

## شناسایی اهداف در تصاویر سنجش از دوری با قدرت تفکیک بالا با استفاده از روش‌های یادگیری عمیق

نیما فرهادی\*<sup>۱</sup>، عباس کیانی<sup>۲</sup>، حمید عبادی<sup>۳</sup>

۱. دانشجوی کارشناسی، گروه نقشه‌برداری، دانشگاه صنعتی نوشیروانی بابل
۲. دانشجوی دکتری، گروه فتوگرامتری و سنجش از دور، دانشگاه صنعتی خواجه نصیرالدین طوسی
۳. استاد گروه فتوگرامتری و سنجش از دور، دانشگاه صنعتی خواجه نصیرالدین طوسی

تاریخ پذیرش مقاله: ۱۳۹۷/۸/۳۰

تاریخ دریافت مقاله: ۱۳۹۷/۳/۱۱

### چکیده

شناسایی عوارض موجود در تصاویر، یکی از مسائل اساسی در حوضه تفسیر تصاویر به‌ویژه در تصاویر سنجش از دوری، به شمار می‌آید. یکی از روش‌های کارآمد و به‌روز در این زمینه، به‌کارگیری شیوه‌های یادگیری عمیق، جهت استخراج و تفسیر است. یک عارضه، مجموعه‌ای از الگوهای منحصر به فرد است که با عوارض مجاور خود متفاوت است، این تفاوت معمولاً در یک یا چند ویژگی به‌طور هم‌زمان اتفاق می‌افتد که می‌توان به تفاوت در شکل، رنگ و درجه خاکستری اشاره نمود. در این راستا، روش یادگیری عمیق با توانایی تحلیل مفاهیم انتزاعی سطح بالا، می‌تواند انتخاب مناسبی در این زمینه باشد. در روش پیشنهادی، ابتدا یک پایگاه داده مطابق با شرایط محیطی و جغرافیایی کشور از برخی از فرودگاه‌های ایران تشکیل شد. سپس با استفاده از شبکه‌های عصبی کانولوشنی به تولید مدل یادگیرنده بهینه اقدام شد. برای این کار، در قسمت پردازش داده‌های خام در کنار استفاده از روش انتقال آموزشی، بردارهایی جهت دسته‌بندی عوارض مورد نظر استخراج و به یک مدل ماشین‌بردار پشتیبان طبقه‌بندی‌کننده، تحویل داده می‌شوند. در ادامه، مقادیر خروجی با مقادیر به‌دست‌آمده از تصویر آزمایشی برای هر عارضه، مقایسه و در یک روند تکرار شونده تحلیل و جهت تطابق ساختاری بررسی می‌شوند. نتایج استخراج شده از اعمال مدل پیشنهادی بر روی چند سری داده آزمایشی، ارزیابی و با روش‌های مشابه مقایسه شد که در نهایت با مقادیر ۹۸/۲۱ درصد برای معیار Precision و ۹۹/۱ درصد برای معیار F1-Measure، قادر به شناسایی عوارض هدف است.

**کلیدواژه‌ها:** یادگیری عمیق، شبکه‌های کانولوشنی، تصاویر سنجش از دوری، یادگیری ماشین، انتقال آموزشی.

۱- مقدمه

امروزه با توجه به پیشرفت صنعت و فناوری، یکی از نیازهای اساسی و جدی در جامعه، بحث خودکار سازی و تسریع کارها است. خودکار سازی در تشخیص عوارض، به معنای انجام یک سری پردازش‌ها، برای شناسایی و انتخاب عوارض مورد نظر به صورت کاملاً خودکار است که دخالت انسان در آن به حداقل رسیده و از نتیجه آن هم می‌توان به تسریع انجام تحقیقات و کاهش هزینه‌های تمام شده اشاره نمود (ghorbani et al., 2017). در حال حاضر، سامانه‌های تشخیص اشیاء دقیقی مانند تشخیص صورت، سامانه‌های شمارش افراد و... به وجود آمده‌اند. این سامانه‌ها همه اشیاء موجود در تصویر را شناسایی و دسته‌بندی نمی‌کنند، بلکه موقعیت هر یک از اهداف مورد نظر را به وسیله ایجاد محدوده‌ای<sup>۱</sup> به دور آن‌ها، مشخص می‌کند (Girshick, 2015). این سامانه‌ها صرف نظر از دقتی که دارند، از زمان آموزش و شناسایی بالایی برخوردارند. در همین راستا، ظهور روش‌های یادگیری عمیق<sup>۲</sup>، در کنار دقت و سرعت پردازش بالا، نتایج قابل قبولی نیز از خود نشان می‌دهند.

یادگیری عمیق، جزء آن دسته از روش‌هایی است که با استفاده از مدل‌های مختلف، اقدام به مشخص کردن موقعیت عوارض موجود در تصاویر می‌کند. این روش، مفاهیم انتزاعی سطح بالا اشیاء را از تصاویر، دریافت و آن‌ها را به لایه‌هایی با مفاهیم سطح پایین‌تر تبدیل می‌کند (LeCun et al., 2015). از مزایای یادگیری عمیق، می‌توان به آموزش کاملاً خودکار ویژگی‌ها به صورت غیر نظارتی (He et al., 2014)، دقت بالا در نتایج خروجی و زمان آموزش پایین اشاره نمود (Diehl et al., 2015). به عنوان مثال، در بحث خودکار سازی، قابلیت استخراج خودکار ویژگی‌ها به یک موضوع بسیار مهم در این امر تبدیل شده است (Deng and Yu, 2014).

علم یادگیری عمیق، به چند زیرشاخه اصلی تبدیل می‌شود که شامل روش‌های مبتنی بر لایه‌های

کانولوشنی، روش‌های مبتنی بر RBM<sup>۳</sup>ها، روش‌های مبتنی بر Autoencoderها و روش‌های مبتنی بر Sparse Codingها است (He et Vincent et al., 2008; al., 2014). شبکه‌های عصبی کانولوشن، یکی از مهم‌ترین روش‌های یادگیری عمیق هستند که در آن‌ها چندین لایه با روشی قدرتمند آموزش می‌بینند (Ciregan et al., 2012). این روش، بسیار کارآمد بوده و یکی از رایج‌ترین روش‌ها در کاربردهای مختلف بینایی ماشین است. شبکه‌های عصبی کانولوشنی دارای دو روش مشتق شده R-CNN<sup>۴</sup> و FCN<sup>۵</sup> است. R-CNN در تحقیقات مربوط به شناسایی عوارض، کاربرد زیادی دارد و به معماری کلی برای الگوریتم‌های نوین جهت شناسایی اشیاء، تبدیل شده است. اما FCN بیشتر در تحقیقات قطعه‌بندی معنایی تصاویر، لبه یابی و غیره مورد استفاده قرار گرفته است (Xie and Tu, 2015; Yoo et al., 2015).

نخستین بار در سال ۲۰۱۴، گیرشیک و همکاران مدل R-CNN را معرفی کردند (Girshick et al., 2014). این مدل مرزهای عارضه مورد نظر را از تصاویر آموزشی برداشت کرده و با اعمال لایه‌های شبکه هوش مصنوعی بر روی مرزهای استخراج شده، موقعیت اشیاء موجود در تصویر را شناسایی می‌کند. اما مدل R-CNN فرایند ساده‌ای دارد و از طرفی، اعمال شبکه هوش مصنوعی بر روی هر مرز استخراج شده، زمان آموزش را بسیار بالا می‌برد. با این حال در تحقیقی دیگر، گیرشیک مدل Fast-R-CNN را ارائه کرد که از R-CNN الهام گرفته است، با این تفاوت که مشکل کندی فرایند را تا حدودی رفع کرده است (Girshick, 2015). این مدل به جای تولید انبوهی از شبکه‌های هوش مصنوعی و لایه‌های SVM<sup>۶</sup>، تنها از یک شبکه هوش مصنوعی و

1. Bounding Box
2. Deep Learning
3. Restricted Boltzmann Machines
4. Regions with CNN features
5. Fully convolutional network
6. Support Vector Machine

