه را نشان می‌دهد.

۳-۱-۲- شبکه هرمی ویژگی (FPN)

شبکه هرمی ویژگی، توسط فیس بوک معرفی شد [۱۷]. شبکه هرمی ویژگی یک معماری شبکه عصبی است که به دنبال مدیریت واریانس مقیاس در اشیاء درون یک تصویر است. این ایده از هرم‌های تصویری مرسوم در تکنیک‌های کلاسیک بینایی کامپیوتر الهام گرفته شده‌است که واریانس مقیاس را با نمونه‌برداری از یک تصویر در هرم‌های تصویری مختلف و اجرای الگوریتم مورد نظر بر روی هر تصویر دوباره نمونه‌برداری شده، مدیریت می‌کند. شبکه هرمی ویژگی با استفاده از تمایل لایه‌های عمیق‌تر شبکه‌های باقیمانده برای داشتن ویژگی‌های وضوح پایین‌تر، اما با اطلاعات معنایی غنی‌تر، به اثر مشابهی دست می‌یابند. بنابراین برای دستیابی به مکان دقیق اشیاء در مقیاس‌های مختلف، لایه‌های ویژگی متعدد در معماری شبکه‌های باقیمانده به همراه لایه‌های ویژگی چند مقیاسی غنی‌شده انتخاب می‌شوند. لایه‌های ویژگی چند مقیاسی غنی‌شده از ترکیب لایه ویژگی کم‌عمق‌تر با لایه ویژگی عمیق‌تر بعدی به کمک نزدیک‌ترین همسایه هرم تصویری ایجاد می‌شوند. شبکه هرمی ویژگی، یک شبکه پیچشی دارای سه نوع اتصال جانبی، پایین به بالا و بالا به پایین است تا نقشه ویژگی کانولوشن برای هر تصویر ورودی ساخته شود.

۳-۱-۳- شبکه‌های فرعی برای طبقه‌بندی و رگرسیون

خروجی نقشه‌های ویژگی ایجاد شده توسط شبکه هرمی ویژگی با شبکه‌های فرعی کاملاً کانولوشن تغذیه می‌شوند که هم کلاس و هم مکان اشیاء مختلف را در تصویر تخمین می‌زنند. آشکارساز معماری RetinaNet از لنگرها به عنوان نقطه شروع برای تخمین جعبه مرزی استفاده می‌کند. لنگرها، جعبه‌های مرزی از پیش تعریف‌شده با مقیاس‌ها و نسبت‌های مختلف هستند. بنابراین هر دو زیرشبکه یک خروجی برای هر جعبه مرزی تولید می‌کنند. شبکه فرعی پیش‌بینی کلاس دارای ابعاد خروجی W*H*K*A است که در آن K تعداد کلاس‌ها، A تعداد لنگرها، W و H نشان دهنده عرض و ارتفاع تصویر خروجی است. در این پژوهش K=2، (کلاس پس‌زمینه و کلاس خودرو سواری) است. شبکه فرعی تخمین مکان دارای ابعاد خروجی W*H*4*A است که در آن چهار پارامتر برای هر لنگر، انحرافات برای قرار دادن مجدد لنگر بر روی شیئ شناسایی شده‌است.

زیرشبکه طبقه‌بندی، یک شبکه کاملاً پیچیده است که به هر سطح از شبکه هرمی ویژگی متصل است، این شبکه شامل چهار لایه کانولوشن ۳*۳، یک تابع فعال‌سازی واحد خطی همسوشده (ReLU[۱۲])، به دنبال آن یک لایه کانولوشن ۳*۳ دیگر و در نهایت تابع سیگمایی[۱۳] است. هدف آن پیش‌بینی وجود شیئ در یک موقعیت خاص در تصویر است.

زیرشبکه رگرسیون، یک شبکه کاملاً پیچیده است که موقعیت یک شیئ را با محاسبه افست بین لنگر و شیئ واقعی پیش‌بینی می‌کند. طراحی شبکه با زیرشبکه طبقه‌بندی یکسان است، و تنها تفاوت آن تابع فعال‌سازی نهایی است که در این مورد، یک تابع خطی است [۱۱].

۳-۲- تابع ضرر کانونی

کمک اصلی به تشخیص شیئ، توسط تابع ضرر کانونی[۱۴] انجام می‌شود [۱۱]. تابع ضرر کانونی مشکل عدم تعادل کلاس طبیعی را که در تشخیص شیئ وجود دارد، بررسی می‌کند، جایی که رایج‌ترین کلاس، کلاس پس‌زمینه بی‌اهمیت است. برای رسیدگی به عدم تعادل بین کلاس‌های پس‌زمینه و پیش‌زمینه، تابع ضرر کانونی، سهم زیان را از نمونه‌های با اعتماد بالا، کاهش می‌دهد. در نتیجه باعث می‌شود شبکه از نمونه‌های سخت بیشتر بیاموزد و نمونه‌های آسان را نادیده بگیرد. تابع زیان کانونی در رابطه (۱) بیان شده‌است:

$$Focal\ Loss(Pt) = \alpha_t(1-Pt)^{\gamma}\log Pt \qquad (۱)$$

اگر شیئ در کلاس واقعی خودش تشخیص داده شده باشد، Pt برابر مقدار احتمال تشخیص درست یعنی P است. اگر کلاس اشتباه تشخیص داده شده باشد، Pt برابر 1-P در نظر

۱۲ Rectified Linear Unit
۱۳ Sigmoid

۱۴ Focal loss function

گرفته می‌شود. α یک پارامتر متعادل‌کننده کلاس است که به طور تجربی برای حالتی که برابر با دو تنظیم شود، بهینه است. γ پارامتری است که با استفاده از اعتبارسنجی تنظیم می‌شود [۱۱].

