

Recognition and Tracking of Aerial Targets Using Convolutional Neural Network

A. Jahed Saravani^{ID}

*Assistant Professor, Khatam Al-Anbiya Air Defense University, Tehran, Iran

(Received: 2022/09/08, Revised: 2022/12/17, Accepted: 2023/01/01, Published: 2023/01/21)

DOR: <https://dor.isc.ac/dor/20.1001.1.23454024.1401.10.2.2.3>

Abstract

Automatic recognition and tracking systems of aerial targets are of particular importance in the battle field. These types of systems use visual sensors, have the ability to be installed on various military systems, and are suitable for discovering and tracking low-altitude targets. In this manuscript, a convolutional neural network was designed to recognize the type of aerial targets (cargo, aerobatics, fighter and missile) and then target tracking using a pre-trained network (GoogLeNet) and transfer learning in the form of a region with convolutional neural network was done. The recognition accuracy of aerial targets in the test data set is 96.3%. On the other hand, the overlap value between the actual and predicted bounding box of target in the test data set for cargo and aerobatics plane, fighter and missile is 0.61, 0.66, 0.64 and 0.51, respectively, which shows the desirable accuracy of the developed model for targets tracking in consecutive frames.

Keywords: : Aerial targets, Target recognition, Target tracking, CNN, RCNN.

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Publisher: Imam Hussein University

Authors



*Corresponding Author Email: A_jahed@khadu.ac.ir

تشخیص و ردیابی اهداف هوایی با استفاده از شبکه‌های عصبی پیچشی

علی جاهد سراوانی^{id*}

استادیار دانشکده مهندسی برق، دانشگاه پدافند هوایی خاتم‌الانبیاء (ص)، تهران، ایران

(دریافت: ۱۴۰۱/۰۶/۱۷، بازنگری: ۱۴۰۱/۰۹/۲۶، پذیرش: ۱۴۰۱/۱۰/۱۱، انتشار: ۱۴۰۱/۱۱/۰۱)

DOR: <https://dor.isc.ac/dor/20.1001.1.23454024.1401.10.2.2.3>



* این مقاله یک مقاله با دسترسی آزاد است که تحت شرایط و ضوابط مجوز (CC BY) Creative Commons Attribution توزیع شده است.

نویسندگان



ناشر: دانشگاه جامع امام حسین (ع)

چکیده

سیستم‌های تشخیص و ردیابی خودکار اهداف هوایی از اهمیت ویژه‌ای در صحنه نبرد برخوردار هستند. این نوع از سیستم‌ها از سنسورهای بصری استفاده کرده، قابلیت نصب بر روی سامانه‌های مختلف نظامی را داشته و برای کشف و ردیابی اهدافی با ارتفاع پست مناسب هستند. در این مقاله، یک شبکه عصبی پیچشی برای تشخیص نوع اهداف هوایی (هواپیمای باری، نمایشی، جنگنده و موشک) طراحی گردیده و در ادامه ردیابی هدف با استفاده از یک شبکه پیش آموزش‌دیده (GoogLeNet) و یادگیری انتقالی در قالب شبکه عصبی پیچشی مبتنی بر ناحیه انجام شد. دقت شناسایی اهداف هوایی تعریف شده در داده‌های تست برابر با ۹۶/۳٪ می‌باشد. از طرف دیگر میزان همپوشانی چارچوب واقعی و پیش‌بینی‌شده هدف در داده‌های تست برای هواپیمای باری و نمایشی، جنگنده و موشک به ترتیب برابر ۰/۶۱، ۰/۶۶، ۰/۶۴ و ۰/۵۱ می‌باشد که نشان از دقت مطلوب مدل توسعه داده شده برای ردیابی اهداف در قاب‌های متوالی می‌باشد.