کانولوشنی و BING<sup>۳</sup> استفاده کردند (Cheng et al., 2014; Wu et al., 2015). آن‌ها استخراج مرزهای پیشنهادی را با استفاده از BING و دسته‌بندی مرزها را با استفاده از شبکه عصبی کانولوشنی انجام دادند. با توجه به تصاویر ماهواره‌ای، هواپیما جز آن دسته از عوارض است که از لحاظ چرخش دارای درجه آزادی بالایی است. در همین راستا، ژنگ و همکاران از روش EHO<sup>۴</sup> برای به دست آوردن ویژگی‌های ثابت چرخشی جدید استفاده کرد (Zhang et al., 2014). همچنین ونگ و همکاران برای به دست آوردن ویژگی‌های چرخشی، ماتریس ثابت چرخشی RIM<sup>۵</sup> را ارائه کردند (Wang et al., 2017). دستاوردهای اشاره شده در بالا فقط در مناطق خاصی تأثیرگذار هستند و معمولاً این روش‌ها از راه‌هایی برای شناسایی هواپیما استفاده می‌کنند که با محاسبات فراوانی همراه هستند. یکی از مزایای شبکه‌های عصبی، فرآیند شناسایی بر پایه ویژگی‌های غنی برداشت شده از داده‌های آموزشی است که حجم محاسبات را به شدت کاهش می‌دهد (Li et al., 2018). در همین راستا، دیائو و همکاران برای شناسایی هواپیما از یک شبکه باور عمیق<sup>۶</sup> که به وسیله ماشین بولتزن محدود آموزش دیده بود، استفاده کردند (Diao et al., 2016). اما این روش نیز تا حدودی برای شناسایی اشیاء، بهینه‌سازی نشده است (XU and HUO, 2018). برای جبران نقص‌های اشاره شده، ردمون و همکاران روش YOLO<sup>۷</sup> را ارائه کردند که در این روش با استفاده از یک شبکه عصبی عمیق، موقعیت‌یابی اشیاء را انجام دادند (Redmon et al., 2016). سرعت شناسایی و موقعیت‌یابی اشیاء، در این روش به‌طور چشم‌گیری

یک‌لایه SVM برای استخراج ویژگی و دسته‌بندی عوارض استفاده می‌کند، در نتیجه سرعتی به مراتب بالاتر از مدل R-CNN خواهد داشت. مدل Fast-R-CNN همانند مدل قبل خود از الگوریتم جست‌وجوی انتخابی برای انتخاب مرز عوارض استفاده می‌کند که این الگوریتم سرعت پایینی در انتخاب مرز از خود به جا می‌گذارد. برای حل این مشکل، رن و همکاران، مدل Faster R-CNN را ارائه داده‌اند (Ren et al., 2015). این مدل، همانند مدل Fast-R-CNN از یک شبکه هوش مصنوعی و Softmax layer برای استخراج ویژگی و دسته‌بندی عوارض استفاده می‌کند، اما در انتخاب مرز عوارض از شبکه عصبی سریع به جای الگوریتم جست‌وجوی انتخابی که سرعت پایینی دارد، استفاده می‌کند.

مدل Faster R-CNN تنها مدل تشخیص دقیق اشیاء نیست، اما در حال حاضر می‌توان آن را یکی از بهترین مدل‌های ارائه شده در نظر گرفت که قدرت پردازش بالایی دارد و از طرفی دیگر نسبت به دو مدل Fast R-CNN و R-CNN سرعت و دقت بیشتری در انتخاب عوارض از خود به جای گذاشته است (Ren et al., 2015). در سال‌های اخیر، تحقیقاتی در زمینه Faster R-CNN صورت گرفته است. به عنوان مثال، چن و گوپتا با در نظر گرفتن قانون NMS<sup>۱</sup> و کلاسه‌بندی، مرزهای مورد نظر را استخراج کردند (Chen and Gupta, 2017). باین حال، در این روش، امکان انتخاب عارضه اشتباه و یا عوارضی که در پس زمینه قرار دارند، وجود دارد (Lu et al., 2017). یوان و همکاران، برای بهبود دقت و سرعت شناسایی شبکه Faster R-CNN، از روش پالایش تکراری مرزهای پیشنهادی و روش LSTM<sup>۲</sup> استفاده کرده‌اند و اطلاعات بین تکرارهای مختلف را به دست آورده‌اند (Yuan et al., 2017).

دسته‌های دیگر از محققان، شناسایی اشیاء را در زمینه‌های تخصصی‌تر، مانند شناسایی هواپیما در تصاویر سنجش از دوری، انجام داده‌اند. به طور مثال، وو و همکاران برای شناسایی هواپیما از یک شبکه عصبی

1. Non-Maximal Suppression
2. Long Short Term Memory
3. Binarized Normed Gradients for objectness estimation
4. Extending Histogram Oriented Gradients
5. Rotation-Invariant Matrix
6. Deep Belief Network
7. You Only Look Once

استخراج شده از داده‌های خروجی را در لایه‌های سطح پایین، پشت سر هم مدل می‌کند. یادگیری عمیق از دسته‌بندی‌های مختلفی تشکیل شده است که پرکاربردترین آن روش‌های مبتنی بر شبکه عصبی کانولوشنی است. R-CNN و FCN، الگوهای مشتق شده از روش‌های مبتنی بر شبکه عصبی کانولوشنی است که در واقع همگی زیرشاخه یادگیری ماشین هستند. هدف اصلی R-CNN، شناسایی و انتخاب عوارض، با استفاده از ویژگی‌های به دست آمده از داده‌های آموزشی است. به طوری که، مدل‌های شبکه عصبی کانولوشنی را به عنوان یک استخراج گر ویژگی، مورد استفاده قرار داده و هیچ تغییری در شبکه اعمال نمی‌کند (Girshick, 2015).

**۲-۱- شبکه‌های عصبی کانولوشن مبتنی بر مرز**  
مدل R-CNN، با تلفیق مرزهای استخراج شده و لایه‌های شبکه هوش مصنوعی، موقعیت اشیاء موجود در تصویر را شناسایی می‌کند. این مدل در گام نخست، تصویر ورودی را برای یافتن اشیاء موجود با استفاده از الگوریتم جست و جوی انتخابی پردازش می‌کند و برای نمونه، ۲۰۰۰ مرز پیشنهادی استخراج می‌کند (Girshick, 2015). در گام بعد، لایه‌های شبکه هوش مصنوعی را بر روی مرزهای استخراج شده، اعمال می‌کند. در نهایت، با دسته‌بندی عوارض، پروسه تکمیل خواهد شد. مدل R-CNN، فرایند ساده‌ای دارد اما از طرفی، اعمال شبکه هوش مصنوعی بر روی هر مرز استخراج شده، زمان آموزش را بسیار بالا می‌برد. مدل Fast-R-CNN، مدلی است که از R-CNN الهام گرفته است، با این تفاوت که مشکل کندی فرایند را تا حدودی رفع کرده است. این مدل در گام اول، عملیات برداشت ویژگی را قبل از انتخاب مرز عوارض انجام می‌دهد، به این معنی که یک شبکه هوش مصنوعی را به جای ۲۰۰۰ شبکه هوش مصنوعی، بر روی تصویر ورودی اعمال می‌کند. گام دوم، جایگزین کردن یک لایه Softmax layer به جای مدل SVM است، به عبارتی با گسترش دادن شبکه هوش مصنوعی، اقدام به

بهبود یافت اما با این حال، دقت بالایی ندارد و تصاویر ورودی آن باید در اندازه‌های ثابت و یکسان مانند ۵۰۰×۵۰۰ پیکسل تعیین شده باشند که با توجه به ابعاد بالای تصاویر سنجش از دوری، برای شناسایی هواپیما مناسب نیستند (Xu et al., 2016).