بطور کلی برای اشیایی که به درستی طبقه‌بندی شده‌اند، Pt مقدار بالایی دارد، بنابراین ارزش خطای کانونی کوچک است. به این معنی که خطای گرادیان‌ها کم است، بنابراین شبکه نباید با در نظر گرفتن این نمونه‌ها وزن‌ها را تغییر دهد. شبکه باید از نمونه‌هایی با مقدار خطای زیاد درس بگیرد و سپس خطا را بر این اساس کاهش دهد.

۳-۳- برآورد دقت یادگیری

مدل در هر دوره یادگیری با استفاده از توابع دقت و ضرر اعتبارسنجی می‌شود. تابع دقت نشان‌دهنده دقت مدل در طبقه‌بندی تصاویر اعتبارسنجی است، در حالی که تابع ضرر نشان‌دهنده عدم دقت پیش‌بینی توسط مدل است. اگر یادگیری مدل موفقیت‌آمیز باشد، مقدار تابع ضرر، کم و مقدار تابع دقت، زیاد است. با این حال، اگر مقدار تابع ضرر در طول یادگیری زیاد شود، نشان‌دهنده بیش برازش[۱۵] است. معیار دقت[۱۶] و معیار یادآوری یا حساسیت[۱۷] رایج‌ترین شاخص‌های ارزیابی برای تشخیص اهداف در یادگیری عمیق هستند. هر چه مقدار این معیارها بالاتر باشد، توانایی پیش‌بینی قوی‌تر است. فرمول‌های محاسبه دقت تشخیص و معیار یادآوری به ترتیب در معادلات (۲) و (۳) نشان داده شده است:

$$P = \frac{TP}{TP + FP} \qquad (۲)$$

$$R = \frac{TP}{TP + FN} \qquad (۳)$$

مثبت واقعی (TP)[۱۸]، منفی کاذب (FN)[۱۹] و مثبت کاذب (FP)[۲۰] بر اساس ماتریس ابهام محاسبه می‌شوند که تعداد تشخیص‌های مختلف را نشان‌می‌دهند. TP تعداد اهدافی را نشان می‌دهد که به درستی شناسایی شده‌اند و FP تعداد اهدافی را نشان می‌دهد که هدف نبوده‌اند، اما به اشتباه به عنوان هدف شناسایی شده‌اند. FN تعداد اهدافی را نشان می‌دهد که هدف هستند، اما شناسایی نشده‌اند [۱۸]. معیار F1-Score یک معادل بین این دو پارامتر ارزیابی مدل است و به عنوان میانگین هارمونیک آن‌ها تعریف می‌شود (معادله ۴):

$$F1_{Score} = 2 \times \frac{P \times R}{P + R} \qquad (۴)$$

معیار میانگین دقت (AP) برابر میانگین مقادیر معیار یادآوری(حساسیت) محاسبه شده برای بازه بین صفر و یک پارامتر اشتراک بر اجتماع (IoU[۲۱]) می‌باشد.

۴- نتایج و بحث

در این پژوهش ابتدا تصاویر ماهواره‌ای گوگل ارث برای منطقه مورد مطالعه در شهر اهواز از سامانه گوگل ارث با بالاترین بزرگنمایی به صورت زمین‌مرجع دانلود شده‌است. بر روی تصویر ماهواره‌ای موقعیت خودروها با جعبه‌های مرزی (اشکال مستطیل) برای داده‌های آموزشی و اعتبارسنجی مشخص شده‌است. سپس تصاویر ماهواره‌ای با ابعاد ۱۲۸ در ۱۲۸ پیکسل و با گام ۶۴ پیکسل بریده شده‌است و تصاویری با ابعاد ۱۲۸ در ۱۲۸ پیکسل تولید شده‌است. ابعاد انتخابی پس از بررسی و اجرای مدل با درنظر گرفتن سرعت یادگیری و دقت انتخاب شده‌است. با توجه به اینکه شناسایی اهداف کوچک از تصویر دارای چالش‌های متعددی می‌باشد، سعی گردیده‌است مدلی انتخاب و ارائه گردد که بتواند نتایج بهینه‌ای ارائه دهند. مدل پیشنهای بر اساس معماری آشکارساز RetinaNet می‌باشد. این مدل‌ها در سه حالت مختلف بر اساس شبکه‌های یادگیری عمیق باقیمانده با تعداد لایه‌های ۱۸ (ResNet18)، ۳۴ (ResNet34) و ۵۰ (ResNet50) به عنوان ستون فقرات معماری آشکارساز RetinaNet در ۵۰ دوره تکرار پردازش شده‌اند. موقعیت خودروهای شناسایی شده با ضریب اطمینان بین صفر تا یک محاسبه می‌شود. ضریب اطمینان برای داده‌های اعتبارسنجی بزرگتر از ۰/۶ در نظر گرفته شده است، پس امتیاز زیر ۰/۶ به این معنی است که لنگر حاوی یک شیئ مرتبط (در این پژوهش خودرو) نیست، بلکه فقط اطلاعات پس زمینه را در بر می‌گیرد. هر سه مدل توسط داده‌های آموزشی، آموزش داده شده‌اند. جدول (۱) زمان پردازش مدل‌های پیشنهادی به همراه میانگین دقت

