نشریه علمی علوم و فنون نقشه برداری، دوره یازدهم، شماره ۱، شهریور ماه ۱۴۰۰ - مقاله پژوهشی

توسعه مدلی مبتنی بر تشدید گرادیان در شبکه‌های کانولوشنی عمیق به‌منظور شناسایی اهداف در تصاویر سنجش‌ازدوری

نیما فرهادی*[۱]، عباس کیانی[۲]، حمید عبادی[۳]

[۱] دانشجوی کارشناسی ارشد فتوگرامتری و سنجش‌ازدور - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی
farhadinima75@email.kntu.ac.ir

[۲] استادیار دانشکده عمران - دانشگاه صنعتی نوشیروانی بابل
a.kiani@nit.ac.ir

[۳] استاد دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه‌نصیرالدین طوسی
ebadi@kntu.ac.ir

(تاریخ دریافت مرداد ۱۳۹۸، تاریخ تصویب اردیبهشت ۱۴۰۰)

چکیده

پیشرفت‌های صورت گرفته در فناوری تصویربرداری ماهواره‌ای امکان تهیه اطلاعات متنوع برای شناسایی اهداف را فراهم می‌کند. چنین اطلاعاتی فرآیند تفسیر تصاویر سنجش‌ازدوری نوری را تسهیل می‌بخشد. نوع خاصی از این تفاسیر به فعالیت‌های مربوط به شناسایی اهداف ختم می‌شود که امروزه اکثر تحقیقات انجام‌شده در این حوزه با استفاده از شبکه‌های عصبی و تکنیک‌های یادگیری عمیق صورت می‌گیرد. نحوه طراحی شبکه عصبی کانولوشن مورداستفاده، در دقت شناسایی نقش بسزایی دارد. تحقیقات اخیر درزمینه یادگیری عمیق و شبکه‌های کانولوشن نشان می‌دهد که عمیق‌تر کردن این شبکه‌ها باعث افزایش دقت آن‌ها می‌شود؛ اما گاهی بیش‌ازحد عمیق‌تر شدن باعث به وجود آمدن مشکلاتی ازجمله بالا رفتن پارامترهای آموزشی، محو شدن گرادیان آموزشی، بلااستفاده ماندن بسیاری از ویژگی‌های تولیدشده و... می‌شود که در پی آن کاهش دقت در شناسایی اهداف موردنظر را خواهد داشت. به این منظور در این تحقیق روشی توسعه داده‌شده است که در آن سعی گردید با حفظ ویژگی‌های تولیدشده توسط لایه‌های کانولوشن و انتقال آن‌ها به لایه‌های بعدی، بر این مشکل غلبه گردد. این نوع ارتباط بین لایه‌ها، اجازه عمیق‌تر کردن شبکه‌های کانولوشنی با افت گرادیان کمتر را می‌دهد. معماری ارائه‌شده علاوه بر کم‌رنگ کردن مشکل ناپدید شدن گرادیان، باعث می‌شود تعداد پارامترها و همچنین مدت‌زمان موردنیاز برای آموزش یک مدل یادگیری عمیق کاهش یابد. بدین منظور در ابتدا با استفاده از تصاویر سنجش‌ازدوری، مجموعه‌ای از داده‌های آموزشی آماده و پس از پردازش‌های اولیه، عوارض هدف برچسب‌گذاری شده است. سپس روش پیشنهادشده را به‌عنوان استخراج گر ویژگی مدل Faster R-CNN تعریف کرده و بر روی داده‌های آموزشی، آموزش داده می‌شود. جهت ارزیابی روش پیشنهادی نیز، بخشی از فرودگاه بین‌المللی پکن چین به‌عنوان مطالعه موردی اول و بخشی از فرودگاه بین‌المللی امام خمینی (ره) به‌عنوان منطقه موردمطالعه دوم انتخاب‌شده است و مقادیر معیار F1-Measure برای هر دو منطقه به ترتیب برابر ۹۷/۹ و ۹۳/۷ می‌باشد. درنهایت نتایج حاصله از اعمال مدل پیشنهادی، با مدل‌های مختلف شبکه مطرح موجود، مقایسه شده است. نتایج به‌دست‌آمده، دلالت بر قابل‌اعتماد بودن و مؤثر بودن روش ارائه‌شده دارند.

واژگان کلیدی: یادگیری عمیق، شبکه‌های کانولوشن، شناسایی اهداف سنجش‌ازدوری، استخراج‌گر ویژگی

---

* نویسنده رابط

۱- مقدمه

شناسایی اهداف در تصاویر ماهواره‌ای یکی از موضوعات تحقیقاتی است که در سال‌های اخیر موردتوجه بسیاری از محققان سنجش‌ازدوری قرارگرفته است؛ ازاین‌رو، در این حوزه، روش‌های متعددی توسعه داده‌شده است که روش‌های مربوط به یادگیری ماشین جزء قابل اعتمادترین آن‌ها هستند. در زمینه‌ی یادگیری ماشین، شبکه‌های عصبی کانولوشن یک روش پرکاربرد طی سالیان گذشته می‌باشد که در قرن بیستم میلادی برای اولین بار معرفی‌شده است [۱]. این شبکه‌ها به‌عنوان استخراج‌گر ویژگی از تصاویر رقومی در مدل‌هایی همچون Faster R-CNN[۱] [۲]، SSD[۲] [۳] و YOLO[۳] [۴] برای شناسایی اهداف مورداستفاده قرار می‌گیرند که در میان آن‌ها مدل Faster R-CNN با دقت و ضریب اطمینان بالاتری عوارض موجود در تصاویر را شناسایی می‌کند. در سال‌های اخیر تحقیقات فراوانی در زمینه‌ی بهبود Faster R-CNN در تصاویر طبیعی صورت گرفته است. به‌عنوان‌مثال Chen و Gupta [۵] از NMS[۴] در انتخاب مرزهای موردنظر از کلاسه‌بندی مرزها استفاده کردند. بااین‌حال در این روش امکان انتخاب عارضه اشتباه و یا عوارضی که در پس‌زمینه قرار دارند، به‌عنوان عارضه موردنظر وجود دارد [۶]. Li و همکاران [۷] با اعمال عملگرهای کانولوشن کوچک‌تر میزان محاسبات روش Faster R-CNN تا حد قابل‌توجهی کاهش دادند. دسته‌ای دیگر از محققان شناسایی اهداف را در زمینه‌های تخصصی‌تری، مانند شناسایی هواپیما در تصاویر سنجش‌ازدوری، انجام داده‌اند. به‌طور مثال Farhadi و همکاران [۸] در سال ۲۰۱۹ از شبکه‌های پیش آموخته و ترکیب سلسله مراتبی از چندین شبکه Faster R-CNN برای شناسایی هواپیما استفاده کرده‌اند. Wu و همکاران [۹] برای شناسایی هواپیما از یک روش مبتنی بر گرادیان باینری و نرمال شده[۵] و شبکه‌های عصبی کانولوشن استفاده کردند. Zhang و همکاران [۱۰] به‌منظور شناسایی هواپیما در تصاویر سنجش‌ازدوری، یک چهارچوب قابل‌آموزش نیمه نظارتی را با استفاده از شبکه‌های کانولوشن دوگانه، ارائه کردند. Zeng و همکاران [۱۱] در سال ۲۰۱۹ با تلفیق مفاهیم یادگیری عمیق و تحلیل‌های مکانی یک روش سلسله مراتبی را به‌منظور شناسایی هواپیما ارائه کردند. Xu و همکاران [۱۲] از ترکیب ویژگی‌های چندلایه در شبکه‌های کانولوشنی تمام متصل باهدف شناسایی هواپیما بهره گرفتند؛ اما بااین‌حال بیشتر این روش‌ها برای عوارض بزرگ‌مقیاس و با وضوح بالا مناسب هستند، اما در تصاویر با پس‌زمینه پیچیده و عوارض کوچک، دقت پایین‌تری را حاصل می‌کنند [۱۳]. از طرفی دیگر، شبکه‌های عصبی کانولوشن بکار رفته در این روش‌ها، ویژگی‌های سطح پایین استخراج‌شده از داده‌های ورودی را فراگرفته و ویژگی‌های سطح بالا را تولید می‌کنند و با تلفیق این ویژگی‌ها، الگوهای مناسبی را پدید می‌آورند که ازاین‌جهت گزینه مناسبی برای شناسایی اهداف بصری به‌حساب می‌آیند [۱۴]. عمق این شبکه‌ها تأثیر بسیار زیادی بر روی نحوه‌ی ترکیب ویژگی‌های سطح پایین دارد؛ شبکه با عمق بیشتر دارای لایه‌ها و سطوح ویژگی بیشتر است. پیشرفت‌های اخیر در زمینه‌ی قدرت پردازشی سخت‌افزارها، امکان آموزش شبکه‌های عمیق مانند VGG ۱۹ لایه [۱۵] و ResNet[۶] ۱۵۲ لایه [۱۶] را به محققان می‌دهند.

شبکه‌های عمیق‌تر از دقت طبقه‌بندی بالاتری نسبت به شبکه‌های کم عمق‌تر برخوردارند؛ اما چنانچه عمق شبکه‌ها به‌صورت قابل‌توجهی زیاد شود، مشکل جدیدی به نام محو شدن گرادیان به وجود می‌آید [۱۷]. محو شدن گرادیان زمانی رخ می‌دهد که تعداد لایه‌های کانولوشن درون شبکه افزایش یابد؛ به تعبیری دیگر با افزایش لایه‌های کانولوشن درون شبکه اطلاعات ورودی به لایه‌های بالاتر، کاهش می‌یابد و گاهی در انتهای شبکه به صفر میل می‌کند [۱۸]. مقالات زیادی که اخیراً چاپ شده‌اند به این موضوع اشاره دارند [۱۴]، [۱۹]، [۲۰]. همگی این مقالات با ارائه روش‌های مختلف سعی در حل مشکل گرادیان دارند. روش‌های موجود توپولوژی و پروسه آموزشی مختلفی دارند اما یک مفهوم اصلی را با خود به یدک می‌کشند و آن‌هم ایجاد یک مسیر میانبر برای اتصال لایه‌های قبل به لایه‌های بعد است که باهدف جلوگیری از محو شدن گرادیان دنبال می‌شود.