بررسی و مطالعه تحقیق‌های صورت گرفته در گذشته، بیانگر ضرورت استفاده از مدل‌های یادگیری عمیق در تشخیص اشیاء است. این پژوهش‌ها، نشان‌دهنده این است که در انتخاب درست عوارض مانند هواپیما، عوامل متعددی همچون رنگ، شکل، اندازه و... موثر هستند. در مجموع، شناسایی اشیاء با استفاده از شبکه‌های مبتنی بر مرز، از زمان پیدایش تا به حال دستخوش تغییرات زیادی بوده و سمت و سوی این عمده تغییرات، بهبود در دقت شناسایی و کاهش زمان آموزش مدل است. در این مقاله، برای پوشش کاستی‌های اشاره شده به روشی برای بهبود Faster R-CNN و استخراج سریع و دقیق ویژگی‌ها از تصاویر با ابعاد بالا پرداخته شده است.

هدف از انجام این تحقیق، ارائه یک روش مناسب و کارآمد برای استخراج خودکار عوارض است که در همین راستا استفاده از الگوریتم یادگیری عمیق، جهت استخراج هواپیما از تصاویر با قدرت تفکیک مکانی بالا در کنار روش‌های انتقال آموزشی، ارائه شده است. مقاله حاضر، در پنج بخش کلی ارائه شده است که در بخش اول، در مقدمه، کلیاتی در مورد هدف از نوشتن مقاله و خلاصه‌ای از آن آمده است؛ سپس در بخش دوم، مبانی نظری تحقیق و در بخش سوم روش پیشنهادی بیان شده است. در بخش چهارم، نتایج و در نهایت در بخش پنجم نتیجه‌گیری و پیشنهادهای ارائه می‌شود.

## ۲- مبانی نظری تحقیق

یادگیری عمیق، زیرمجموعه‌ای از الگوریتم‌های یادگیری ماشین است که هدف آن، کشف چندین سطح از بازنمودهای توزیع شده از داده‌ورودی است. در واقع، یادگیری عمیق مفاهیم انتزاعی، سطح بالای

حل یک مسئله و اعمال آن بر مسئله دیگر، تمرکز دارد. انتقال آموزشی، یک فناوری بسیار مؤثر بوده و هدف آن منتقل کردن اطلاعات مفید از مسئله‌ای حل شده به یک مسئله جدید برای حل آن است (Chen et al., 2018). در انتقال آموزشی، ابتدا یک شبکه پایه به‌وسیله مجموعه داده‌های پایه آموزش می‌بیند و سپس ویژگی‌های آموزش داده شده به شبکه هدف انتقال داده می‌شوند که به این عمل، انطباق دقیق<sup>۳</sup> گفته می‌شود (Yosinski et al., 2014). انطباق دقیق یکی از مهم‌ترین ابزارهای انتقال آموزشی است. تعداد انبوهی از مسائل مربوط به دسته‌بندی و شناسایی، از انطباق دقیق استفاده می‌کنند و دلیل آن نیز کمبود داده آموزشی است. در اکثر موارد هم گزارش شده است که نتایج به‌دست‌آمده از انطباق دقیق شبکه‌های هوش مصنوعی، به‌مراتب دقیق‌تر از روش‌های سنتی آموزش آن‌ها است. به‌عنوان مثال، در چند سال اخیر، کارهای زیادی در زمینه شناسایی اشیاء که مبتنی بر شبکه‌های هوش مصنوعی هستند، صورت گرفته است، درحالی‌که شبکه به‌کاررفته در آن‌ها معمولاً از شبکه‌هایی مانند ResNet، VGG، CIFAR10Net، GoogLeNet، ZF هستند (Chen et al., 2018). این مدل‌های از پیش آموزش داده شده، در کار شناسایی اشیاء استفاده می‌شود. ImageNet یک پایگاه داده بصری انبوه است که تقریباً بیشتر مدل‌های از پیش آموزش داده شده، بر روی این پایگاه داده، آموزش دیده‌اند (Krizhevsky et al., 2012). مدل استفاده شده در این مقاله، CIFAR10Net است که بر روی ۵۰۰۰۰ داده آموزشی در ۱۰ کلاس آموزش دیده است (Krizhevsky et al., 2014). سپس این مدل، به‌وسیله داده‌های آموزشی به‌دست‌آمده هواپیما، یکنواخت سازی می‌شود تا توانایی شناسایی هواپیما را داشته باشد.

دسته‌بندی عوارض می‌کند. با این کار، از تولید یک مدل جدید SVM نیز جلوگیری می‌شود. این مدل به‌جای تولید انبوهی از شبکه‌های هوش مصنوعی و لایه‌های SVM، تنها از یک شبکه هوش مصنوعی و یک لایه SVM برای استخراج ویژگی و دسته‌بندی عوارض استفاده می‌کند، در نتیجه سرعتی به‌مراتب بالاتر خواهد داشت.

مدل Fast-R-CNN نیز از الگوریتم جست‌وجوی انتخابی، برای انتخاب مرز عوارض استفاده می‌کند که این الگوریتم، سرعت پایینی در انتخاب مرزها از خود به‌جا می‌گذارد (Ren et al., 2015). برای حل این مشکل نیز مدل Faster R-CNN ارائه شده است. مدل Faster R-CNN همانند مدل Fast-R-CNN از یک شبکه هوش مصنوعی و Softmax layer برای استخراج ویژگی و دسته‌بندی آن استفاده می‌کند اما در انتخاب مرز عوارض، از شبکه عصبی سریع به‌جای الگوریتم کند جست‌وجوی انتخابی استفاده می‌کند که نام این شبکه عصبی، RPN<sup>۱</sup> است. نحوه عملکرد یک RPN به‌گونه‌ای که پس از اعمال شبکه هوش مصنوعی به تصویر ورودی، در لایه آخر همان شبکه یک ماسک با ابعاد  $n$  در  $n$  (معمولاً  $n$  برابر ۳) به نقشه عوارض تولید شده، توسط شبکه هوش مصنوعی، اعمال می‌شود و ابعاد آن‌ها را کاهش می‌دهد. سپس ماسک اعمال شده، چندین مرز امکان‌پذیر را بر اساس محدوده‌های پیش‌فرض تولید می‌کند و در نهایت نیز مرزهایی که شامل عارضه موردنظر هستند، انتخاب و مختصاتی که توسط محدوده‌های اطراف عارضه مشخص شده است، ارائه می‌شود. مدل Faster R-CNN، مقادیر به‌دست‌آمده از RPN را جایگزین مقادیر جست‌وجوی انتخابی کرده و در گام‌های بعد، همانند مدل Fast R-CNN، لایه‌های مرسوم را اضافه می‌کند (Chen and Gupta, 2017).

## ۲-۲- انتقال آموزشی<sup>۲</sup>

انتقال آموزشی یک موضوع، مطالعاتی در زمینه یادگیری ماشین است که بر روی ذخیره دانش به‌دست‌آمده در

1. Region Proposal Network
2. Transfer Learning
3. Fine Tuning

### ۳- روش پیشنهادی

#### ۳-۱- آماده‌سازی داده‌ها

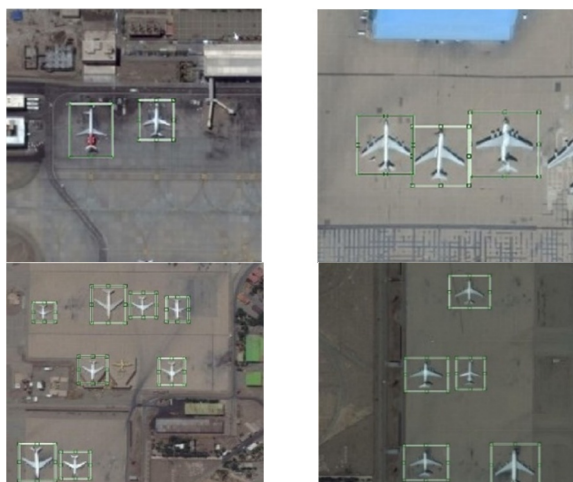
داده‌های آموزشی که برای آموزش شبکه عصبی مبتنی بر مرزهای پیشنهادشده استفاده می‌شود، از طریق تصاویر گوگل ارث، تأمین شده است. ابعاد و شکل داده‌ها در ارتباط مستقیم با نتایج حاصل از الگوریتم هستند، به‌گونه‌ای که هر چه قدر کیفیت یا ابعاد داده‌های آموزشی بیشتر شود، کیفیت تصویر ارزیابی نیز باید افزایش یابد. به‌همین دلیل، داده‌هایی با ابعاد پایین جهت آموزش شبکه استفاده شده است تا هم در مقدار زمان آموزش و حداقل سخت‌افزار موردنیاز صرفه‌جویی شود.