۱۵ Over fitting
۱۶ Precision
۱۷ Recall
۱۸ True Positive
۱۹ False Negative
۲۰ False Positive
۲۱ Intersetion over Union

داده‌های اعتبارسنجی را بر اساس حد آستانه ۰/۱ برای پارامتر اشتراک بر اجتماع و ضریب اطمینان بزرگتر از ۰/۵ نشان می‌دهد. همانطور که مشاهده می‌شود با افزایش تعداد لایه‌ها در شبکه زمان پردازش برای یادگیری مدل افزایش یافته‌است. بطورکلی هرچه تعداد لایه‌ها بیشتر باشد، شبکه عمیق‌تر شده و زمان یادگیری بیشتر خواهد شد. میانگین دقت آموزش هر سه مدل بیش از ۷۰ درصد بوده‌است. مدل پیشنهادی Stuparu و همکاران بر مبنای آشکارساز RetinaNet دارای دقت مشابه بوده‌است [۱۴]. جدول (۲) میانگین دقت برای مدل‌های پیشنهادی در ۵۰ دوره تکرار را نشان می‌دهد. تعداد تکرار تا زمانی‌که روند آموزش می‌تواند بهبود یابد، ادامه می‌یابد. در خصوص مدل‌های پیشنهادی تعداد تکرار کمتر در آموزش تا ۳۰ دوره‌ی تکرار تاثیر زیادی بر روی نتایج نخواهد داشت.

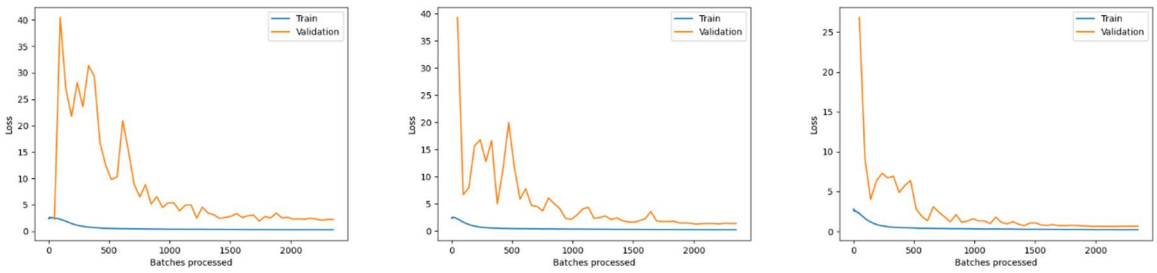
جدول ۱- زمان آموزش و میانگین دقت آموزش برای مدل‌های پیشنهادی

نام مدل	زمان پردازش (ساعت)	دقت
RetinaNet - ResNet18	۷/۰	۰/۷۳
RetinaNet – ResNet34	۷/۵	۰/۷۴
RetinaNet – ResNet50	۹/۵	۰/۷

جدول ۲- میانگین دقت در دوره‌های آموزش برای مدل‌های پیشنهادی

تعداد تکرار/ نام مدل	۱	۵	۱۰	۱۵	۲۰	۲۵	۳۰	۳۵	۴۰	۴۵	۵۰
RetinaNet –ResNet18	۰/۰۲۶	۰/۴۸۳	۰/۶۶۲	۰/۶۸۱	۰/۷۰۵	۰/۷۰۴	۰/۷۱۲	۰/۷۲۰	۰/۷۲۸	۰/۷۲۱	۰/۷۲۵
RetinaNet –ResNet34	۰/۲۸۱	۰/۵۹۵	۰/۶۸۲	۰/۷۱۴	۰/۷۳۴	۰/۷۳۳	۰/۷۳۶	۰/۷۴۹	۰/۷۴۲	۰/۷۴۴	۰/۷۴۵
RetinaNet –ResNet50	۰/۳۵۳	۰/۵۱۷	۰/۶۵۶	۰/۷۰۸	۰/۷۰۳	۰/۷۲۸	۰/۷۱۰	۰/۷۳۱	۰/۷۱۵	۰/۷۰۶	۰/۷۰۴

شکل (۶) روند یادگیری مدل بر مبنای تابع ضرر را نشان می‌دهد که بیان‌کننده آن است که به دنبال افزایش تعداد دوره‌های آموزشی، مقادیر دقت افزایش و مقادیر تابع ضرر کاهش یافته‌است. این نتایج نشان می‌دهد که همه مدل‌ها بیش از حد به مجموعه داده‌ها برازش نشده‌اند و با موفقیت ویژگی‌های تصاویر قطعه شده را یاد گرفته‌اند.