کلیدواژه‌ها: اهداف هوایی، شناسایی هدف، ردیابی هدف، RCNN، CNN

۱- مقدمه

حرکت دوربین، تغییرات ناگهانی در وضعیت شیء هدف و حتی مخفی شدن آن در پس زمینه دارد [۳، ۵-۸]. در کاربردهای نظامی اهداف هوایی رادار گریز توسط سنسورهای بصری به سادگی تشخیص داده شده و با دقت بالا ردیابی می‌شوند [۹]. سنسورهای بصری در برابر تجهیزات راداری بسیار ارزان قیمت بوده و از عملکردی منفعل برخوردارند. همچنین، این سنسورها در برابر اختلال^۵ بسیار مقاوم بوده و قابلیت فراهم نمودن اطلاعاتی با جزئیات بسیار زیاد از هدف را دارا می‌باشند. از دیگر ویژگی‌های کلیدی این نوع از سنسورها، قابلیت نصب آنها بر روی تجهیزات مختلف از جمله پهپادها می‌باشد [۱۰]. در این حالت پهپادها در صورت روبرو شدن با اهداف پرنده مختلف قادر به تشخیص و ردیابی آنها به صورت خودکار می‌باشند [۱۱]. با توجه به ویژگی‌های متمایز این نوع از سنسورها و همچنین پیشرفت‌های

تشخیص و ردیابی بصری اشیاء مختلف به صورت برخط از موضوعاتی است که کاربردهای گسترده‌ای در دوربین‌های نظارتی، پهپادها، سیستم‌های نوری^۲ و سامانه‌های ATR^۳ از جمله تجزیه و تحلیل حرکت و شناسایی نوع فعالیت هدف مورد نظر دارد [۴-۱]. هدف از ردیابی بصری تخمین مرز و مسیر حرکت شیء در دنباله‌ای از قاب‌ها می‌باشد [۲]. مهمترین چالش در الگوریتم‌های ردیابی برخط، تغییرات تدریجی رخ داده در شیء هدف و پس زمینه آن در طی زمان می‌باشد. این تغییرات دلایل متفاوتی چون تغییرات شدت روشنایی، تار شدن قاب‌های متوالی،

* رایانامه نویسنده مسئول: A_jahed@khadu.ac.ir

² Optical systems

³ Automatic target recognition

⁴ Frames

⁵ Jamming

چشمگیر در علم بینایی کامپیوتر^۱ مطالعات بسیار زیادی در دهه گذشته در این زمینه انجام شده است.

تشخیص ناحیه مطلوب (ROI)^۲ در تصویر بر اساس ویژگی‌های مشابه در پیکسل‌های تصویر و یا تطبیق مشخصه^۳ صورت می‌پذیرد. در روش اول از تفاوت ویژگی‌های پیکسل‌های شی و پس زمینه مانند کنتراست، رنگ و ... در هر قاب استفاده می‌شود. کنتراست جزء آن دسته از ویژگی‌هایی است که با استفاده از روش‌های لبه‌یابی^۴ کلاسیک و هوشمند مشخص گردیده و در ادامه با اعمال فیلترهای مورفولوژیکی شی مورد نظر از پس‌زمینه به سادگی تفکیک می‌گردد. فیلترهای مورفولوژیکی طیف وسیعی از فیلترهای پردازش تصویر دیجیتال بوده که بر اساس شکل تصاویر عمل می‌کنند. در این فیلترها، هر پیکسل تصویر بر اساس پیکسل‌های همسایه خود پردازش می‌شود [۱۲، ۱۳]. از دیگر ویژگی‌های قابل استفاده، تفاوت رنگ می‌باشد. ناحیه‌بندی بر اساس رنگ^۵ با استفاده از خوشه‌بندی^۶ انجام می‌گیرد. در این روش پیکسل‌های تصویر در دو خوشه پیش‌زمینه (شی مورد نظر) و پس زمینه با استفاده از روش‌های خوشه‌بندی کلاسیک (مانند K-means) و یا هوشمند (مانند FCM^۷ یا SOM^۸) قرار گرفته و از یکدیگر تفکیک می‌شوند. این فرآیند در قاب‌های متوالی تکرار شده و شی مورد نظر ردیابی می‌شود [۱۴، ۱۵].