داده‌های آموزشی به‌صورت بخش‌های تصویری با ابعاد  $100 \times 100$  تا  $500 \times 500$  انتخاب و سپس در پایگاه داده ذخیره شده است. در این مقاله تمامی تصاویر آموزشی از فرودگاه‌های داخل کشور انتخاب شده‌اند، اما برای داده‌های ارزیابی علاوه بر فرودگاه‌های داخلی، از فرودگاه‌هایی مانند فرودگاه بین‌المللی پکن و فرودگاه بین‌المللی استانبول ترکیه به دلیل بالابودن تعداد هواپیماها و مناسب برای ارزیابی، استفاده شده است. در حقیقت، یک پایگاه داده که تطابق بالایی با ویژگی‌های محیطی داخل کشور دارد، تهیه شده است. با توجه به ساختار و روش پیشنهادشده جهت شناسایی

اهداف، این تحقیق به‌خوبی قابلیت تعمیم به مناطق دیگر را داراست. در این تحقیق، تعداد ۵۷۰ بخش تصویری شامل ۱۲۹۰ عارضه موردنظر به‌عنوان داده‌های آموزشی و آزمایشی استفاده شد و شبکه نیز از این عوارض معرفی‌شده برای برداشت ساختار و در ادامه برای شناسایی خودکار عوارض موردنظر استفاده می‌کند. شکل ۱، نمونه‌ای از عوارض معرفی‌شده به شبکه را نمایش می‌دهد.

#### ۳-۲- طراحی شبکه

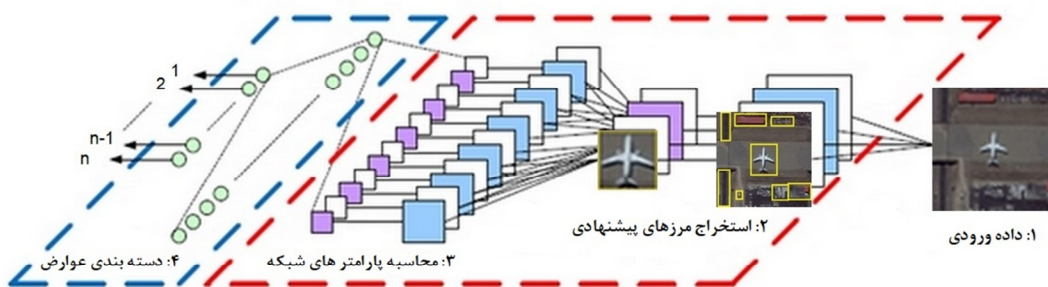
شبکه‌ای که در این مقاله پیشنهادشده است بر اساس شبکه‌های درختی یا سلسله‌مراتبی ۱ است و به این معناست که در چنین معماری‌هایی هر لایه، ورودی خود را از خروجی لایه قبلی دریافت می‌کند که باعث استحکام محتوایی و جلوگیری از کاهش اطلاعات مفید می‌شود (Ciregan et al., 2012). البته عمق شبکه، یکی از عوامل مؤثر از بین رفتن اطلاعات مفید در ورودی لایه‌ها است که باید نسبت مناسب بین داده‌های آموزشی و عمق شبکه حفظ شود. در این مقاله، از چهار لایه Faster R-CNN به‌صورت سلسله‌وار استفاده شده است که هرکدام لایه‌های ورودی خود را از Faster R-CNN قبلی دریافت می‌کند و این لایه‌ها در هر بار گذر از



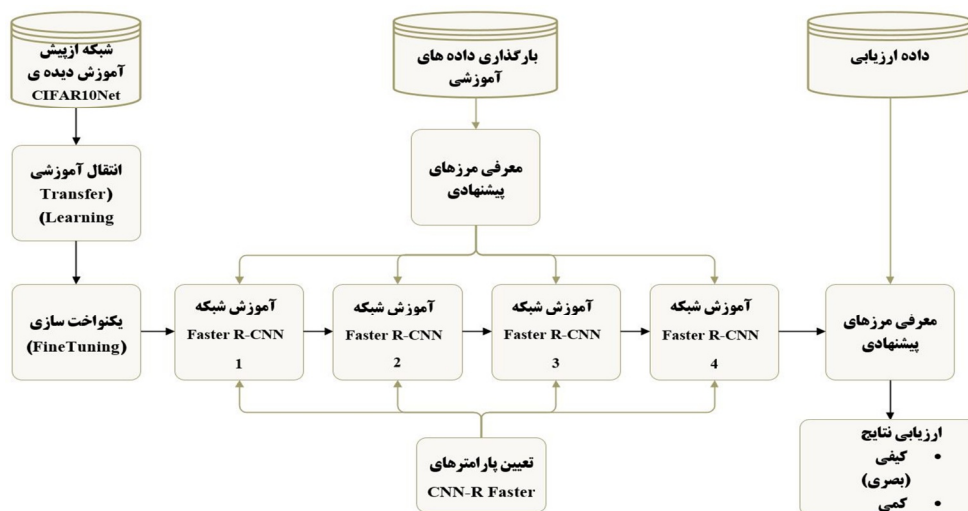
شکل ۱. نمونه‌ای از مرزهای پیشنهادی

شبکه شناسایی ارائه شده متشکل از چهار شبکه FASTER R-CNN است که هر کدام از آن‌ها دارای ۱۵ لایه کانولوشنی که شامل ۲ لایه Fully-connected هستند. این معماری به جهت پایین بودن زمان شناسایی انتخاب شده است. شکل (۲)، ۱۵ لایه کانولوشنی اشاره شده را به اختصار نمایش می‌دهد. روند اجرایی و بخش‌های مختلف روش پیشنهادی در شکل ۳ ارائه شده است. این روند نما، نمایی کلی از قسمت‌های مختلف شبکه است که گام اول استخراج لایه‌های آموزش دیده شبکه از پیش آموزش دهنده CIFAR10Net، گام دوم تعیین پارامترهای Faster R-CNN، گام سوم آموزش شبکه و گام آخر نیز شناسایی عوارض موضعی و ارزیابی نتایج است.

Faster R-CNN به روزرسانی می‌شوند تا در مرحله آخر برای شناسایی، دقت بالایی داشته باشند. لایه‌های ابتدایی که وارد اولین شبکه می‌شوند دارای معماری مناسبی برای دسته‌بندی اشیاء در تصاویر سنجش از دوری هستند و به وفور در کارهای تشخیص اشیاء، مورد استفاده قرار می‌گیرند که به آن‌ها شبکه پایه گفته می‌شود. در این مقاله، از CIFAR10Net به عنوان شبکه پایه بهره گرفته شده است. CIFAR10Net برای راحت آموزش دیدن، طراحی شده و مناسب برای انطباق دقیق است. لایه‌های آموزش دیده CIFAR10Net، به عنوان لایه‌های اولین Faster R-CNN معرفی می‌شود. این لایه‌ها با استفاده از ۵۰۴ تصویر هواپیما، انطباق دقیق و به روزرسانی می‌شوند و در گام‌های بعدی خطاهای تشخیص کاهش می‌یابد.