RetinaNet - ResNet18 RetinaNet - ResNet34 RetinaNet - ResNet50

شکل۶- روند یادگیری مدل‌ها در دوره‌های تکرار

۴-۱- ارزیابی مدل‌ها بر روی داده‌های تست

برای ارزیابی دقت مدل‌های پیشنهادی تصاویر ماهواره‌ای با قدرت تفکیک مکانی بالا از مناطق مختلف شهر اهواز حاوی بیش از ۱۵۰۰۰ خودرو از گوگل ارث دانلود گردیده‌است. موقعیت خودروها به صورت بصری و در محیط نرم‌افزار ArcGIS ترسیم و در پایگاه داده مکانی ذخیره گردید. شکل (۷) نمونه‌ای از تصاویر ماهواره‌ای دانلود شده برای تست مدل‌ها را نشان می‌دهد. مدل‌های آموزش دیده بر روی تصاویر ماهواره‌ای تست اجرا شده‌است تا موقعیت خودروها را استخراج نماید. مدل‌ها مکان خودرو را با ضریب اطمینان بین صفر تا یک محاسبه می‌کنند. موقعیت هر خودرو با یک کادر مستطیلی بر روی تصویر ماهواره‌ای ژئورفرنس شده، نشان داده می‌شود. از بین قطعاتی که دارای همپوشانی زیادی هستند، قطعاتی که دارای ضریب اطمینان بالاتری هستند انتخاب و بقیه حذف می‌شوند. برای اینکار از روش سرکوب غیر حداکثری (NMS[۲۲]) برای حذف قطعه‌های تکراری که روی همان شیء همپوشانی دارند، استفاده می‌شود [۱۹].

۲۲ Non-Maximum suppression

پارامتر امکان همپوشانی روش سرکوب غیرحداکثری ۲۵ درصد اعمال شده‌است. زمان انجام پردازش برای مدل دارای ۱۸ لایه برابر ۱۸ دقیقه، ۳۴ لایه برابر ۲۰ دقیقه و ۵۰ لایه برابر ۲۷ دقیقه بوده‌است.

برای برآورد معیارهای دقت، بازیابی و F1، تعداد مثبت واقعی (موقعیت خودروهایی را که به درستی شناسایی شده‌اند)، تعداد مثبت کاذب (اهدافی را که خودرو نبوده‌اند، اما به اشتباه به عنوان خودرو شناسایی شده‌اند) و تعداد منفی کاذب (اهدافی را که خودرو هستند، اما شناسایی نشده‌اند)، محاسبه شده‌است.

معیارهای ارزیابی شامل دقت، میانگین دقت، بازیابی و F1 برای داده‌های تست در جدول (۳) نشان داده شده‌است.

جدول ۳- برآورد معیارهای ارزیابی دقت برای داده‌های تست

نام مدل	دقت	بازیابی	معیار F1	میانگین دقت
RetinaNet ResNet18	۰/۵۳	۰/۸۷	۰/۶۶	۰/۵۳
RetinaNet ResNet34	۰/۶۷	۰/۹۹	۰/۸۰	۰/۷۶
RetinaNet ResNet50	۰/۷۰	۰/۹۹	۰/۸۲	۰/۸۶

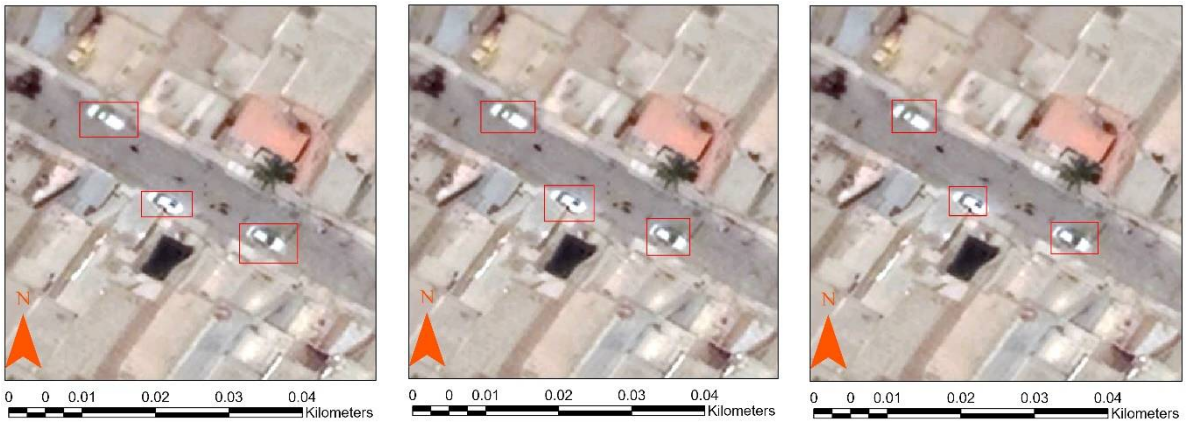
ارزیابی نتایج بر روی داده‌های تست نشان می‌دهد که با افزایش لایه‌های مدل، دقت افزایش یافته است. مقدار بالای معیار بازیابی نشان می‌دهد که مدل‌ها در شناسایی موقعیت خودروها موفق عمل می‌کنند. هرچند معیار دقت نشان می‌دهد که تعداد اشیا شناسایی شده اشتباه به عنوان خودرو می‌تواند تاثیر منفی در برآورد دقت نهایی داشته باشد. معمولا اشیا شناسایی شده اشتباه به جای خودرو خارج از محدوده شبکه معابر بوده و در صورت داشتن شبکه معابر می‌توان آن‌ها را به صورت خودکار شناسایی و حذف نمود و در نتیجه معیار دقت هم افزایش خواهد یافت. نتایج نشان می‌دهد که معماری RetinaNet با شبکه باقیمانده ۵۰ لایه دارای دقت بالاتری نسبت به حالت ۱۸ و ۳۴ لایه است.