تحقیق در زمینه‌ی معماری‌های شبکه، قسمت جداناپذیر از تاریخچه شبکه‌های عصبی از بدو تولد تابه‌حال بوده است. تفاوت اصلی بین شبکه‌های مدرن و سنتی، افزایش تعداد لایه‌های مخفی بکار رفته در معماری و نحوه

۱ Faster Region-based Convolutional Neural Networks
۲ Single Shot multibox Detector
۳ You Look Only Once
۴ Non-Maximal Suppression
۵ BInarized Normed Gradients (BING)
۶ RESidual NETworks

اتصال آن‌ها به هم است. اولین مفاهیم به وجود آمده از شبکه‌های عصبی را می‌توان به دهه ۹۰ میلادی ارجاع داد. در سال ۱۹۹۸ LeCun و همکاران [۱] شبکه‌های عصبی کانولوشن را برای اولین بار به‌منظور دسته‌بندی اعداد مندرج در کد پستی معرفی کردند و اسم شبکه خود را LeNet گذاشتند. از شبکه عصبی آن‌ها در سرویس‌های پستی استفاده فراوانی شد اما اعداد عارضه‌های بسیار ساده‌ای هستند و شبکه LeNet قابلیت دسته‌بندی داده‌های پیچیده‌تر مانند تصاویر بصری دوبعدی را نداشت [۲۱]. به‌منظور پوشش این نقص Krizhevsky و همکاران [۲۲] در سال ۲۰۱۲ با تغییراتی اساسی در شبکه عصبی کانولوشن توانستند خطای روش‌های قبل در دسته‌بندی تصاویر بر روی مجموعه داده ImageNet [۲۳] را به میزان ۱۰ درصد کاهش دهند. Simonyan و Zisserman [۱۵] در سال ۲۰۱۴ شبکه‌ای عمیق‌تر با نام VGG را معرفی کردند. استراتژی به‌کاررفته در این معماری استفاده از فیلتر با ابعاد کوچک و تعداد لایه‌های بیشتر بوده است. در همان سال، Szegedy و همکاران [۲۴, ۲۵] بلوک‌های مفهومی[۱] را در معماری شبکه‌ها به کار بردند و شبکه GoogLeNet را با استفاده از چندین مدل مفهومی پشت سرهم تشکیل دادند. اساس طراحی این بلوک‌ها بر پایه راندمان بالا در محاسبات می‌باشد. آن‌ها پارامترهای به‌کاررفته در کل شبکه را به ۵ میلیون کاهش دادند. مادامی‌که این شبکه‌ها رفته‌رفته عمیق و عمیق‌تر می‌شدند، اندازه گرادیان نیز روبه‌زوال و صفر شدن میل می‌کرد. سرانجام در سال ۲۰۱۵، He و همکاران [۱۶] روشی را در طراحی شبکه‌ها ابداع کردند که اجازه استفاده از ۱۵۲ لایه کانولوشنی را با انتشار گرادیان مناسب، می‌دهد. آن‌ها بلوک‌های باقی‌مانده[۲] را که شامل دولایه کانولوشن بود به وجود آوردند. در این بلوک‌ها مقدار ورودی با یک عمل جمع ساده به خروجی بلوک اضافه می گردد؛ که همین امر باعث بهتر جریان یافتن گرادیان در کل شبکه می‌شود. Larsson و همکاران [۲۶] در سال ۲۰۱۶ با ترکیب خطی از چندین لایه موازی بلوک‌های Fractal را به وجود آوردند که با تکرار این بلوک‌ها می‌توان عمق شبکه را بدون استفاده از روش باقی‌مانده‌ها زیاد کرد. مفهوم اصلی بلوک‌های Fractal، انتقال مؤثر اطلاعات از لایه‌های اولیه به لایه‌های پایین‌تر از طریق میانبرهای کوتاه است. سرانجام در سال ۲۰۱۷ Huang و همکاران [۱۸] ایده‌ی جالبی را در طراحی شبکه‌های عصبی کانولوشن به کار بردند که باعث کاهش تعداد پارامترهای موردنیاز در آموزش شبکه می‌شد. آن‌ها تمام لایه‌های کانولوشن را در بلوک‌هایی به نام بلوک چگال[۳] به هم متصل کردند. با استفاده از این ایده مشکل شارش گرادیان تا حدودی حل‌شده و علاوه بر این، استفاده دوباره از ویژگی‌های استخراج‌شده‌ی لایه‌های قبل در لایه-های بعدی، باعث تفکیک‌پذیری بهتر ویژگی‌ها توسط شناساگر می‌شود.

با توجه به تحقیقات متعددی که در زمینه‌ی بهبود شبکه‌های کانولوشنی صورت گرفته است، اما همچنان این شبکه‌ها با مشکلاتی همچون هزینه پردازشی بالا و محو شدن گرادیان روبه‌رو هستند. از طرفی، به دلیل این‌که شبکه‌های CNN به‌عنوان پایه و اساس روش‌های شناسایی عارضه ایفای نقش می‌کنند، بهبود عملکرد و هزینه پردازشی موردنیاز این دسته از شبکه‌ها، باعث بهبود کارایی روش‌های شناسایی عارضه می‌گردد. در همین راستا، در تحقیق پیش رو، هدف، ارائه یک معماری جدید با نحوه اتصال متفاوت لایه‌های کانولوشن است که محو شدن گرادیان در آن به حداقل مقدار خود می‌رسد. در این معماری اطلاعات هر لایه مهم و در طول آموزش شبکه مؤثر هستند. خروجی لایه آخر هر سلول در کنار دیگر خروجی‌ها قرارگرفته و تا انتهای شبکه حفظ می‌شود.

۲- مبانی نظری تحقیق

یادگیری عمیق زیرمجموعه‌ای از الگوریتم‌های یادگیری ماشین می‌باشد که هدف آن کشف چندین سطح از بازنمودهای توزیع‌شده از داده‌ی ورودی است؛ درواقع یادگیری عمیق مفاهیم انتزاعی سطح بالای استخراج‌شده از داده‌های ورودی را در لایه‌های سطح پایین پشت سر هم مدل می‌کند [۲۷]. مکانیسم استخراج مفاهیم استفاده‌شده در مجموعه‌ای به نام شبکه عصبی انجام می‌شود. نوعی از شبکه‌های عصبی، شبکه‌های عصبی کانولوشن است که مجموعه‌ای از فیلترها را در خود جای داده است. بسته به ابعاد این فیلترها، پارامترهایی تعریف می‌شود که نیازمند بهینه‌سازی هستند. این پارامترها در طول پروسه آموزش تغییر می‌کنند تا بهترین مفاهیم را از تصاویر آموزشی

[۱] Inception Blocks
[۲] Residual Blocks
[۳] Dense Block

استخراج کنند. سپس مفاهیم از لایه‌های اول به لایه‌های بعدی منتقل می‌شود تا مفاهیم کلی تشکیل شوند.

شناسایی اهداف یکی از مسائل مهم و پرمخاطب در زمینه‌ی یادگیری عمیق است [۲۸, ۲۹] و کاربرد فراوانی در علوم مختلف نقشه‌برداری دارد. روش‌های متنوعی برای این موضوع شده است که ازجمله مهم‌ترین آن‌ها می‌توان به Faster R-CNN، SDD و YOLO اشاره کرد که همه‌ی این روش‌ها از شبکه‌های عصبی کانولوشنی به‌عنوان استخراج ویژگی استفاده می‌کنند. در این مقاله از روش Faster R-CNN برای شناسایی اهداف استفاده‌شده است که نخستین بار توسط Girshick در سال ۲۰۱۵ توسعه یافته است.