در روش‌های ذکر شده، ناحیه‌بندی بر اساس ویژگی‌های پیکسل‌های پیش زمینه و پس زمینه در هر قاب به صورت مستقل انجام شده و شی مورد نظر کشف می‌شود. در روش‌های مبتنی بر تطبیق مشخصه که امروزه از آن‌ها به صورت گسترده استفاده می‌شود، در ابتدا هدف در قاب ابتدایی انتخاب شده، مشخصات آن استخراج گردیده، و در ادامه این مشخصات در دنباله‌ای از قاب‌ها برای شناسایی کاندیدای هدف جستجو می‌شوند. در نهایت هدف یافته شده با مشخصات مشابه از پس‌زمینه تصویر متمایز می‌گردد. در این الگوریتم، ردیابی مشخصه‌ها و چگونگی ردیابی آنها از اهمیت بسزایی برخوردار بوده و نقشی تعیین کننده در عملکرد نهایی الگوریتم دارد. در

الگوریتم‌های SIFT^۹ و HOG^{۱۰}، ویژگی‌های سطح پایین مثل شکل، محتوا و ... استخراج گردیده و این ویژگی‌ها در قاب‌های بعدی به منظور انتخاب صحیح هدف بر اساس معیارهای مشابهت‌سنجی مانند فاصله اقلیدسی یا همبستگی متقاطع^{۱۱} جستجو و یافت می‌شود. استفاده از ویژگی‌های سطح پایین در این الگوریتم‌ها منجر به عدم پایداری آنها در برابر تغییرات رخ داده شده در قاب‌های متوالی مانند جابجایی، چرخش، تغییر زاویه دید، تغییر شدت روشنایی، تغییر مقیاس و ... می‌شود [۱۶-۲۲]. ASIFT^{۱۲} نسخه‌ای اصلاح شده الگوریتم SIFT بوده که از ثبات بیشتری در برابر تغییرات ذکر شده برخوردار است و قادر به استخراج نقاط کلیدی بیشتری می‌باشد [۱۵]. روش تطبیق الگو^{۱۳} معمولاً در تصاویری استفاده می‌شوند که دارای ویژگی‌هایی آشکار نیستند. در این روش، تصویر الگو در ابتدا در اختیار الگوریتم قرار گرفته و شبیه‌ترین الگو در قاب‌های بعدی یافت می‌شود. محاسبات بسیار سنگین و زمان‌بر و همچنین ناتوانی این الگوریتم در برابر کلاتر^{۱۴} و مخفی شدن هدف از اشکالات این روش می‌باشد [۲۳، ۲۴]. برای مقابله با پوشیدگی و یا انسداد^{۱۵} نسبی هدف و آشکارسازی دوباره آن از روش‌های مبتنی بر هسته استفاده می‌شود. در این روش، عامل موثر اطلاعات ذخیره شده هدف در قاب‌های قبلی می‌باشد. بنابراین، در مواقعی که قسمتی از هدف در پس‌زمینه مخفی می‌شود، موقعیت آن با توجه به قاب‌های قبلی جستجو می‌گردد. از فیلتر کالمن نیز می‌توان به عنوان ابزاری برای جستجوی و تخمین هدف انفرادی و نه چندگانه استفاده نمود [۲۵]. روش انتقال میانگین^{۱۶} یک رویکرد تخمین گرادیان چگالی غیرپارامتریک می‌باشد که برای ردیابی نیازمند تعریف مدل هدف می‌باشد. این مدل بر اساس هیستوگرام رنگ هدف متحرک بدست می‌آید. کاندیدای هدف نیز به همین روش مشخص می‌گردد. این الگوریتم در برابر تغییرات تدریجی تصویر از عملکردی مطلوب برخوردار بوده اما به سادگی همگرا نشده و دارای سرعت مطلوبی نیست [۲۶، ۲۷].