شکل ۲. نمایش گرافیکی از معماری شبکه عصبی چندلایه مبتنی بر مرز Faster R-CNN



شکل ۳. روند نما کلی روش پیشنهادی

نیز از معیار F-Measure که ترکیبی از دو معیار قبلی است بهره گرفته شده است (Akçay and Aksoy, 2008; Sun et al., 2012). برای انجام این کار، به صورت دستی، تمام هدف‌های موردنظر در داده‌های ارزیابی مشخص شده است و شبکه برای هر هدف، مقادیری با عنوان‌های TP<sup>1</sup>، FP<sup>1</sup>، NP<sup>2</sup> اختصاص می‌دهد. برچسب TP، زمانی به یک هدف تشخیص داده شده، به وسیله شبکه داده می‌شود که محدوده منسوب به آن، حداقل ۵۰٪ همپوشانی با هدفی که به صورت دستی مشخص شده است، داشته باشد (Felzenszwalb et al., 2010). در غیر این صورت، به هدف موردنظر برچسب FP داده خواهد شد. معیارهای Precision و Recall درصد آشکارسازی اهداف موردبررسی را نشان داده و به صورت زیر تعریف می‌شوند.

$$\text{Recall} = \frac{TP}{NP} \quad \text{رابطه (۱)}$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{رابطه (۲)}$$

در رابطه (۱)، NP تعداد کل اهداف مکانی است که در داده‌های واقعی تولید شده‌اند. TP موقعیت‌های صحیح شناسایی شده است و در رابطه (۲)، FP تعداد موقعیت‌های غلط شناسایی شده در تمام تصاویر مورد آزمایش است. دو معیار بالا، پس از محاسبه در نمودار رسم می‌شوند. معیارهای Precision و Recall به تنهایی نمی‌توانند به طور جامع، کیفیت شناسایی را بازگو نمایند، به همین دلیل، از معیاری با عنوان F1-Measure استفاده شده است (Han et al., 2014). مطابق رابطه زیر، معیارهای Recall و Precision را با میانگین‌گیری با وزن  $\beta^2$  ترکیب نموده و یک مقدار واحد حاصل می‌شود.

$$F_{\beta} = (1 + \beta^2) \frac{\text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}} \quad \text{رابطه (۳)}$$

در رابطه (۳)، مقدار  $\beta^2 < 1$  میزان اهمیت معیار

1. True Positives sample
2. False Positive samples
3. Negative Positive samples

### ۳-۲-۱- آموزش شبکه عصبی مبتنی بر مرزهای پیشنهادی

شبکه CIFAR10Net، یک شبکه از پیش آموزش دیده را برای ما فراهم می‌کند، اما باین حال برای شناسایی هواپیما مناسب نیست و نیاز به آموزش بیشتر دارد و پارامترهای صحیح باید تخمین زده شوند. همچنین لازم به ذکر است که از پایگاه داده‌های معروف مانند PASCAL VOC 2007، 2012 برای آموزش شبکه استفاده نشده است و دلیل آن نیز اختلاف بین تصاویر مربوط به هواپیما در این پایگاه داده‌ها و تصاویر سنجش از دوری همین عارضه، است (Radovic et al., 2017).

### ۴- پیاده‌سازی و ارزیابی نتایج

ارزیابی یک شبکه، نیازمند ارزیابی دقت طبقه‌بندی آن بر روی کلاس مورد مطالعه است. تصاویر کلاس موجود در این مقاله، از برنامه گوگل ارث تهیه شده‌اند و در آن تنوع نوع هواپیماها، تنوع اندازه از هواپیما، با قدرت تفکیک پایین تا قدرت تفکیک بالا در نظر گرفته شده است. سیستم مورد استفاده برای آموزش و ارزیابی شبکه دارای کارت گرافیکی NVIDIA GDDR5 GeForce 960 M با حافظه ۴ گیگابایتی و پردازنده INTEL CORE I7 6700HQ است. شکل ۴، لایه‌های شبکه عصبی بکار رفته را به تصویر نشان می‌دهد. شبکه عصبی موردنظر به سه لایه اصلی تقسیم می‌شود و هر لایه نیز شامل زیر لایه‌هایی است که برای ورود اطلاعات، پردازش و خروج آن مورد استفاده قرار می‌گیرند.

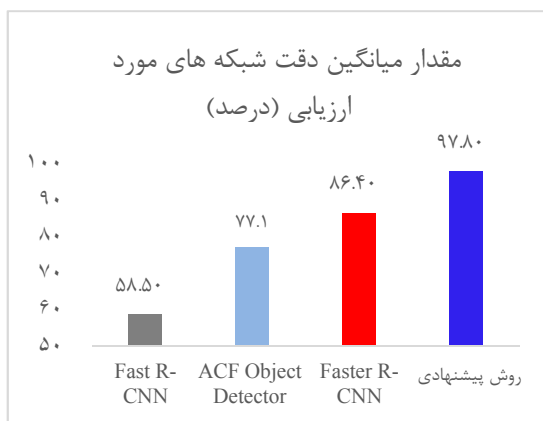
به منظور ارزیابی جامع تر روش پیشنهادی، علاوه بر داده‌های منطقه مورد مطالعه، دو سری داده متفاوت دیگر نیز موردبررسی قرار گرفته‌اند. داده دوم از تصاویر با قدرت تفکیک مکانی مشابه داده‌های آموزشی و از یک منطقه دیگر با ساختار متفاوت انتخاب شده‌اند. در ادامه، داده‌های دیگری که از لحاظ قدرت تفکیک بالا با دو داده نخست متفاوت است، جهت قیاس استفاده شدند.

همچنین برای نمایش میزان دقت الگوریتم پیشنهادی، از دو معیار استاندارد Recall و Precision و



مساحت زیر منحنی‌های این نمودار نشان‌دهنده m.AP یا میانگین دقت چند دسته است که در این مقاله، از یک دسته استفاده شده است. مقدار m.AP برای روش پیشنهادی برابر ۹۷/۸٪، برای روش Faster R-CNN اولیه برابر ۸۶/۴٪، برای روش ACF Object Detector برابر ۷۷/۱٪ و برای روش Fast R-CNN برابر ۵۸/۵٪ است. انتخاب m.AP روش پیشنهادی با مقدار ۹۷/۸٪ نشان‌دهنده، بالاتر بودن دقت و شناسایی بهتر هواپیما نسبت به سایر روش‌های موجود در این زمینه است. جهت قیاس درون‌سامانه‌ای نیز شبکه پیشنهادی با استفاده از انتقال آموزشی و بدون استفاده از انتقال آموزشی، مورد ارزیابی قرار داده شد و شکل ۶، نمودار آن را نمایش می‌دهد.

مطابق شکل ۶، روش ارائه شده بدون استفاده از انتقال آموزشی دارای m.AP کمتری از روش پیشنهادی با استفاده از انتقال آموزشی است. مطابق با نتایج حاصل شده، روش ارائه شده در این تحقیق با استفاده از شبکه‌های چندلایه و انتقال آموزشی به صورت هم‌زمان می‌تواند عملکرد مناسبی در شناسایی اهداف و بهبود دقت ارائه دهد.

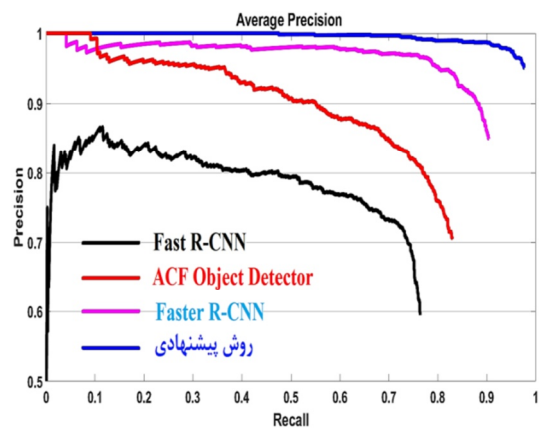


شکل ۷. نتایج شبکه‌های مورد نظر بر روی داده‌های ارزیابی

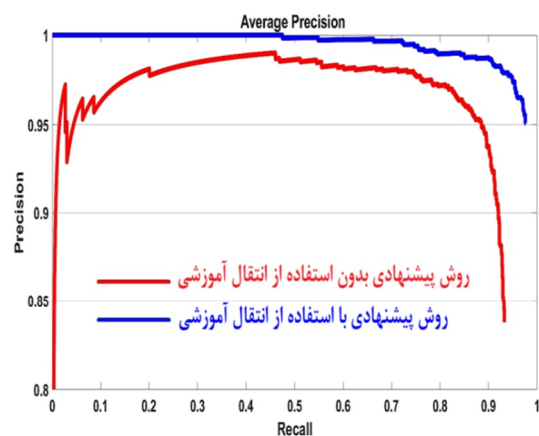
1 Mean Average Precision

Precision را افزایش خواهد داد و بالعکس. مطابق با اکثر تحقیقاتی که در زمینه شناسایی اهداف، انجام شده است مقدار  $\beta^2$  را برابر با ۱ در نظر گرفته‌اند که میزان تأثیرگذاری هر دو معیار با یکدیگر برابر شود (Ok et al., 2013).