۲-۱- شناساگر مبتنی بر مرز

در زمینه بینایی ماشین یکی از مسائل پایه و بحث‌برانگیز شناسایی اهداف است. در سال ۲۰۱۶، Girshick و همکاران [۳۰] شبکه کانولوشنی را ارائه کردند که با استفاده از مرزهای مشخص‌شده در داده‌های آموزشی، اهداف موردنظر را شناسایی می‌کند. شناساگر ارائه‌شده توسط آن‌ها میانگین دقت حاصله از روش‌های قبل را تا ۵۰ درصد بهبود دادند [۳۱]. این روش را می توان ترکیبی از چندین مکانیسم مختلف شامل یک شبکه عصبی کانولوشن برای استخراج ویژگی، یک ماشین بردار پشتیبان برای دسته‌بندی و یک مکانیسم ارائه اهداف دانست [۳۲]. R-CNN با نتایج خوب خود توانست بسیاری از رقیبان خود را پشت سر بگذارد اما به دلیل افزونگی محاسبات ناشی از همپوشانی مرزهای ارائه‌شده، عملکردی کند را از خود به‌جای می‌گذارد. به‌منظور افزایش سرعت و دقت در شناسایی، Girshick و همکاران نسخه بروز شده‌ی R-CNN، به نام Fast R-CNN را ارائه کردند [۳۳]. این مدل در گام اول، عملیات برداشت ویژگی را قبل از انتخاب مرز عوارض انجام می‌دهد، به این معنی که یک شبکه کانولوشنی را به‌جای اعمال دنباله‌ای از شبکه کانولوشنی، بر روی تصویر ورودی اعمال می‌کند. در این مدل، گام دوم، جایگزین کردن یک‌لایه Softmax به‌جای مدل SVM مورداستفاده در R-CNN است؛ با این کار از تولید یک مدل جدید SVM نیز جلوگیری می‌شود. یکی از وجوه اشتراک بین R-CNN و Fast R-CNN وجود الگوریتم جستجوی انتخابی است این الگوریتم ذاتاً نیاز به عملیات محاسباتی بالایی دارد و باعث کندی شبکه می شود [۳۲]. Ren و همکاران [۲] در سال ۲۰۱۵ مدل Faster R-CNN را ارائه کردند. آن‌ها برای بهبود سرعت یادگیری، یک شبکه RPN طراحی و به‌جای الگوریتم جستجوی انتخابی استفاده کردند. نحوه عملکرد RPN به‌گونه‌ای است که پس از اعمال شبکه عصبی به تصویر ورودی، در لایه آخر همان شبکه یک ماسک با ابعاد n در n (معمولاً n برابر ۳) به نقشه ویژگی تولیدشده توسط شبکه کانولوشنی، اعمال می‌شود و ابعاد آن‌ها را کاهش می‌دهد. سپس ماسک اعمال‌شده، چندین مرز امکان‌پذیر را بر اساس محدوده‌های پیش‌فرض تولید می‌کند و درنهایت نیز مرزهایی که شامل عارضه موردنظر هستند، انتخاب و مختصاتی که توسط محدوده‌های اطراف عارضه مشخص‌شده است، ارائه می‌گردد. مدل Faster R-CNN، مقادیر به‌دست‌آمده از RPN را جایگزین مقادیر جست‌وجوی انتخابی کرده و همین موضوع باعث افزایش کارایی و سرعت شناسایی عارضه مدل مذکور می‌گردد [۲]. دقت و کارایی قابل‌قبول این مدل، دلیل اصلی انتخاب آن به‌عنوان روش شناسایی عارضه می‌باشد که در این مقاله استفاده شده است. یکی از وجوه اشتراک بین هر سه مدل، شبکه‌های عصبی کانولوشن به‌عنوان استخراج‌گر ویژگی هستند که نحوه طراحی آن در سال‌های اخیر موردتوجه بسیاری از محققان قرارگرفته است.

۲-۲- طراحی شبکه کانولوشن

فرض کنید X یک تصویر آموزشی گذرنده از شبکه کانولوشن با L لایه بوده که هرکدام از این لایه‌ها یک انتقال غیرخطی $H_L(X)$ را اعمال می‌کنند. انتقال غیرخطی $H_L(X)$ می‌تواند ترکیبی از لایه‌های کانولوشن یا عملگرهایی مانند Batch Normalization [۳۴]، Rectified Linear Unit (ReLU) [۳۵] یا Pooling [۱] باشد. خروجی هر لایه $L-$ ام نیز، X_L در نظر گرفته می‌شود.

$$X_L = H_L(X_{L-1}) \quad (1)$$

رابطه (۱) نحوه جریان اطلاعات در لایه‌های شبکه‌های سنتی را نمایش می‌دهد که هر لایه ورودی خود را از لایه قبل دریافت می‌کند. شبکه ResNet با ایجاد یک میانبر بین ورودی هر لایه و خروجی آن، انتقال غیرخطی را به نحوی دور می‌زند که منجر به انتقال بهتر اطلاعات می‌شود [۱۸].

$$X_L = H_L\left(X_{L-1}\right) + X_{L-1} \qquad (۲)$$

رابطه (۲)، معادله به‌کاررفته در شبکه ResNet می باشد. مزیت اصلی این تابع، انتقال مناسب گرادیان از لایه‌های ابتدایی به لایه‌های بعدی است؛ اما بااین‌حال ارتباط بین ورودی و خروجی تابع با استفاده از یک عمل جمع ساده انجام می‌شود که ممکن است جریان اطلاعات در شبکه را مختل کند [۱۸].

۳- روش پیشنهادی

معماری پیشنهادی ترکیبی از چند بلوک است که پشت سر هم قرارگرفته‌اند. بلوک‌ها ورودی خود را از بلوک قبل دریافت می‌کنند و محاسبات مربوطه را بر روی آن‌ها انجام می‌دهند. هر بلوک شامل چند سلول می‌شود که این سلول‌ها دارای دو لایه کانولوشن هستند. به‌منظور استفاده بهینه از تمام ویژگی‌های تصاویر آموزشی، فیلترهای بکار رفته در لایه‌های کانولوشن دارای ماسک‌هایی با اندازه‌های ۱×۱ و ۳×۳ هستند که خروجی لایه ۳×۳ در لایه ترکیب‌کننده[۱] با اطلاعات لایه‌های قبلی ادغام می‌شود. معماری شبکه‌های سنتی به‌گونه‌ای است که هر لایه کانولوشن ورودی خود را از خروجی لایه قبل دریافت می کند. این فرآیند باعث ایجاد ویژگی جدید می‌شود اما در کنار آن اطلاعات موردنیاز برای آموزش شبکه در فرآیند آموزش از بین می‌رود. معماری هر سلول در شبکه پیشنهادی به‌گونه‌ای است که تمام ویژگی‌های استخراج‌شده را در خود حفظ و نگه‌داری می‌کند تا در سلول‌های بعدی مورداستفاده قرار گیرد. رابطه (۳)، نحوه ورود اطلاعات به هر لایه کانولوشن در معماری پیشنهادی را مشخص می‌کند.

$$X_L = H_L(X_0, X_1, \dots, X_{L-1}) \qquad (۳)$$

طبق رابطه فوق، ورودی هر لایه ترکیبی از تمام خروجی لایه‌های قبلی است. همانند معماری ResNet، بلوک‌های در نظر گرفته‌شده برای شبکه پیشنهادی، ۴ بلوک بوده است که به‌منظور کاهش محاسبات و بهبود فرایند یادگیری، میان هر بلوک، از یک لایه‌ی Average Pooling با ماسک دو پیکسل در دو پیکسل و Stride برابر ۲ استفاده‌شده است تا اندازه مکانی نقش‌های ویژگی[۲] را نصف کند. جدول (۱) تفاوت بین دو معماری ResNet و معماری پیشنهادی را نمایش می‌دهد. اعداد داخل براکت نشان‌دهنده‌ی ابعاد و تعداد لایه‌های مورداستفاده در معماری می‌باشد. همچنین متغیر N معرف تعداد فیلترهای ابتداییِ لایه‌های کانولوشنی است. تعداد پارامترهای استفاده‌شده در شبکه ResNet تقریباً ۲/۵ برابر شبکه پیشنهادی است. عمق شبکه پیشنهادی نیز با در نظر گرفتن تعداد سلول‌های استفاده‌شده تعیین می‌گردد. هر سلول دارای دو لایه کانولوشن است. با فرض داشتن ۴ بلوک، $\boldsymbol{n}$ سلول در شبکه، یک لایه کانولوشن در ابتدا و یک لایه FC[۳]، شبکه شامل $2n+2$ لایه می‌گردد. جزییات بیشتر در ادامه موردبررسی قرار خواهند گرفت.

جدول ۱- ترتیب قرارگیری لایه‌ها در معماری پیشنهادی

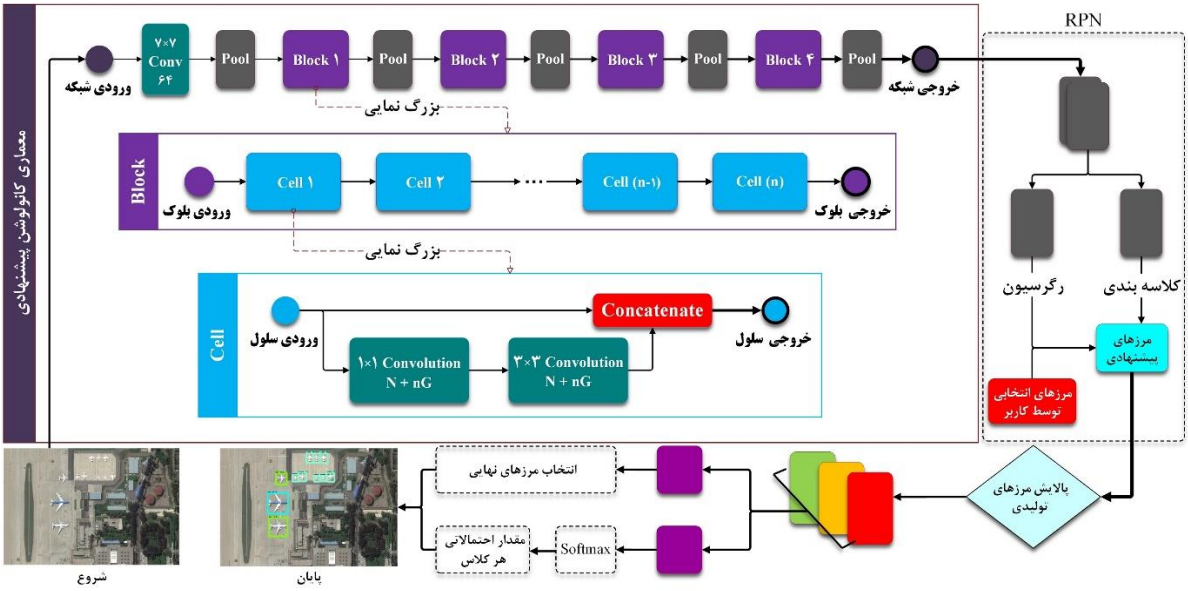
مرحله	خروجی	ResNet-50	معماری پیشنهادی ۵۰ لایه G = ۱۶
کانولوشن اول	۲۵۶×۲۵۶	Stride ۲ ،۶۴ ،۷×۷	
لایه Pooling	۱۲۸×۱۲۸	Max pool ،Stride ۲ ،۳×۳	
بلوک اول	۱۲۸×۱۲۸	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, N+nG\\3\times3, N+nG\end{bmatrix}\times4$
لایه Pooling	۶۴×۶۴	----	Average pooling ۲×۲, Stride ۲
بلوک دوم	۶۴×۶۴	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, N+nG\\3\times3, N+nG\end{bmatrix}\times6$
لایه Pooling	۳۲×۳۲	----	Average pooling ۲×۲, Stride ۲
بلوک سوم	۳۲×۳۲	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, N+nG\\3\times3, N+nG\end{bmatrix}\times10$
لایه Pooling	۱۶×۱۶	----	Average pooling ۲×۲, Stride ۲
بلوک چهارم	۱۶×۱۶	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, N+nG\\3\times3, N+nG\end{bmatrix}\times4$
----	۱×۱	Global Average Pool, ۳d-FC, Softmax	
تعداد پارامتر		۲۵/۵ × ۱۰^۶	۱۰/۶ × ۱۰^۶