شکل ۷- نمونه‌ای از تصاویر ماهواره‌ای برای تست مدل‌ها

برای ارزیابی دقیق رفتار مدل‌های پیشنهادی سه حالت متفاوت زیر با توجه به تراکم خودرو در نظر گرفته شده‌است. هدف ارزیابی دقیق نقاط ضعف مدل در شرایط متفاوت است.

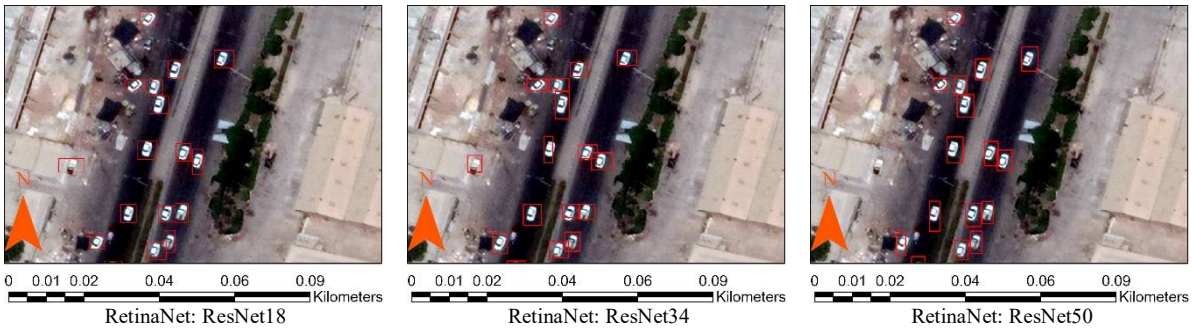
- ✓ معابر با تراکم کم
- ✓ معابر با تراکم متوسط
- ✓ معابر با تراکم بالا

شکل (۸) نتیجه اعمال مدل‌های پیشنهادی بر روی یک قسمت از تصویر ماهواره‌ای (داده تست) حاوی معابر با تراکم کم خودرو (در معابر شهری یا برون‌شهری که فاصله خودروها از هم زیاد است) را نشان می‌دهد. فقط نتایجی که دارای ضریب اطمینان بالا بوده‌اند، برای نمایش نهایی موقعیت خودروها استفاده شده‌اند.



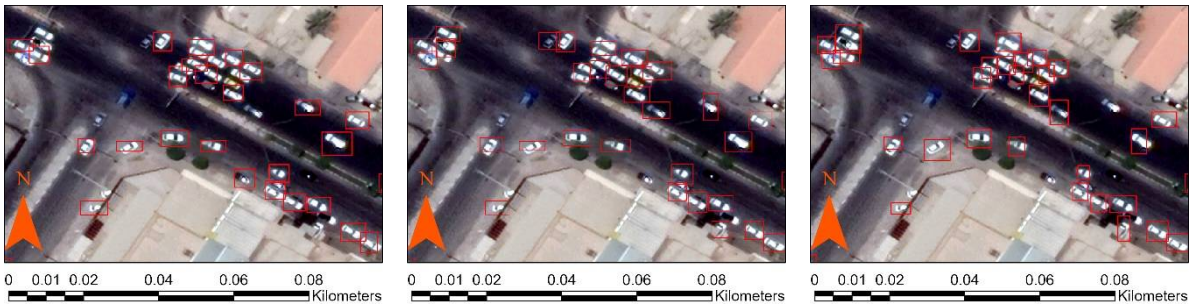
شکل ۸- نتیجه تست مدل‌ها بر روی نمونه‌ای از تصاویر ماهواره‌ای حاوی معابر شهری با تراکم کم خودرو

نتایج اعمال مدل‌ها برروی تصاویر ماهواره‌ای دارای خودرو با تراکم پایین نشان می‌دهد که مدل‌های پیشنهادی بر مبنای معماری آشکارساز RetinaNet با ستون فقرات شبکه‌های عصبی باقیمانده با تعداد لایه‌های ۱۸، ۳۴ و ۵۰ در تشخیص خودرو یکسان عمل کرده و قابلیت تشخیص بیشتر خودروهای سواری آموزش داده‌شده از تصاویر ماهواره‌ای گوگل ارث را دارند. شکل (۹) نتیجه اعمال مدل‌ها بر روی تصاویر ماهواره‌ای حاوی معابر شهری با تراکم متوسط خودرو را نشان می‌دهد. نتایج نشاندهنده دقت بالا در شناسایی موقعیت خودروها برای همه مدل‌ها می‌باشد. در معابر شهری با تراکم متوسط، هر سه مدل موفق عمل کرده‌اند.