جهت قیاس نسبی از روش‌های (Girshick, 2015) ACF Object (Dollár et al., 2014), Fast R-CNN) Detector و Faster R-CNN (Ren et al., 2015) به منظور ارزیابی روش پیشنهادی استفاده گردید. روش‌ها با داده‌های آموزشی یکسانی آموزش دیده‌اند و همچنین از داده‌های مشابه برای ارزیابی آن‌ها استفاده شده است. شکل ۵ نمودار Precision/recall را نمایش می‌دهد.



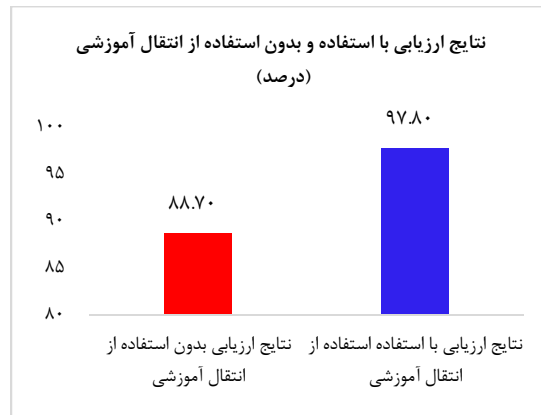
شکل ۵. نتایج شبکه‌های مورد نظر بر روی داده‌های ارزیابی



شکل ۶. نتایج ارزیابی با استفاده و بدون استفاده از انتقال آموزشی

شکل ۷، نمایش نمودار میله‌ای نتایج به دست آمده از مقایسه شبکه‌های مورد نظر بر روی داده‌های ارزیابی است. همچنین شکل ۸، نیز نمودار میله‌ای نتایج حاصله از ارزیابی شبکه با استفاده و بدون استفاده از انتقال آموزش را به تصویر می‌کشد. شایان ذکر است برای درک و نمایش بهتر، مقیاس نمودارها متناسب با مقدار نمایش داده شده تنظیم شده است.

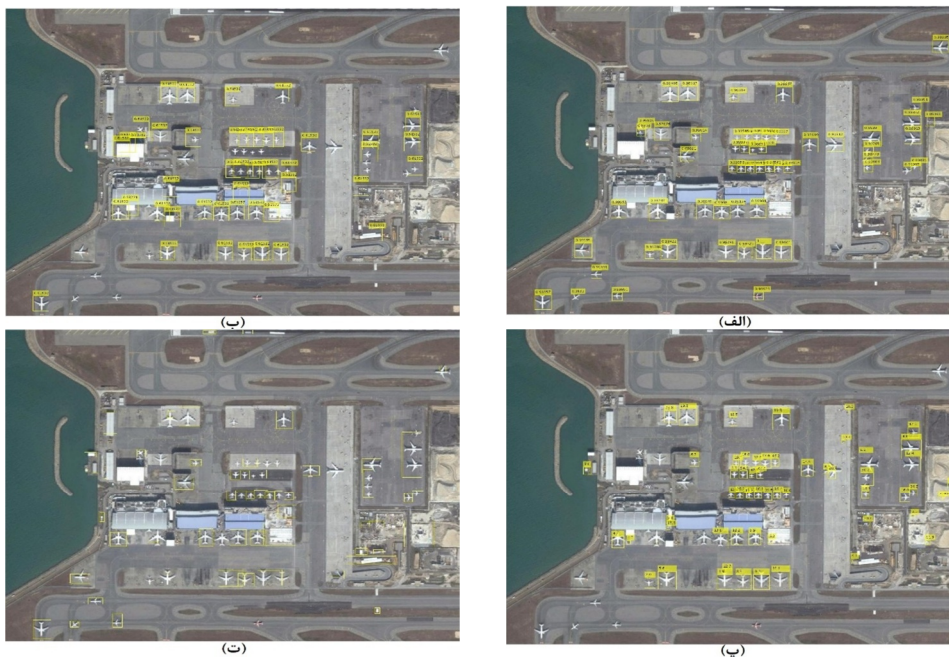
همچنین جدول ۱، نتایج کمی به دست آمده از روش پیشنهادی و سایر روش‌ها را ارائه نموده است. همان‌طور که از اعداد به دست آمده مشخص است، روش پیشنهادی، نتایجی دقیق‌تر را بر روی منطقه مورد مطالعه ارائه می‌دهد.



شکل ۸. نتایج ارزیابی با استفاده و بدون استفاده از انتقال آموزشی

جدول ۱. نتایج ارزیابی روش پیشنهادی با دیگر روش‌های مرتبط با شناسایی عوارض بر روی منطقه مورد مطالعه

معیار ارزیابی	شکل ۹ (الف)	شکل ۹ (ب)	شکل ۹ (پ)	شکل ۹ (ت)
TP	۵۵	۳۵	۳۸	۳۱
FP	۱	۱۱	۱۹	۲۱
NP	۵۵	۵۵	۵۵	۵۵
Recall	۱۰۰	۶۳/۶۴	۶۹	۵۶/۳۶
Precision	۹۸/۲۱	۷۶/۱	۶۶/۱	۵۹/۶۲
F1-measure	۹۹/۱	۶۹/۳۱	۶۷/۸	۵۷/۹۴



شکل ۹. شناسایی هواپیما در تصویر منطقه مورد مطالعه، (الف) نتایج حاصل از روش پیشنهادی؛ (ب) نتایج حاصل از روش

Faster R-CNN ابتدایی؛ (پ) نتایج حاصل از روش ACF Detector؛ (ت) نتایج حاصل از روش Fast R-CNN

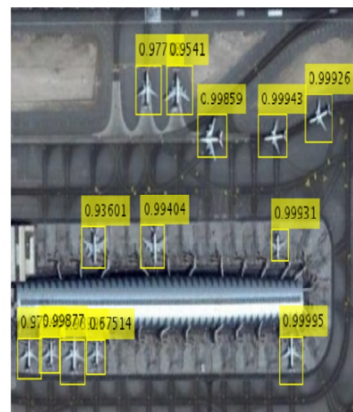
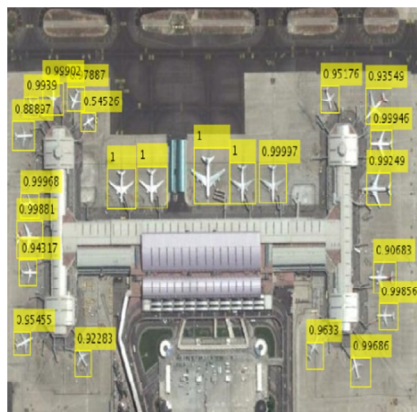
شبهه الگوی آموزشی باشد. در نتیجه با توجه به ابعاد داده‌های آموزشی که در دسترس هستند، انتظار می‌رود از انتخاب داده‌های ارزیابی با قدرت تفکیک بالا، نتایج قابل قبولی حاصل نشود. یکی دیگر از مزایای روش پیشنهادی قابلیت تشخیص عوارض موجود در تصاویر با قدرت تفکیک‌های مختلف است. حال اگر داده‌های ارزیابی ما دارای ابعاد کوچکی باشند، با استفاده از انتقال آموزشی می‌توان از تصاویر با رزولوشن بالا نیز نتیجه‌ی مناسبی به دست آورد.

شکل ۱۱، نتایج حاصل‌شده از اعمال شبکه طراحی‌شده بر روی تصاویر ارزیابی با قدرت تفکیک مکانی بسیار بالا را نمایش می‌دهد. نتایج به‌دست آمده نشان‌دهنده آن است که شبکه، دقت بالایی در شناسایی اهداف بسیار کوچک را دارد.

شکل ۹، نتایج بررسی شبکه پیشنهادی و روش‌های مقایسه‌ای را بر روی منطقه مورد مطالعه نمایش می‌دهد. روش پیشنهادی با توانایی شناسایی همه اهداف (۵۵ هواپیما از ۵۵ هواپیما) موجود در منطقه، روش کارآمدتری نسبت به سایر روش‌ها است.