[۱] Concatenation Layer

[۲] Feature Maps

[۳] Fully Connected layer

تعداد فیلترهای هر لایه کانولوشن استفاده‌شده در شبکه از یک عمل جمع ساده پیروی می‌کند که به‌عنوان یک پارامتر آزاد قابل تغیر و بهینه شدن است. تعداد فیلترهای هر بلوک در معماری ResNet ثابت هستند ولی نسبت به بلوک قبل خود دو برابرند درحالی‌که در معماری پیشنهادی روند دیگری برای افزایش فیلترها نسبت به عمیق شدن شبکه در پیش گرفته‌شده است. در این شبکه تعداد فیلترهای هر لایه، حاصل جمع متغیر N با پارامتر نرخ رشد (G) است که در پایان هر سلول به آن اضافه می گردد. شکل (۱)، فلوچارت روش Faster R-CNN را نمایش می‌دهد که معماری کانولوشن ارائه‌شده در آن، وظیفه آموزش و استخراج ویژگی‌های سطح بالا را در یک فرآیند سلسله مراتبی، بر عهده دارد.



شکل ۱- فلوچارت روش Faster R-CNN به همراه معماری کانولوشن ارائه‌شده به عنوان استخراج‌گر ویژگی

۴- پیاده‌سازی و ارزیابی نتایج

۴-۱- آماده‌سازی داده‌ها

داده‌هایی که برای آموزش شبکه Faster R-CNN استفاده‌شده است، از فرودگاه‌های مختلف در سرتاسر دنیا اخذشده است. فرودگاه‌هایی مانند فرودگاه بین‌المللی امام خمینی (ره)، دبی، مهرآباد و فرودگاه بین‌المللی پکن ازجمله مهم‌ترین آن‌ها هستند. مجموعه داده جمع‌آوری‌شده شامل ۳۲۰ تصویر آموزشی و ۷۵ تصویر جهت ارزیابی معماری ارائه‌شده است که سه کلاس هواپیمای دو موتور، چهار موتور و هواپیما با موتور در عقب را شامل می‌شوند.

شبکه‌های مبتنی بر مرز جهت یادگیری عوارض موردنظر نیازمند مکان دقیق عارضه همراه باکلاس مرتبط با آن عارضه هستند. لازم است در داده‌های آموزشی، محل قرارگیری عوارض در تصویر با استفاده از یک چارچوب خاص مشخص شود تا شبکه عصبی کانولوشن به‌کاررفته در روش‌های شناسایی اهداف، بتواند عارضه هدف را از دیگر عوارض موجود در تصویر آزمایشی تشخیص دهد. به‌طور معمول مشخص کردن عوارض در تصاویر آموزشی با استفاده از ایجاد یک چهارضلعی دور آن انجام می‌شود. هر یک از رئوس بیانگر یک x و y به‌عنوان مختصات پیکسلی موقعیت عارضه هستند. به هر یک از آن‌ها یک برچسب که نمایانگر کلاس عارضه‌ی موردنظر است داده می‌شود تا در پروسه دسته‌بندی عوارض مورداستفاده قرار گیرد.

در پاره‌ای از تحقیقات [۳۶-۳۸] برای کاهش هزینه محاسباتی، تصویر مورد ارزیابی را به چندین بخش تقسیم کرده و پردازش‌های موردنظر را بر روی این قطعات انجام می‌دهند. همچنین این روش باعث کاهش خطا در شناسایی اهداف کوچک می‌شود [۳۶]. تعداد تقسیمات صورت گرفته در این مقاله با توجه به‌اندازه تصویر سنجش‌ازدوری، میزان همپوشانی[۱] و اندازه هر بخش محاسبه شده است.

۱ Overlap

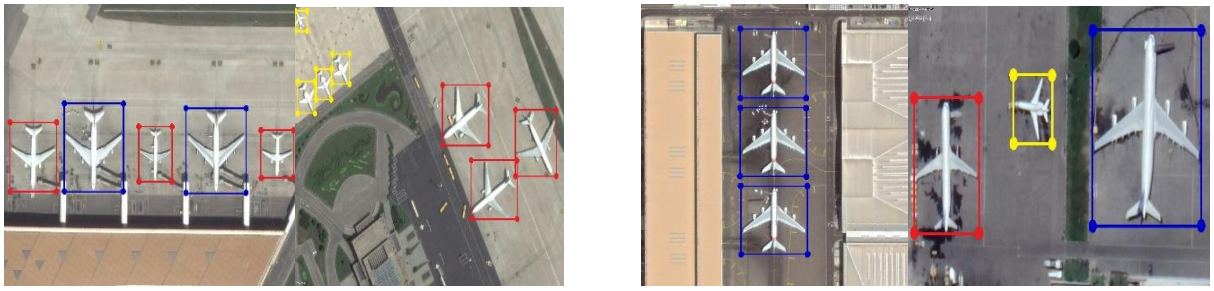
$$m = \left[\frac{\text{همپوشانی} - \text{طول تصویر منطقه}}{\text{همپوشانی} - \text{طول بخش}}\right] \quad (۴)$$

$$n = \left[\frac{\text{همپوشانی} - \text{عرض تصویر منطقه}}{\text{همپوشانی} - \text{عرض بخش}}\right] \quad (۵)$$

در رابطه (۴)، m تعداد بخش‌های موجود در طول و در رابطه (۵)، n تعداد بخش‌های موجود در عرض تصویر منطقه مطالعاتی می‌باشد. در روابط بالا، عبارت همپوشانی، فاصله پوششی بین بخش‌ها است. منطق و دلیل تعیین میزان همپوشانی، شناسایی هواپیماهایی است که بین مرز دو بلوک قرار دارند [۳۶]. در این مقاله برای منطقه مطالعاتی اول، تصویر ماهواره‌ای رزولوشن بالا با ابعاد ۴۸۰۰×۱۱۰۰۰ پیکسل مورد ارزیابی قرار گرفته است. برای هر بخش از این تصویر، ابعاد و همپوشانی به ترتیب برابر ۱۶۰۰×۱۱۰۰ و ۲۰ پیکسل به‌صورت دلخواه منظور شده است که با توجه به روابط (۴) و (۵)، مقادیر ۱۰ و ۳ به ترتیب برای m و n حاصل می‌گردد. شکل (۲) منطقه مطالعاتی اول را به همراه بخش‌های آن، نمایش می‌دهد.

شکل ۲- تصویر منطقه مورد مطالعه اول پس از عملیات بخش‌بندی

شکل (۳) نمونه‌ای از چند تصویر آموزشی و آزمایشی را به تصویر می‌کشد که هرکدام از هواپیماها در دسته‌های خود مشخص گردیده‌اند. قدرت تفکیک مکانی تصاویر ماهواره‌ای موجود در حدود ۱۵ سانتیمتر است. عوامل مؤثر بر کیفیت عکس مانند روشنایی، قدرت تفکیک و... به‌گونه‌ای انتخاب و کنترل شده‌اند که اکثر حالت‌های رؤیت شدن هواپیما را دربر می‌گیرد. منطقه مطالعاتی اول انتخاب‌شده، فرودگاه بین‌المللی پکن است که در این فرودگاه از سه نوع کلاس استفاده‌شده، وجود دارد. شکل (۴) تصویر ماهواره‌ای فرودگاه امام خمینی (ره) را نمایش می‌دهد که این تصویر نیز به‌عنوان تصویر مطالعاتی دوم انتخاب و پردازش شده است. ابعاد کلی این تصویر شامل ۳۰۰۰۰ × ۴۰۰۰ پیکسل می‌باشد. با توجه به روابط (۴) و (۵)، مقادیر ۲، ۶ و صفر به ترتیب برای m، n و پارامتر همپوشانی حاصل می‌گردد که درنتیجه هرکدام از قسمت-های ایجاد شده دارای ابعاد ۵۰۰۰ × ۲۰۰۰ هستند.