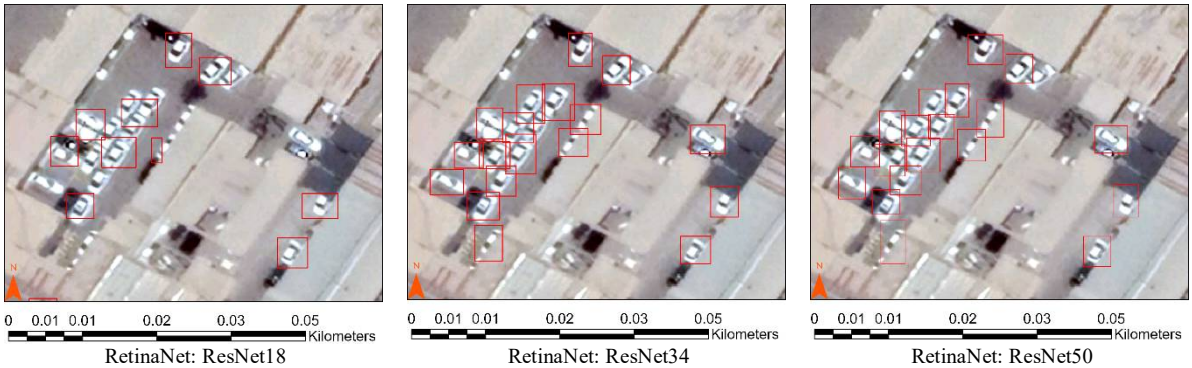


شکل۹- نتیجه تست مدل‌ها بر روی نمونه‌ای از تصاویر ماهواره‌ای حاوی معابر شهری با تراکم متوسط خودرو

شکل (۱۰) نتیجه اعمال مدل‌ها بر روی تصاویر ماهواره‌ای حاوی معابر شهری با تراکم زیاد خودرو مانند چهارراه‌ها یا ترافیک سنگین را نشان می‌دهد. نتایج نشاندهنده دقت قابل قبول مدل در شناسایی موقعیت خودروها برای همه مدل‌ها می‌باشد. در مناطقی که تراکم خودرو زیاد است معماری مبتنی بر روش‌های ۳۴ و ۵۰ لایه نسبت به حالت ۱۸ لایه دقت بهتری دارند. با توجه به نزدیکی خودروها در این حالت امکان تشخیص دقیق تعداد خودروها وجود ندارد، هرچند به ترتیب معماری مبتنی بر ۵۰ لایه نسبت به ۳۴ لایه و به همین ترتیب ۳۴ لایه نسبت به ۱۸ لایه عملکرد بهتری دارند.



شکل۱۰- نتیجه تست مدل‌ها بر روی نمونه‌ای از تصاویر ماهواره‌ای حاوی معابر شهری با تراکم زیاد خودرو



شکل۱۱- نتیجه تست مدل‌ها بر روی نمونه‌ای از تصاویر ماهواره‌ای حاوی پارکینگ روباز

شکل (۱۱) نتایج عملکرد مدل‌های پیشنهادی معماری RetinaNet با ستون فقرات مبتنی بر شبکه‌های باقیمانده ۱۸، ۳۴ و ۵۰ لایه، را بر روی مکان‌های پارکینگ خودرو را نشان می‌دهد. با توجه به نزدیکی خودروها و قدرت تفکیک مکانی ۰/۵ متری تصاویر ماهواره‌ای دانلود شده، امکان تشخیص خودروها کاهش یافته‌است. با توجه به افزایش لایه‌های معماری مدل عمیق‌تر و حساسیت مدل در تشخیص خودروهای ناقص افزایش یافته‌است. لذا در محل پارکینگ خودروها، حتی خودروهایی که قسمتی از آن‌ها دیده می‌شود، توسط مدل ۳۴ و ۵۰ لایه شناسایی شده‌اند. در صورتی که فاصله خودروها خیلی کم باشد، دقت مدل ۵۰ لایه بهتر است.