شکل ۱۰، نتایج به‌دست‌آمده از اعمال روش پیشنهادی بر روی تعدادی داده ارزیابی دیگر را نمایش می‌دهد. این داده‌ها به‌صورت کاملاً تصادفی از فرودگاه‌های داخلی و خارجی به‌دست‌آمده‌اند.

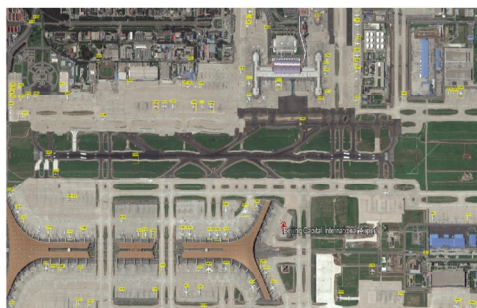
شبکه‌های Faster R-CNN، توانایی تشخیص عوارضی را دارند که الگوی آن عارضه مطابق با الگوی به‌دست‌آمده از داده‌های آموزشی باشد و این بدان معناست که اگر شبکه به‌وسیله الگوی یک عارضه با تعداد پیکسل پایین آموزش داده‌شده باشد، فقط توانایی شناسایی عوارضی را دارد که الگوی آن عارضه تقریباً



شکل ۱۰: نمونه‌ای از نتایج به‌دست‌آمده از روش پیشنهادی با استفاده از انتقال آموزشی



(ب)



(الف)



(ت)

شکل ۱۱: اعمال روش پیشنهادی با انتقال آموزشی بر روی تصاویر ارزیابی با قدرت تفکیک بسیار بالا، (الف) فرودگاه بین‌المللی پکن چین، (ب) فرودگاه بین‌المللی مهرآباد، (ت) فرودگاه بین‌المللی استانبول ترکیه

جدول ۲. نتایج ارزیابی روش پیشنهادی

	شکل ۱۱ ارزیابی پارامترها	شکل ۱۱ (ب)	شکل ۱۱ (الف)
Precision	۹۳/۵۲	۱۰۰	۹۱/۶۱
Recall	۹۵/۲۸	۹۷/۱۴	۹۶/۷۷
F1-measure	۹۴/۳۹	۹۸/۵۵	۹۴/۱۲

داده شد. قدرت تفکیک مکانی داده‌های اخذ شده علاوه بر تأثیر مستقیم در قدرت پردازشی مورد نیاز، بر دقت شناسایی نیز تأثیر خواهد گذاشت. انتخاب داده‌های مناسب برای هر عمل تشخیص اشیاء، یک فرایند دقیق است که نیاز به آزمایش‌ها و سعی و خطای فراوانی دارد. در ادامه، می‌توان با به‌کارگیری روش‌های ارائه شده در این مقاله مانند یادگیری عمیق، شبکه‌های کانولوشنی، انتقال آموزشی و... شناسایی هر نوع عارضه‌ای را انجام داد که در مقاله حاضر هواپیما به‌عنوان عارضه هدف انتخاب شده است.

در نهایت، پیشرفت‌های به‌دست آمده در این مقاله را می‌توان در دو بخش دنبال کرد. اولین بخش این است

همان‌طور که قبلاً گفته شد در اغلب کاربردها با توجه به محدودیت‌های سخت‌افزاری که با آن مواجه هستیم، قابلیت آموزش شبکه با تصاویر حجیم و قدرت تفکیک بالا وجود ندارد. در نتیجه اکثر روش‌هایی که اقدام به تشخیص عوارض می‌کنند، توانایی پردازش تصاویر سنجش از دوری با قدرت تفکیک بالا را نداشته و تقریباً هیچ خروجی از آن‌ها به دست نمی‌آید و فقط قادر به پردازش تصاویری هستند که هم‌اندازه داده‌های آموزشی باشند. این کاستی نه فقط در شبکه‌های هوش مصنوعی مبتنی بر مرز بلکه در بقیه روش‌ها و الگوریتم‌های تشخیص عوارض نیز دیده می‌شود. روش پیشنهادی در این مقاله با استفاده از روش انتقال آموزشی در کنار شبکه اصلی قادر خواهد بود این نقص را به حداقل برساند. جدول ۲، نتایج به‌دست آمده از اعمال روش پیشنهادی بر روی داده‌های با قدرت تفکیک مکانی و ابعاد تصویری بالا را نمایش می‌دهد.

در این مقاله، در مورد عوامل مؤثر بر نحوه انتخاب داده‌های آموزشی و داده‌های ارزیابی به‌طور مفصل شرح

## ۶- منابع

- Akçay, H.G. and Aksoy, S., 2008, **Automatic detection of geospatial objects using multiple hierarchical segmentations**, IEEE Transactions on Geoscience and Remote Sensing 46, 2097-2111.
- Chen, X. and Gupta, A., 2017, **An Implementation of Faster RCNN with Study for Region Sampling**, arXiv preprint arXiv:1702.02138.
- Chen, Z., Zhang, T. and Ouyang, C., 2018, **End-to-End Airplane Detection Using Transfer Learning in Remote Sensing Images**, Remote Sensing 10, 139.
- Cheng, M.-M., Zhang, Z., Lin, W.-Y. and Torr, P., 2014, **BING: Binarized normed gradients for objectness estimation at 300fps**, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3286-3293.
- Ciregan, D., Meier, U. and Schmidhuber, J., 2012, **Multi-column deep neural networks for image classification**, Computer vision and pattern recognition (CVPR), 2012 IEEE conference on. IEEE, pp. 3642-3649.
- Deng, L. and Yu, D., 2014, **Deep learning: methods and applications**, Foundations and Trends® in Signal Processing 7, 197-387.
- Diao, W., Sun, X., Zheng, X., Dou, F., Wang, H. and Fu, K., 2016, **Efficient saliency-based object detection in remote sensing images using deep belief networks**, IEEE Geoscience and Remote Sensing Letters 13, 137-141.
- Diehl, P.U., Neil, D., Binas, J., Cook, M., Liu, S.C. and Pfeiffer, M., 2015, **Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In Neural Networks (IJCNN)**, 2015 International Joint Conference on IEEE, pp. 1-8.
- Dollár, P., Appel, R., Belongie, S. and Perona, P., 2014, **Fast feature pyramids for object detection**, IEEE Transactions on Pattern Analysis and Machine Intelligence 36, 1532-1545.

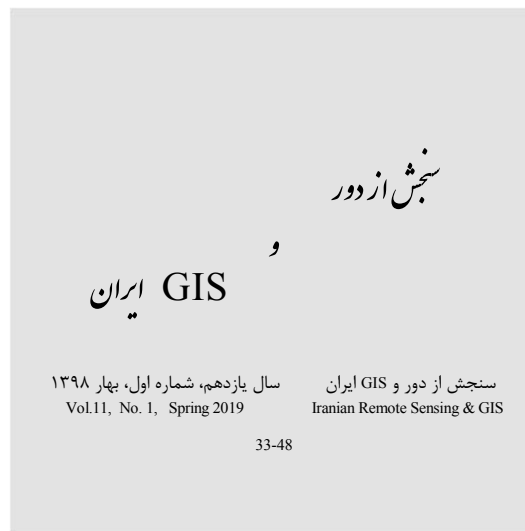
که اگر مسائلی مانند شناسایی چند عارضه به‌طور هم‌زمان مطرح باشد، می‌توان نتایج خوبی را با پیشرفت‌های حاصله از این مقاله، کسب کرد، فقط کافی است پایگاه داده را مطابق عوارض هدف ایجاد کرد. در بخش دوم، می‌توان از شبکه‌های از پیش آموزش‌دیده متفاوتی مانند GoogLeNet, ZF, CIFAR10Net, VGG و ResNet برای انتقال آموزشی بهره برد. شبکه مورد استفاده در مقاله حاضر، CIFAR10Net است.