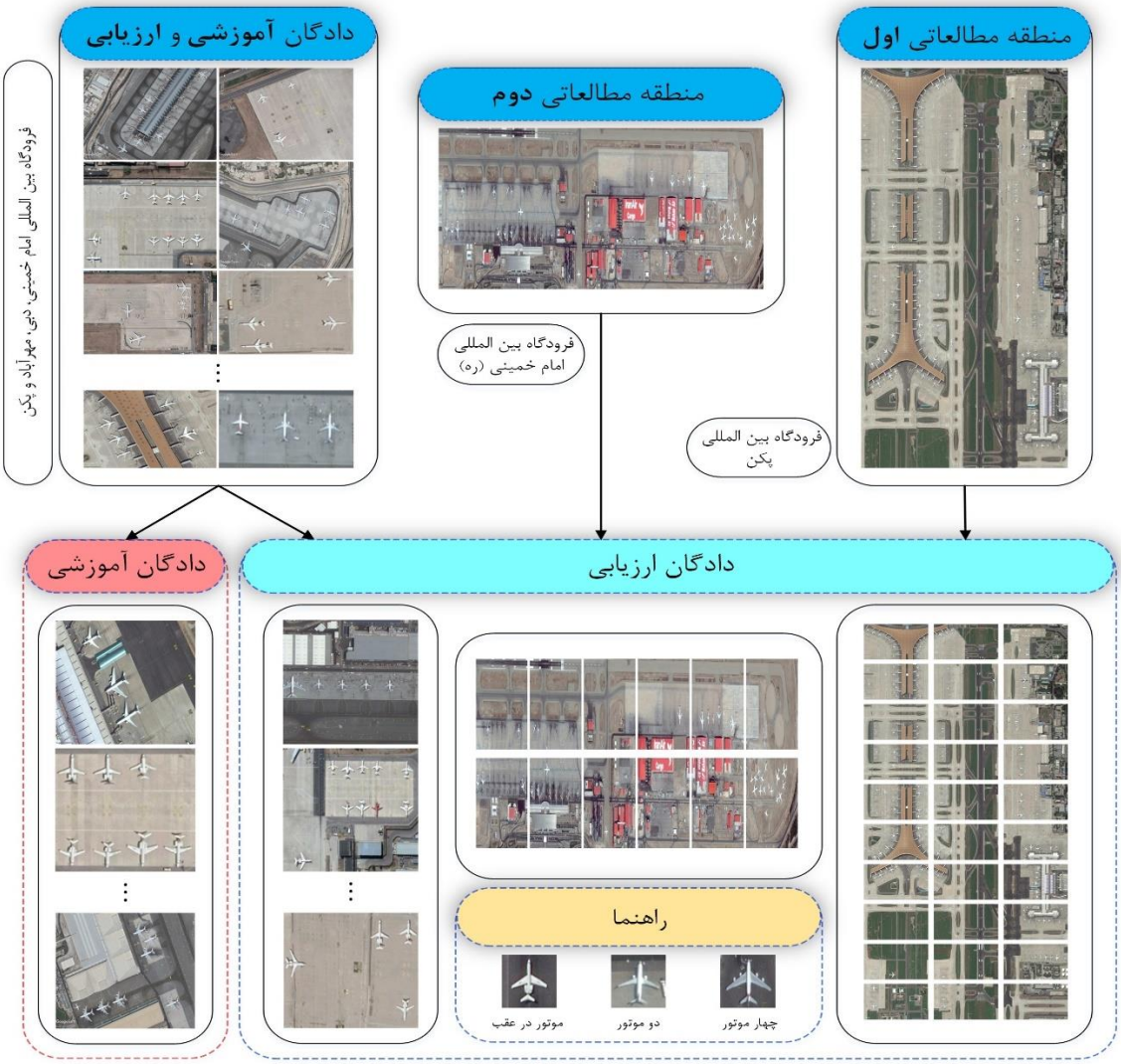


شکل ۳- نمونه ای از داده های آموزشی؛ تصاویر سمت راست، داده‌های آموزشی و تصاویر سمت چپ، داده‌های ارزیابی را نمایش می‌دهد. کادرهای آبی، قرمز و زرد به ترتیب معرف هواپیماهای چهار موتوره، دو موتوره و موتور در عقب هستند

شکل ۴- تصویر سنجش از دوری منطقه مطالعاتی دوم پس از انجام فرآیند بخش‌بندی

در این پژوهش، به‌منظور ارزیابی هرچه بهتر معماری پیشنهادی و معماری‌های دیگر، از سه دسته داده اعتبار-سنجی استفاده‌شده است. دسته اول مربوط به داده‌هایی از جنس داده‌های آموزشی است ولی با این تفاوت که در طی فرآیند آموزش شبکه، از آن‌ها استفاده‌نشده است. در این مجموعه تعداد ۷۵ نمونه، در دسترس است. دسته دوم و سوم جزء داده‌های با ابعاد بالا هستند که از فرودگاه‌های بین‌المللی پکن و امام خمینی (ره) اخذ گردیده و پس از اعمال بخش‌بندی جهت ارزیابی مورداستفاده قرار می-گیرند.

شکل (۵) شمای کلی از نحوه بودجه‌بندی تصاویر ماهواره‌ای باقدرت تفکیک بالا به مجموعه داده‌های آموزشی و ارزیابی است. شایان‌ذکر است تصحیحات اتمسفری بر روی تمامی داده‌های آموزشی، ارزیابی، تصاویر منطقه مطالعاتی اول و دوم اعمال‌شده است.



شکل ۵- دسته‌بندی تصاویر سنجش‌ازدوری به داده‌های آموزشی و ارزیابی

۴-۲- آموزش و پیاده‌سازی معماری ارائه‌شده

معماری ارائه‌شده در این تحقیق با استفاده از الگوریتم بهینه‌سازی SGD[۱] [۳۹] آموزش دیده است. نرخ یادگیری برای آن در ابتدای آموزش ۰/۰۰۵ قرار داده شد و در گام‌های ۴۰۰۰۰ و ۴۶۰۰۰، مقدار آن تقسیم‌بر ۱۰ شد. مقدار پارامترهای Weight Decay، Momentum و Batch Size الگوریتم بهینه‌ساز به ترتیب برابر ۰/۰۰۰۱، ۰/۹ و ۱۶ قرار داده‌شده‌اند. همچنین در معماری ارائه‌شده، از Dropout [۴۰] استفاده‌نشده است. پارامتر آزاد نرخ رشد برابر ۲ و پارامتر N برابر ۲۴ مقداردهی شده‌اند تا شبکه بهترین پوشش را روی‌داده‌ها داشته باشد. چارچوب مورداستفاده جهت آموزش نیز Keras با پس‌زمینه‌ی Tensorflow، CUDA نسخه ۹ و CuDNN نسخه ۷ می‌باشد و حافظه RAM و حافظه گرافیکی (VRAM) قابل‌دسترس به ترتیب برابر ۱۲ و ۱۶ گیگابایت هستند.

پس از مشخص کردن پارامترهای الگوریتم بهینه‌ساز، نوبت به انجام فرآیند آموزشی می‌رسد. در این فرآیند داده‌های آموزشی از لایه‌های کانولوشنی طراحی‌شده در شبکه عبور می‌کنند و در متغیرهای ترکیب‌کننده ذخیره می‌شوند. بعد از آن که فرایند پردازشی هر سلول به اتمام رسید، خروجی ترکیب‌کننده‌ها در درون سلول از یک لایه Average pooling عبور می‌کنند. داده‌های آموزشی، این فرآیند را دنبال می‌کنند تا به لایه چگال برسند. در این لایه به تعداد کلاس‌های موجود، نورون وجود دارد که وظیفه دسته‌بندی ویژگی‌های برداشت‌شده از دسته داده-

[۱] Stochastic Gradient Descent

های آموزشی را دارند. دسته‌بندی کلاس‌ها با استفاده از مقداردهی به هرکدام از نورون‌ها انجام می‌شود. نورونی که مقدار بیشتری داشته باشد، به‌عنوان کلاس پیش‌بینی‌شده برای تصویر انتخاب می‌شود. میزان دقت کلاس پیش‌بینی شده با تابع هزینه اندازه‌گیری می‌شود.

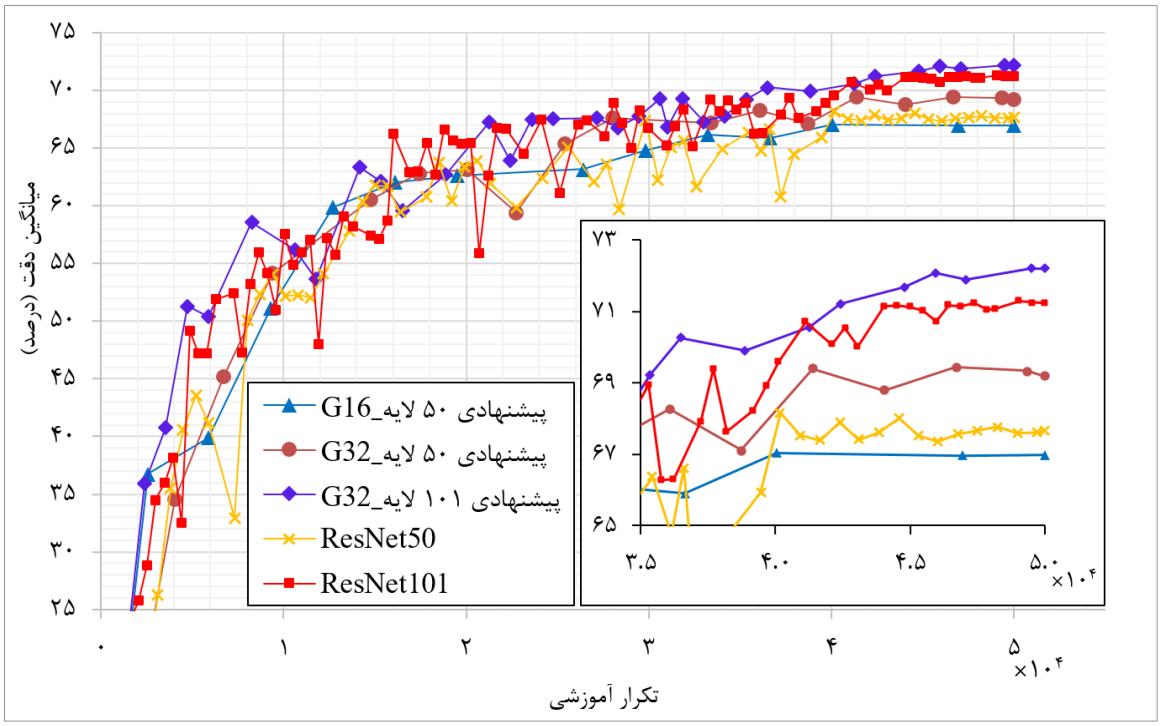
$$L = -\left[y \log(z) + (1-y)\log(1-z)\right] \quad (۶)$$

تابع هزینه استفاده‌شده در این تحقیق Cross-entropy است که رابطه (۶) بیانگر آن است. y و z به ترتیب مقدار صحیح برای کلاس داده و مقدار پیش‌بینی شده هستند.