۵- نتیجه‌گیری و پیشنهادات

شناسایی اهداف کوچک از تصاویر ماهواره‌ای بدلیل نسبت ابعاد این اهداف به قدرت تفکیک مکانی تصاویر ماهواره‌ای یکی از چالش‌های مهم سنجش از دور می‌باشد. در این بین شناسایی تعداد و موقعیت خودروها در داخل شهرها و جاده-های بین شهری جهت کمک به مدیریت ترافیک و پیش‌بینی نحوه گسترش شبکه حمل و نقل ضروری است. تصاویر ماهواره‌ای با قدرت تفکیک مکانی بالا چون در محدوده وسیع و ارزانتر در دسترس هستند، می‌توانند نقش مهمی در شناسایی موقعیت خودروها داشته و متعاقب آن می‌توان با استفاده از سیستم‌های اطلاعات مکانی تراکم و تحلیل شبکه ترافیک را در سطح مکانی بزرگتر انجام داد. نتایج می‌توانند در توسعه شهری و حتی شبکه حمل و نقل بین شهری استفاده شود. در این پژوهش با توجه به در دسترس بودن تصاویر ماهواره‌ای گوگل ارث از این تصاویر استفاده شده است، هرچند چالش استخراج خودروها بدلیل کیفیت این تصاویر افزایش می‌یابد. با توجه به توانایی شبکه‌های عصبی یادگیری عمیق در شناسایی اشیا از تصاویر، در این پژوهش برای استخراج خودروها از شبکه‌های عصبی یادگیری عمیق مبتی بر معماری آشکارساز RetinaNet بر مبنای شبکه‌های عصبی باقیمانده با تعداد لایه ۱۸، ۳۴ و ۵۰ استفاده گردید.

نتایج حاصل از آموزش مدل‌های پیشنهادی نشان می-دهد که میانگین دقت آموزش برای هر سه مدل بیشتر از ۰/۷ برای داده‌های آموزشی بوده‌است. آموزش مدل در ۵۰ دوره تکراری با ضریب اطمینان بالای ۰/۶ و پارامتر اشتراک بر اجتماع برابر ۰/۱ انجام شده‌است. روند تغییرات تابع ضرر نشان می‌دهد که مدل‌ها به خوبی به داده‌های آموزشی برازش داده شده‌اند. دقت مدل با لایه‌های بیشتر در فرآیند آموزشی اندکی بیشتر از مدل با لایه‌های کمتر بوده است، هر چند زمان پردازش مدل با لایه‌های بیشتر از مدل با لایه‌های کمتر، بیشتر بوده‌است.

نتایج بر روی داده‌های تست نشان‌دهنده عملکرد مناسب مدل پیشنهادی آشکارساز RetinaNet با شبکه عصبی باقیمانده ۵۰ لایه می‌باشد. مدل ۵۰ لایه از نظر معیار میانگین دقت با ۰/۸۶، معیار دقت با ۰/۷، معیار بازیابی با ۰/۹۹ و معیار F1 با ۰/۸۲ بهترین عملکرد را داشته‌است. مدل ۱۸ لایه نیز از نظر معیار میانگین دقت با ۰/۵۳، معیار دقت با ۰/۵۳، معیار بازیابی با ۰/۸۷ و معیار F1 با ۰/۶۶ کمترین دقت را داشته‌است. مدل ۳۴ لایه نیز دقت بهتری از مدل ۱۸ لایه داشته‌است. بنابراین می‌توان نتیجه‌گیری نمود که روش ۵۰ لایه نسبت به دو مدل با لایه کمتر (۱۸ و ۳۴ لایه) عملکرد بهتری دارد. از نظر زمان پردازش هر چند مدل ۱۸ لایه دارای زمان کمتری است، اما دقت پایین‌تری هم دارد. اما مدل ۳۴ لایه دارای دقت نسبتا مناسب و زمان پردازش مناسبی است، لذا در مواردی که زمان پردازش اهمیت دارد می‌توان از مدل ۳۴ لایه استفاده نمود.

ارزیابی عملکرد مدل‌ها بر اساس تراکم خودور نشان داد که هر سه مدل در استخراج خودروها از تصاویر ماهواره‌ای که خودروها از هم فاصله دارند و متراکم نیستند، می‌توانند نتایج مطلوبی ارائه نمایند. اما در مناطق پرترافیک مانند توقفگاه‌ها و پشت چراغ قرمزها که خودروها بدلیل فاصله نمونه‌برداری زمینی تصاویر ماهواره‌ای بسیار نزدیک به هم و در موازی چسپیده هستند، مدل ۵۰ لایه بهتر عمل می‌کند. چالش اصلی در مناطق دارای تراکم بالای خودور می‌باشد، که امکان تشخیص دقیق تعداد خودروها را کاهش می‌دهد اما سطح اشغال را بهتر برآورد می‌کند. لذا در صورتیکه هدف برآورد سطح اشغال باشد، دقت مدل بیشتر خواهد بود. همچنین در مواردی که فقط بخشی از خودرو در تصاویر ماهواره‌ای دیده می‌شود، مدل ۵۰ لایه عملکرد بهتری نشان می‌دهد. بطور کلی می‌توان نتیجه گرفت که می‌توان از شبکه‌های عصبی آشکارساز Retina Net بر اساس شبکه‌های یادگیری عمیق باقیمانده برای استخراج خودرو از تصاویر ماهواره‌ای گوگل ارث با دقت مناسب استفاده نمود. استفاده از شبکه‌های یادگیری عمیق با ۵۰ لایه با میانگین دقت ۸۶ درصد نتایج رضایت‌بخشی دارد. کیفیت تصاویر مورد استفاده در این