## ۵- نتیجه‌گیری

در این تحقیق، روشی جهت شناسایی و دسته‌بندی عوارض موجود به‌منظور تفسیر معنایی تصاویر باتا‌کید بر عارضه هواپیما در تصاویر سنجش‌ازدوری ارائه می‌شود. این تحقیق، از سه مرحله آموزش، پردازش و شناسایی تشکیل شده‌است که جهت انجام آموزش، مجموعه داده تصویری مناسبی از فرودگاه‌های داخل کشور با استفاده از تصاویر گوگل ارث تهیه شده‌است. روش پیشنهادی از نقطه‌نظر جامعیت، در مواجهه با تصاویر با قدرت تفکیک‌های مختلف و تصاویر با ابعاد بزرگ از کارایی مناسبی برخوردار است. از ویژگی‌های بارز این الگوریتم می‌توان به سرعت بالای آموزش، دقت زیاد آن در تصمیم‌گیری و پایداری آن در برابر انواع تغییرات روشنایی و درجه خاکستری در تصاویر اشاره نمود. در این تحقیق، برای ارزیابی کارایی روش پیشنهادی و روش‌های شناسایی اشیاء، علاوه بر داده ارزیابی منطقه مورد مطالعه، از داده‌های تصادفی برداشت‌شده از فرودگاه‌های دیگر نیز استفاده شد. از جمله ویژگی‌های این تحقیق می‌توان به استفاده هم‌زمان از شبکه‌های عصبی، مبتنی بر مرز و انتقال آموزشی در کنار استفاده از چندین شبکه تکراری پشت سرهم اشاره کرد. نتایج حاصل شده از روش پیشنهادی، برای تصاویر ارزیابی منطقه مورد مطالعه با مقدار  $98/21$  برای معیار Precision و  $99/1$  برای معیار F1-measure نتایج قابل قبولی را از خود بجا گذاشته است.

- Felzenszwalb, P.F., Girshick, R.B., McAllester, D. and Ramanan, D., 2010, **Object detection with discriminatively trained part-based models**, IEEE transactions on pattern analysis and machine intelligence 32, 1627-1645.
- ghorbani, F., Ebadi, H. and Sedaghat, A., 2017, **Classification of the high-resolution remote sensing images for land use and land cover application using SIFT algorithm**, The 2nd National Conference on Geospatial Information Technology (NCGIT). K.N.Toosi University of Technology, Tehran.
- Girshick, R., 2015, **Fast R-CNN**, *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE Computer Society, pp. 1440-1448.
- Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2014, **Rich feature hierarchies for accurate object detection and semantic segmentation**, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580-587.
- Han, J., Zhou, P., Zhang, D., Cheng, G., Guo, L., Liu, Z., Bu, S. and Wu, J., 2014, **Efficient, simultaneous detection of multi-class geospatial targets based on visual saliency modeling and discriminative learning of sparse coding**, *ISPRS Journal of Photogrammetry and Remote Sensing* 89, 37-48.
- He, Y., Kavukcuoglu, K., Wang, Y., Szlam, A. and Qi, Y., 2014, **Unsupervised feature learning by deep sparse coding**, In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pp. 902-910. Society for Industrial and Applied Mathematics.
- Krizhevsky, A., Nair, V. and Hinton, G., 2014, **The CIFAR-10 dataset**.
- Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012, **Imagenet classification with deep convolutional neural networks**, *Advances in neural information processing systems*, pp. 1097-1105.
- LeCun, Y., Bengio, Y. and Hinton, G., 2015, **Deep learning**. *nature* 521, 436-444.
- Li, Y., Fu, K., Sun, H., Sun, X., 2018, **An Aircraft Detection Framework Based on Reinforcement Learning and Convolutional Neural Networks in Remote Sensing Images**, *Remote Sensing* 10, 243.
- Lu, J., Sibai, H. and Fabry, E., 2017, **Adversarial Examples that Fool Detectors**, arXiv preprint arXiv:1712.02494.
- Ok, A.O., Senaras, C. and Yuksel, B., 2013, **Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery**, *IEEE Transactions on Geoscience and Remote Sensing* 51, 1701-1717.
- Radovic, M., Adarkwa, O., Wang, Q., 2017, **Object Recognition in Aerial Images Using Convolutional Neural Networks**, *Journal of Imaging* 3, 21.
- Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2016, **You only look once: Unified, real-time object detection**, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788.
- Ren, S., He, K., Girshick, R. and Sun, J., 2015, **Faster r-cnn: Towards real-time object detection with region proposal networks**, *Advances in neural information processing systems*, pp. 91-99.
- Sun, H., Sun, X., Wang, H., Li, Y. and Li, X., 2012, **Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model**, *IEEE Geoscience and Remote Sensing Letters* 9, 109-113.
- Vincent, P., Larochelle, H., Bengio, Y. and Manzagol, P.A., 2008, **Extracting and composing robust features with denoising autoencoders**, In *Proceedings of the 25th international conference on Machine learning*, pp. 1096-1103
- Wang, G., Wang, X., Fan, B., Pan, C., 2017, **Feature extraction by rotation-invariant matrix representation for object detection in aerial image**, *IEEE Geoscience and Remote Sensing Letters* 14, 851-855.

- Wu, H., Zhang, H., Zhang, J., Xu, F., 2015, **Fast aircraft detection in satellite images based on convolutional neural networks**, Image Processing (ICIP), 2015 IEEE International Conference on. IEEE, pp. 4210-4214.
- Xie, S., Tu, Z., 2015, **Holistically-nested edge detection**, Proceedings of the IEEE international conference on computer vision, pp. 1395-1403.
- XU, N. and HUO, C., 2018, **Learning Deep Relationship for Object Detection**, IEICE TRANSACTIONS on Information and Systems 101, 273-276.
- Xu, T.-B., Cheng, G.-L., Yang, J. and Liu, C.-L., 2016, **Fast Aircraft Detection Using End-to-End Fully Convolutional Network, Digital Signal Processing (DSP)**, 2016 IEEE International Conference on. IEEE, pp. 139-143.
- Yoo, D., Park, S., Lee, J.-Y. and Kweon, I.S., 2015, **Multi-scale pyramid pooling for deep convolutional representation**, Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on. IEEE, pp. 71-80.
- Yosinski, J., Clune, J., Bengio, Y. and Lipson, H., 2014, **How transferable are features in deep neural networks?**, Advances in neural information processing systems, pp. 3320-3328.
- Yuan, P., Zhong, Y. and Yuan, Y., 2017, **Faster R-CNN with Region Proposal Refinement**.
- Zhang, W., Sun, X., Fu, K., Wang, C. and Wang, H., 2014, **Object detection in high-resolution remote sensing images using rotation invariant parts based model**, IEEE Geoscience and Remote Sensing Letters 11, 74-78.



## Target detection from high-resolution remote sensing images using deep learning methods

Farhadi, N.<sup>\*1</sup>, Kiani, A.<sup>2</sup>, Ebadi, H.<sup>3</sup>

1- Department of Geomatics Engineering, Babol Noshirvani University of Technology, Iran  
2- Ph.D. student in Geomatics Engineering Faculty, K.N.Toosi University of Technology  
3- Professor in Geomatics Engineering Faculty, K.N.Toosi University of Technology

### Abstract

Object detection is one of the fundamental issues in image interpretation process, especially from remote-sensing imagery. One of the most effective and efficient methods in this field is the use of deep learning algorithm for feature extraction and interpretation. An object is a collection of unique patterns that differ with own adjacent properties. This difference usually occurs in one or more features simultaneously, which can be indicated by the difference in shape, color, and gray values. In this regard, the use of deep learning as an efficient branch of machine learning can be useful in generating high-level concepts through learning in different layers. In this research, a database based on the environmental and geographical conditions from some Iranian airports was created. Additionally, an optimal learner model was developed with a convolutional neural network. For this purpose, in the raw data processing section, besides using the transfer learning method, some vectors were extracted to classify the objects and delivered to an SVM model. The output values were compared with the values obtained from the test image for each object, and they were analyzed in a repeatable process for structural matching. Precision of 98.21% and F1-Measure of 99.1% was achieved, for identification of the target objects.

**Keywords:** Deep Learning, Neural Network, Remote Sensing imagery, Machine Learning, Transfer learning