۴-۳- ارزیابی و تحلیل نتایج

ارزیابی یک شبکه کانولوشن نیازمند بررسی دقت طبقه-بندی آن بر روی کلاس‌های موردنظر است. به‌منظور ارزیابی دقت روش ارائه‌شده، قسمتی از مجموعه داده به داده‌های ارزیابی اختصاص‌یافته است. این داده‌های پس از آموزش، مورد ارزیابی قرار می‌گیرد تا میزان اختلاف بین معماری‌های قبلی و معماری ارائه‌شده با استفاده از معیارهایی مشخص شود. رقابت Microsoft COCO [۴۱] ازجمله رقابت‌هایی است که به‌منظور ارزیابی دقت شبکه‌های کانولوشن برگزار می‌شوند. این رقابت شامل چندین معیار برای ارزیابی است که از میانگین دقت کلی (Average Precision) با IOU (Intersection Over Union) در بازه ۰/۵ تا ۰/۹۵ به‌عنوان معیار اصلی یاد می‌شود. طبق [۴۲]، مقادیر AP بر روی داده‌گان ارزیابی محاسبه می‌شود.

شکل (۶) میانگین دقت به‌دست‌آمده از اعمال معماری ارائه‌شده و معماری ResNet با تعداد لایه‌ها و نرخ رشد متفاوت در IOU بین ۰/۵ تا ۰/۹۵ را نمایش می‌دهد. با توجه به این‌که نرخ آموزشی در تکرار ۴۰۰۰۰ و ۴۶۰۰۰ به ده تقسیم می‌شود، الگوی رفتاری شبکه‌ها روان‌تر دنبال می‌شود. برای پایش وضعیت آموزشی، به‌طور منظم و هر ده دقیقه یک‌بار گزارشی از دقت متوسط هر یک از شبکه‌ها ذخیره می‌شود. به‌عبارتی‌دیگر نقاط موجود در شکل (۶) برای هر شبکه، گویای تعداد دفعات ارائه گزارش است. شبکه‌ای که در طول ۵۰۰۰۰ تکرار نقاط بیشتری دارد، حاکی از زمان آموزشی بالاتر آن برای تکمیل پروسه آموزش می‌باشد. در میان شبکه‌های مورد ارزیابی، شبکه پیشنهادی با ۵۰ لایه و نرخ رشد ۱۶، دارای کمترین و ResNet101 دارای بیشترین مدت‌زمان آموزشی می‌باشند. بر این اساس معماری ارائه‌شده در این مقاله توانایی دست-یابی به‌دقت شبکه ResNet را در عین کمتر بودن تعداد پارامتر و زمان لازم برای آموزش، دارا است.



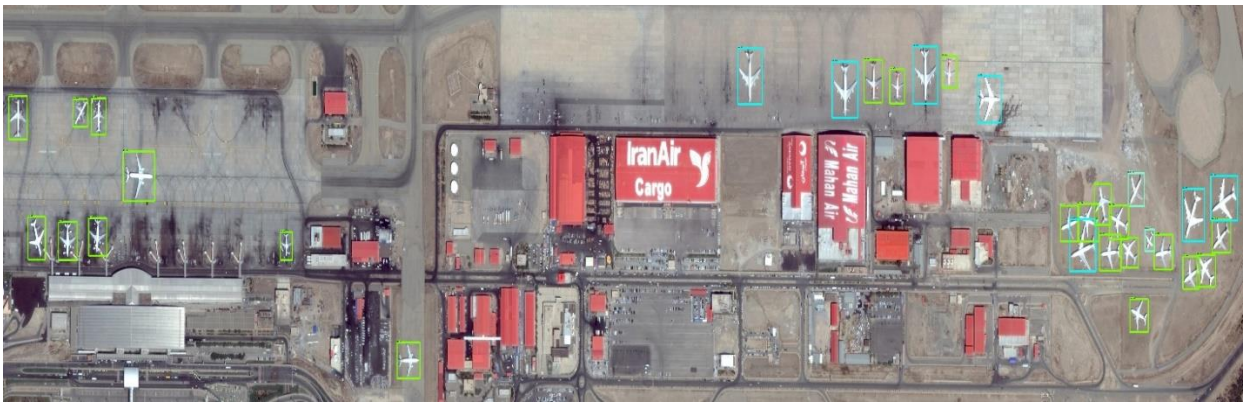
شکل ۶- مقدار میانگین دقت بدست آمده بر روی داده های ارزیابی

جدول ۲- میانگین دقت (AP) حاصله از داده‌های ارزیابی

معماری	تعداد لایه	نرخ رشد	تعداد پارامتر (میلیون)	مدت‌زمان آموزش (دقیقه)	میانگین دقت (درصد)
ResNet	۵۰	----	۲۵/۵	۶۱۶	۶۷/۶
	۱۰۱	----	۴۴/۶	۹۰۲	۷۱/۲
معماری پیشنهادی	۵۰	۱۶	۱۰/۶	۱۵۳	۶۶/۹
	۵۰	۳۲	۱۸/۲	۲۰۹	۶۹/۲
	۱۰۱	۳۲	۳۰/۲	۳۷۰	**۷۲/۱**

جدول (۲) نتایج به‌دست‌آمده از اعمال Faster RCNN همراه با معماری معرفی‌شده در این تحقیق و معماری ResNet را نمایش می‌دهد. بهترین دقت به‌دست‌آمده از داده‌های ارزیابی با قلم درشت مشخص‌شده است. روش ارائه‌شده علاوه بر این‌که دقت بالاتری در تشخیص اهداف دارد، سریع‌تر از سایرین داده‌های آموزشی را پوشش می دهد. حفظ ویژگی‌های حاصله در لایه‌های ابتدایی و استفاده از آن‌ها در لایه‌های بعدی موجب می‌شود میزان هدر رفت در استفاده از ویژگی‌ها کاهش یابد. همین مورد معماری پیشنهادی را از سایر معماری‌ها متمایز می‌کند.

شکل (۷) تصویر سنجش‌ازدوری از منطقه مطالعاتی دوم را نشان می‌دهد. تصویر موردنظر پس از اعمال استخراج‌گر معرفی‌شده در این تحقیق، به‌دست‌آمده است. شناسایی ۳۰ هواپیما از ۳۲ فروند هواپیما، حاکی از دقت مناسب شبکه در مواجه‌شدن با هر نوع تصویر رقومی برای شناسایی اهداف هدف دارد.



شکل ۷- نتیجه بدست آمده از اعمال شبکه پیشنهادی با ۱۰۱ لایه و G=۳۲ بر روی تصویر سنجش از دوری فرودگاه امام خمینی (ره)

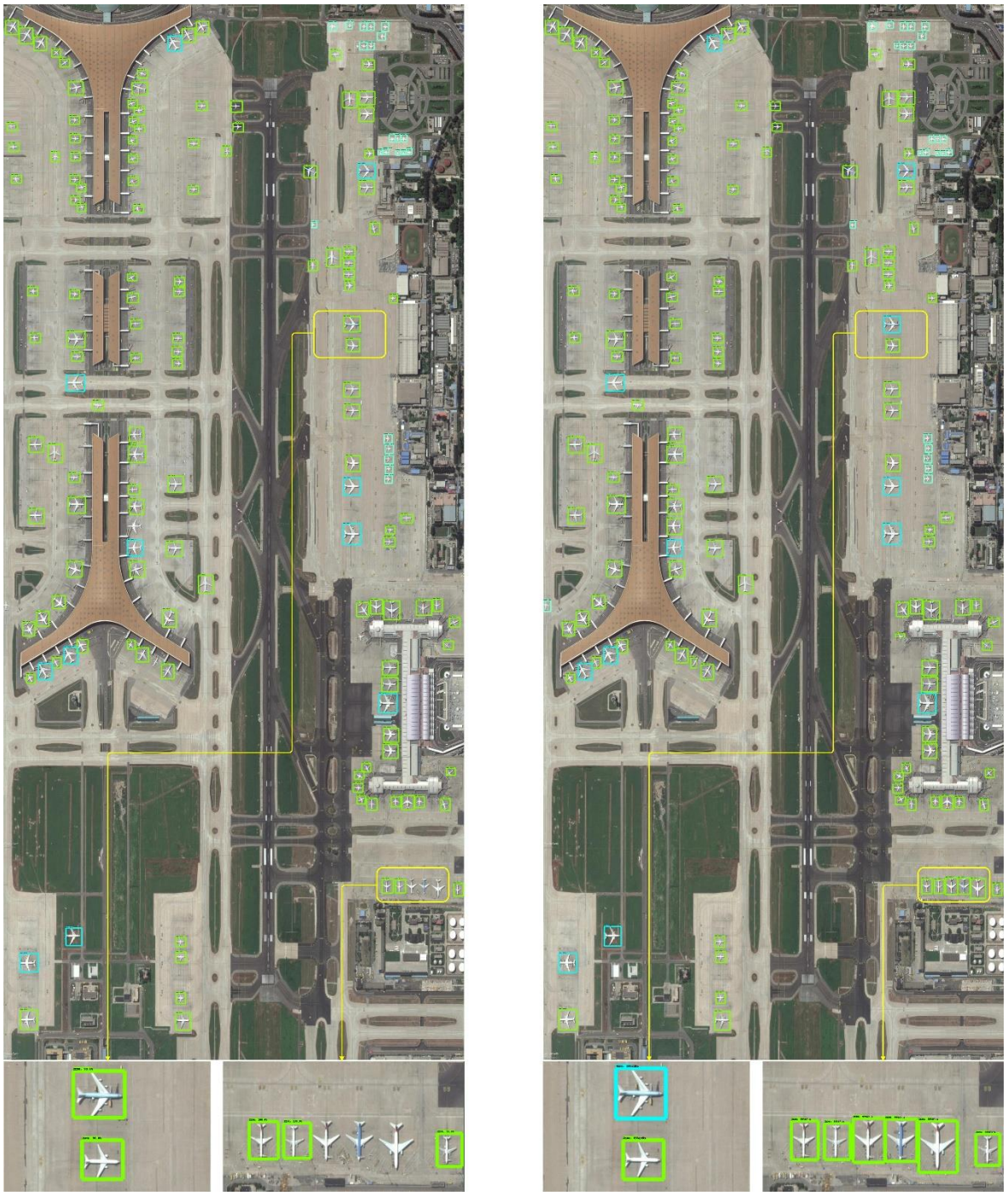
شکل (۸) شامل تصویر خروجی از اعمال معماری پیشنهادی و معماری ResNet با ۱۰۱ لایه بر روی تصویر منطقه مطالعاتی اول است. همان‌طور که مشاهده می‌شود، در سمت راست تصویر، معماری پیشنهادی با شناسایی صحیح ۱۶۵ فروند از ۱۶۹ فروند هواپیما با توجه به کلاس‌های تعیین‌شده، دقت قابل قبولی را در شناسایی عوارض هدف از خود به‌جای گذاشته است. البته شایان‌ذکر است که معماری ارائه‌شده در این مقاله نسبت به معماری ResNet با ۱۰۱ لایه، ۱۴/۴ میلیون پارامتر آموزشی کاهش داشته است. این موضوع، خود باعث کاهش توان پردازشی مورد نیاز هم در فرآیند آموزش و هم در فرآیند شناسایی اهداف می‌گردد.