مطالعه با توجه به اینکه از سامانه گوگل ارث دانلود شده‌است پایین بوده و علت کاهش دقت در مقایسه با نتایج استوپارو و همکاران که به دقت ۹۳ درصد رسیده‌اند [۱۴]، ناشی از اختلاف کیفیت تصاویر ماهواره‌ای مورد استفاده است. همچنین منصور[۲۳] و همکاران نشان دادند که میانگین دقت تشخیص خودرو از تصاویر ماهواره‌ای با استفاده از روش Faster R-CNN برابر ۸۹ درصد و روش SSD برابر ۸۴ درصد است. زمان پردازش روش SSD نسبت به روش روش Faster R-CNN کمتر است [۲۰]. همچنین، نتایج پژوهش نشان می‌دهد که می‌توان با استفاده از شبکه عمیق‌تر به نتایج مطلوب‌تری دست یافت. در این پژوهش از سایر روش‌های شبکه‌های عصبی یک مرحله‌ای مانند SSD و YOLO هم استفاده گردید، اما نتایج ضعیفی داشتند و از روند پژوهش خارج شدند. بدیهی است در صورت استفاده از تصاویر ماهواره‌ای با کیفیت بالاتر یا تصاویر پهپاد می‌توان به نتایج دقیق‌تری دست یافت.

مراجع

[۱] Benjdira, B., Khursheed, T., Koubaa, A., Ammar, A. and Ouni, K. (2018) "Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3." arXiv preprint arXiv:10968/1812.

[۲] Duarte, D., Nex, F., Kerle, N. and Vosselman, G. (2018) "Satellite Image Classification of Building Damages using Airborne and Satellite Image Samples in a Deep Learning Approach." Remote Sensing and Spatial Information Sciences, Riva del Garda, Italy, 4–7, Volume IV-2, PP. 89–96.

[۳] Audebert, N., Le Saux, B. and Lef evre, S. (2017) "Segment-before detect: Vehicle detection and classification through semantic segmentation of aerial images." Remote Sens., vol. 9, no. 4, PP. 368.

[۴] Yu, Y., Gu, T., Guan, H., Li, D. and S. Jin, (2019) "Vehicle detection from high-resolution remote sensing imagery using convolutional capsule networks." IEEE Geosci. Remote Sens. Lett., vol. 16, no. 12, PP. 1894–1898.

[۵] Mattyus, G., Wang, S., Fidler, S. and Urtasun, R. (2015) "Enhancing roads maps by parsing aerial images around the world." In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; PP. 1689–1697.

[۶] Chen, G., Wang, H., Chen, K., Li, Z., Song, Z., Liu, Y., Chen, W. and Knoll, A. (2020) "A survey of the four pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal." IEEE Transactions on Systems, Man, and Cybernetics: Systems.

[۷] Girshick, R., Donahue, J., Darrell, T., Malik, J. and Malik, J. (2014) "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28; PP. 580–587.

[۸] Ren, S., He, K., Girshick, R.B. and Sun, J. (2015) "Faster R-CNN: towards real-time object detection with region proposal networks." corr abs/01497/1506, arXiv preprint arXiv:01497/1506.

[۹] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, Y. and Berg, A. C. (2016) "SSD: Single shot multibox detector." in European conference on computer vision. Springer, PP. 21–37.

[۱۰] Redmon, S., Divvala, R., Girshick, A. and Farhadi, A. (2016) "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, PP. 779–788.

[۱۱] Lin, T.Y., Goyal, P., Girshick, R., He, K. and Doll´ar, P. (2017) "Focal loss for dense object detection," in Proceedings of the IEEE international conference on computer vision, PP. 2980–2988.

[۱۲] Yang, M.Y., Liao, W., LI, X. and Rosenhan, B. (2018) "Deep learning for vehicle detection in aerial images." In: Proceedings of the IEEE International Conference on Image Processing (ICIP). PP. 3079–3083.

[۱۳] Douillard, A. (2020) "Detecting Cars from Aerial Imagery for the NATO Innovation Challenge." Available online: https://arthurdouillard.com/post/nato-challenge/.

[۱۴] Stuparu, D.G., Ciobanu, R.I. and Dobre, C. (2020) "Vehicle Detection in Overhead Satellite Images Using a One-Stage Object Detection Model." Sensors, 20, 6485. https://doi.org/3390/10/s20226485

۲۳ Mansour

[۱۵] Van Etten, A. (2019) “Satellite imagery multiscale rapid detection with windowed networks.” In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 7–11, PP. 735–743.

[۱۶] He, K., Zhang, X., Ren, S. and Sun, J. (2015) “Deep residual learning for image recognition.” arXiv preprint arXiv:03385/1512, 2016.

[۱۷] Lin, T.Y., Doll´ar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) “Feature pyramid networks for object detection.” in Proceedings of the IEEE conference on computer vision and pattern recognition, PP. 2117–2125.

[۱۸] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Doll´ar P. and Zitnick, C.L. (2014) “Microsoft coco: Common objects in context”, In European Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, pp. 740–755.

[۱۹] Ren, S., He, K., Girshick, R. and Sun, J. (2017) “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 1137-1149. https://doi.org/1109/10/TPAMI.2577031/2016

[۲۰] Mansour, A., Hassan, A., Hussein, W. and Said, E. (2019) “Automated vehicle detection in satellite images using deep learning”, 18th International Conference on Aerospace Sciences & Aviation Technology, 610(1):012027. doi:10.1088/1757-899X/610/1/012027