همچنین برای مقایسه شبکه پیشنهادی و دیگر شبکه-ها، از معیار F1-Measure استفاده‌شده است. این معیار ترکیبی غیرخطی از دو معیار Recall و Precision می‌باشد [۴۳, ۴۴]. به‌منظور محاسبه معیار F1-Measure، مکان و کلاس تمام اهداف موردنظر در داده‌های ارزیابی مشخص می‌شود و شبکه برای هر هدف برچسبی با عنوان‌های TP[۱]، FP[۲]، NP[۳] اختصاص می‌دهد. زمانی به یک هدف تشخیص داده‌شده، برچسب TP داده می‌شود که محدوده‌ی شناسایی شده و منسوب به آن، حداقل ۵۰٪ همپوشانی باهدفی که به‌صورت دستی مشخص‌شده است، داشته [۲۸]. در غیر این صورت به هدف موردنظر برچسب FP داده خواهد شد. معیارهای Precision و Recall درصد آشکارسازی اهداف را نشان می‌دهند و به‌صورت زیر تعریف می‌شوند.

$$\text{Recall} = \frac{\text{TP}}{\text{NP}} \qquad (۷)$$

در رابطه (۷)، NP تعداد کل اهداف مکانی است که در داده‌های واقعی تولید شده‌اند. TP موقعیت‌های صحیح شناسایی شده است. این پارامتر همیشه عددی نامنفی است.

۱ True Positives sample
۲ False Positive samples
۳ Negative Positive samples



(الف): روش پیشنهادی با ۱۰۱ لایه و G=۳۲ (۳۰/۲ میلیون پارامتر)

(ب): معماری ResNet با ۱۰۱ لایه (۴۴/۶ میلیون پارامتر)

شکل ۸- نتیجه حاصل شده از اعمال معماری پیشنهادی و معماری ResNet بر روی تصویر منطقه مطالعاتی اول؛ رنگ‌های آبی فیروزه‌ای، زرد و سبز به ترتیب معرف دسته‌های هواپیماهای چهار موتوره، دو موتوره و موتور در عقب هستند.

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (۸)$$

در رابطه (۸)، FP تعداد موقعیت‌های غلط شناسایی شده در تمام تصاویر مورد آزمایش است. با استفاده از Precision و Recall، می‌توان معیار F1-Measure را محاسبه کرد.

$$F_\beta = (1 + \beta^2)\frac{\text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}} \qquad (۹)$$

رابطه‌ی (۹)، معیارهای Recall و Precision را با وزن β^2 ترکیب نموده و یک مقدار واحد و جامع حاصل می‌کند. مطابق با اکثر تحقیقاتی که درزمینه‌ی شناسایی

اهداف انجام‌شده است، مقدار β^2 برابر با ۱ در نظر گرفته‌شده است. در نظر گرفتن این مقدار برای پارامتر ذکرشده باعث می‌شود میزان تأثیرگذاری هر دو معیار یکسان باشد [۴۵].

جداول (۳) و (۴) نتایج خروجی از اعمال شبکه‌های موردبررسی بر روی منطقه مطالعاتی اول و دوم است. بهترین مقادیر برای معیار F1-Measure با قلم درشت و آبی مشخص‌شده است.

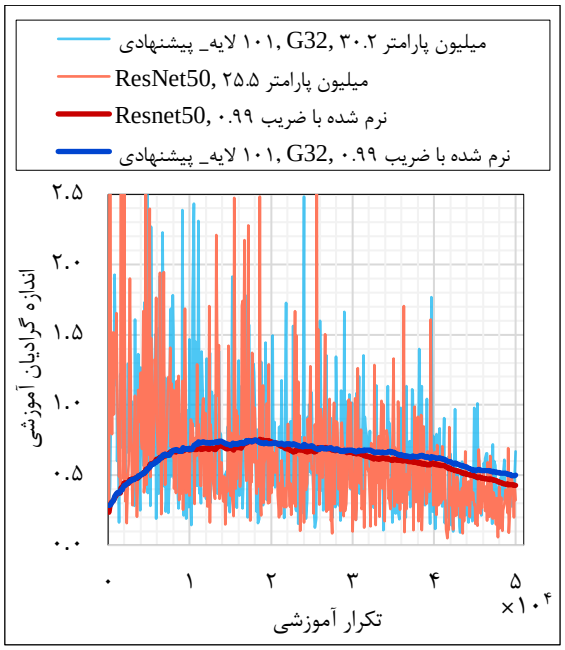
جدول ۳- مقادیر F1-Measure برای منطقه موردمطالعه اول

معماری	تعداد لایه	نرخ رشد	Precision	Recall	F1-Measure
ResNet	۵۰	----	۹۵/۶	۹۴/۸	۹۵/۲
	۱۰۱	----	۹۸/۳	۹۵/۲	۹۶/۷
معماری پیشنهادی	۵۰	۱۶	۹۲/۹	۹۳/۷	۹۳/۳
	۵۰	۳۲	۹۶/۷	۹۵/۹	۹۶/۳
	۱۰۱	۳۲	۹۸/۲	۹۷/۶	**۹۷/۹**

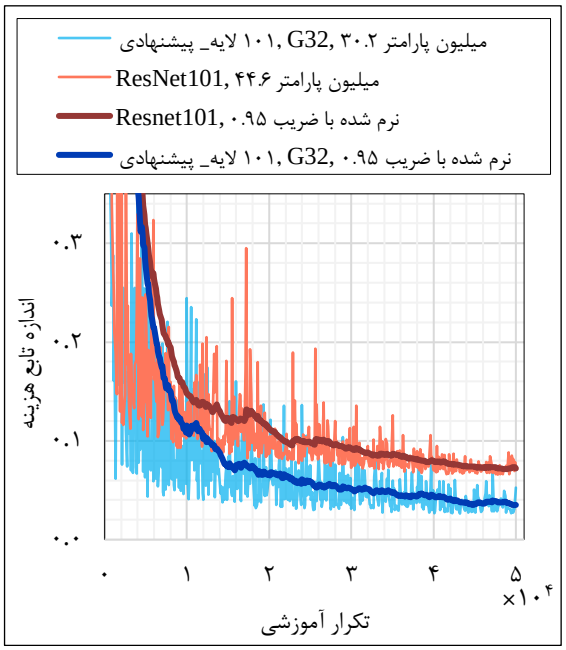
جدول ۴- مقادیر F1-Measure برای منطقه موردمطالعه دوم

معماری	تعداد لایه	نرخ رشد	Precision	Recall	F1-Measure
ResNet	۵۰	----	۸۹/۶	۹۰/۳	۸۹/۹
	۱۰۱	----	۹۲/۹	۹۳/۲	۹۳
معماری پیشنهادی	۵۰	۱۶	۸۸/۱	۸۸/۹	۸۸/۵
	۵۰	۳۲	۹۲/۵	۹۲/۹	۹۲/۷
	۱۰۱	۳۲	۹۳/۷	۹۳/۷	**۹۳/۷**

تعداد لایه بکار رفته در هر شبکه تأثیر بسزایی در میزان محو شدن گرادیان دارد. لایه‌های بیشتر باعث کم شدن اطلاعات منتقل‌شده به لایه‌های انتهایی می‌شود. شکل (۹) اندازه گرادیان را در ۵۰۰۰۰ تکرار آموزشی شبکه ResNet50 و معماری پیشنهادی را نمایش می‌دهد. معماری ارائه‌شده در این تحقیق باوجود تعداد لایه‌های دو برابر نسبت به شبکه مورد مقایسه، از شیب کاهشی کمتری برخوردار است و به‌طور مناسبی مشکل کاهش گرادیان را کم‌رنگ کرده است. این دلیلی بر مفید بودن استفاده چندین‌باره از ویژگی‌های هر لایه در طول فرایند آموزش شبکه است. با توجه به نوسان زیادی که مقدار گرادیان در هر تکرار از خود نشان می‌دهد، با استفاده از توابع نمایی همراه با ضریب ۰/۹۹ نمودار موردنظر نرم شده است تا دید بهتری از اختلاف موجود بین دو معماری به وجود آید. شکل (۱۰) آزمایشی است که برای مقایسه دقت طبقه‌بندی معماری پیشنهادی و ResNet با تعداد لایه‌های برابر انجام‌شده است. نمودارهای موجود در شکل (۱۰) نشان‌دهنده‌ی خطای کمتر معماری طراحی‌شده در این تحقیق، نسبت به معماری ResNet است. این مقایسه در شرایطی کاملاً برابر انجام‌شده است. در تمامی آزمایش‌ها، شبکه ارائه‌شده با تعداد پارامترهای کمتر و مدت‌زمان آموزش و ارزیابی پایین‌تر توانست نتایج بهتری نسبت به معماری ResNet از خود به‌جای بگذارد.



شکل ۹- مقدار نرم مربعی گرادیان حاصل از دو شبکه



شکل ۱۰- مقایسه تابع هزینه دو معماری در فرآیند آموزش

۵- نتیجه‌گیری

در این تحقیق با توجه به اهمیت شناسایی و طبقه-بندی اهداف در تصاویر سنجش‌ازدوری، فرآیندی مبتنی بر شبکه‌های کانولوشنی ارائه گردید که در عین کاهش تعداد پارامترهای آموزشی، با استخراج بهتر ویژگی‌های مناسب طبقه‌بندی، دقت روش‌های شناسایی اهداف را افزایش می‌دهد. افزایش تعداد لایه‌های کانولوشنی یکی از مؤلفه‌های اصلی در افزایش دقت طبقه‌بندی شبکه‌های کانولوشنی عمیق است؛ اما از طرفی افزایش بیش‌ازحد این لایه‌ها موجب محو شدن گرادیان و اطلاعات واردشده در لایه‌های انتهایی می‌گردد. این موضوع در بسیاری از تحقیقات صورت گرفته در زمینه‌ی شبکه‌های کانولوشن مطرح شده و با بررسی توپولوژی‌های مختلف و یا ارائه روش‌های آموزش نوین سعی در مقابله با چالش بوجود آمده کرده‌اند. پیرو دیگر محققان، در این مقاله، معماری کانولوشنی ارائه‌شده است که با حفظ ویژگی‌های تولیدشده در تمام طول شبکه، از محو شدن گرادیان و از بین رفتن ویژگی‌های مستخرج جلوگیری می‌کند. نتیجه‌ی این فرآیند، افزایش توانایی مدل‌های موجود در تفکیک اهداف چند کلاسه است.

فرآیند آموزش معماری کانولوشن پیشنهادی توسط ۳۲۰ قطعه تصویر انجام شد و پس از آن، این معماری بر روی دو منطقه مطالعاتی فرودگاه بین‌المللی پکن و فرودگاه بین‌المللی امام خمینی (ره) مورد ارزیابی قرار گرفت که به ترتیب مقادیر ۹۷/۹ و ۹۳/۷ درصد برای معیار F1-Measure به‌دست‌آمده است. این در حالی است که معماری ResNet با ۱۰۱ لایه کانولوشن و ۱۴/۴ میلیون پارامتر آموزشی بیشتر نسبت به معماری پیشنهادی، مقادیر ۹۶/۷ و ۹۳ درصد را برای معیار مذکور حاصل کرده است که این موضوع حاکی از برتری معماری پیشنهادی هم ازنقطه‌نظر دقت شناسایی و هم از نظر میزان پارامتر آموزشی و توان پردازشی موردنیاز دارد.

در راستای بهبود معماری ارائه‌شده در این مقاله می‌توان از اپراتورهای کانولوشن منبسط‌شده (Dilated Convolution) باهدف استخراج ویژگی‌های بارزتر استفاده کرد. از سوی دیگر باهدف توسعه و عمومی‌سازی هرچه بهتر روش ارائه‌شده، می‌توان آن را در دو مرحله بر روی تصاویر سنجش‌ازدوری با مقیاس کوچک در سطح کشوری و یا حتی بیشتر اعمال کرد؛ به‌طوری که در مرحله‌ی اول هدف شناسایی مکان فرودگاه باشد و در گام بعد هواپیماهای داخل هر فرودگاه توسط روش ارائه‌شده شناسایی شود.

مراجع

[1] Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," Neural computation, vol. 1, no. 4, pp. 541-551, 1989.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, 2015, pp. 91-99.

[3] W. Liu et al., "Ssd: Single shot multibox detector," in European conference on computer vision, 2016: Springer, pp. 21-37.

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

[5] X. Chen and A. Gupta, "An implementation of faster rcnn with study for region sampling," arXiv preprint arXiv:1702.02138, 2017.

[6] J. Lu, H. Sibai, and E. Fabry, "Adversarial examples that fool detectors," arXiv preprint arXiv:1712.02494, 2017.

[7] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, and J. Sun, "Light-head r-cnn: In defense of two-stage object detector," arXiv preprint arXiv:1711.07264, 2017.

[8] N. Farhadi, A. Kiani, and H. Ebadi, "Target detection from high-resolution remote sensing images using deep learning methods," Iranian Journal of Remote Sensing & GIS, vol. 11, no. 1, pp. 48-64, 2019.

[9] H. Wu, H. Zhang, J. Zhang, and F. Xu, "Fast aircraft detection in satellite images based on convolutional neural networks," in 2015 IEEE International Conference on Image Processing (ICIP), 2015: IEEE, pp. 4210-4214.

[10] F. Zhang, B. Du, L. Zhang, and M. Xu, "Weakly supervised learning based on coupled convolutional neural networks for aircraft detection," IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 9, pp. 5553-5563, 2016.

[11] F. Zeng et al., "A hierarchical airport detection method using spatial analysis and deep learning," Remote Sensing, vol. 11, no. 19, p. 2204, 2019.

[12] Y. Xu, M. Zhu, P. Xin, S. Li, M. Qi, and S. Ma, "Rapid airplane detection in remote sensing images based on multilayer feature fusion in fully convolutional neural networks," Sensors, vol. 18, no. 7, p. 2335, 2018.

[13] C. Cao et al., "Research on Airplane and Ship Detection of Aerial Remote Sensing Images Based on Convolutional Neural Network," Sensors, vol. 20, no. 17, p. 4696, 2020.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, p. 436, 2015.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[17] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE transactions on neural networks, vol. 5, no. 2, pp. 157-166, 1994.

[18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700-4708.

[19] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," in Advances in neural information processing systems, 2015, pp. 2377-2385.

[20] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, "Deep networks with stochastic depth," in European conference on computer vision, 2016: Springer, pp. 646-661.

[21] M. Chen, W. Wang, S. Dong, and X. Zhou, "Video Vehicle Detection and Recognition Based on MapReduce and Convolutional Neural Network," in International Conference on Sensing and Imaging, 2018: Springer, pp. 552-562.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 2009: Ieee, pp. 248-255.

[24] C. Szegedy et al., "D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions," IEEE Xplore, 2015.

[25] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818-2826.

[26] G. Larsson, M. Maire, and G. Shakhnarovich, "Fractalnet: Ultra-deep neural networks without residuals," arXiv preprint arXiv:1605.07648, 2016.

[27] Y. Bengio, "Learning deep architectures for AI," Foundations and trends® in Machine Learning, vol. 2, no. 1, pp. 1-127, 2009.

[28] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 9, pp. 1627-1645, 2009.

[29] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," CVPR (1), vol. 1, pp. 511-518, 2001.

[30] N. O. Attoh-Okine, "Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance," Advances in Engineering Software, vol. 30, no. 4, pp. 291-302, 1999.

[31] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580-587.

[32] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei, and H. Zou, "Multi-scale object detection in remote sensing imagery with convolutional neural networks," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 145, pp. 3-22, 2018.

[33] R. Girshick, "Fast R-CNN. In processing of IEEE International Conference on Computer Vision," Santiago, pp. 13-16, 2015.

[34] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167, 2015.

[35] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in Proceedings of the fourteenth international conference on artificial intelligence and statistics, 2011, pp. 315-323.

[36] Z. Chen, T. Zhang, and C. Ouyang, "End-to-end airplane detection using transfer learning in remote sensing images," Remote Sensing, vol. 10, no. 1, p. 139, 2018.

[37] Z. Han, H. Zhang, J. Zhang, and X. Hu, "Fast aircraft detection based on region locating network in large-scale remote sensing images," in 2017 IEEE International Conference on Image Processing (ICIP), 2017: IEEE, pp. 2294-2298.

[38] L. Sommer, N. Schmidt, A. Schumann, and J. Beyerer, "Search Area Reduction Fast-RCNN for Fast Vehicle Detection in Large Aerial Imagery," in 2018 25th IEEE International Conference on Image Processing (ICIP), 2018: IEEE, pp. 3054-3058.

[39] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in International conference on machine learning, 2013, pp. 1139-1147.

[40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," The Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929-1958, 2014.

[41] T.-Y. Lin et al., "Microsoft coco: Common objects in context," in European conference on computer vision, 2014: Springer, pp. 740-755.

[42] J. Hosang, R. Benenson, P. Dollár, and B. Schiele, "What makes for effective detection proposals?," IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 4, pp. 814-830, 2015.

[43] H. G. Akçay and S. Aksoy, "Automatic detection of geospatial objects using multiple hierarchical segmentations," IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 7, pp. 2097-2111, 2008.

[44] H. Sun, X. Sun, H. Wang, Y. Li, and X. Li, "Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model," IEEE Geoscience and Remote Sensing Letters, vol. 9, no. 1, pp. 109-113, 2011.

[45] A. O. Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 3, pp. 1701-1717, 2012.