^۹ Scale-invariant feature transform

^{۱۰} Histogram of oriented gradients

^{۱۱} Cross correlation

^{۱۲} Affine scale-invariant feature transform

^{۱۳} Template Matching

^{۱۴} Clutter

^{۱۵} Occlusion

^{۱۶} Mean-shift

^۱ Computer vision

^۲ Region of interest

^۳ Feature matching

^۴ Edge detection

^۵ Color based segmentation

^۶ Clustering

^۷ Fuzzy C-mean

^۸ Self-organizing map

نوع و ردیابی اهداف هوایی به ترتیب در بخش سوم و چهارم ارائه شده است. در بخش پنجم نیز نتیجه‌گیری نهایی صورت پذیرفته است.

۲- شبکه‌های عصبی پیچشی (CNNs)

شبکه‌های عصبی پیچشی نسخه‌ای تعمیم یافته از شبکه‌های عصبی مصنوعی^۵ بوده که ساختار ارتباطاتی آن از غشا بصری مغز الهام گرفته شده و توانایی استخراج محدوده وسیعی از ویژگی‌های بسیار ساده تا پیچیده تصاویر دیجیتال را با عبور از لایه‌ای به لایه دیگر را دارا می‌باشند. این نوع از شبکه‌ها ظرفیت برخورداری از لایه‌های متعددی را داشته و در نتیجه دست طراحان را برای توسعه ساختارهای متنوعی از آنها باز گذاشته است. در ابتدا تصاویر دیجیتال به لایه ورودی CNN وارد شده و با عبور از لایه‌های متعدد کانولوشن^۶، ادغام^۷ و واحد یکسوشده خطی^۸ پردازش می‌گردند. در لایه‌های ابتدایی، ویژگی‌های سطح پایین (مثل لبه‌ها) و در لایه‌های عمیق‌تر ویژگی‌های پیچیده و سطح بالای تصویر استخراج می‌گردد. هر پیکسل تصویر در شبکه‌های عصبی کلاسیک به عنوان یک ویژگی ورودی در نظر گرفته شده و بنابراین تعداد پارامترهای (وزن‌ها) آنها در مواجهه با تصاویر دو بعدی دیجیتال به شدت افزایش یافته و بالطبع حجم محاسبات و احتمال بیش‌برازش^۹ در پایگاه داده‌هایی با اندازه کوچک، افزایش می‌یابد [۳۹، ۴۰]. تعداد پارامترهای CNN با اعمال فیلتر کانولوشن کنترل شده و ویژگی‌های تصویر توسط آنها بدون از دست دادن اطلاعات تصویر استخراج می‌گردند. در این لایه، فیلتر کانولوشن به صورت افقی و عمودی بر روی تصویر حرکت کرده و ضرب داخلی المان‌های فیلتر و پیکسل‌های تصویر انجام می‌گردد. در واقع ویژگی‌های محلی تصویر در این فرآیند، اسکن و استخراج می‌شود. خروجی فیلتر کانولوشن نقشه ویژگی^{۱۰} نامیده شده و به لایه ادغام منتقل می‌شود. در این لایه فرآیند زیرنمونه‌برداری^{۱۱} انجام شده و اندازه نقشه ویژگی به منظور تسریع هرچه بیشتر محاسبات کاهش می‌یابد. لایه بعدی حاوی یک واحد یکسوشده خطی است که تابعی فعال‌ساز بوده و خروجی غیرخطی تولید می‌کند. در نهایت، تمامی ویژگی‌های استخراج شده به لایه کاملاً همبند^{۱۲} انتقال و در آنجا ترکیب می‌شوند. این لایه متشکل از تعدادی نرون بوده که هر کدام به

در روش تطبیق مشخصه، ویژگی‌های تصویر با استفاده از روش‌های مختلف پردازش و استخراج می‌گردند. در الگوریتم‌های توسعه داده شده، هر الگوریتم قادر به استخراج دسته‌ای از ویژگی‌های تصویر و نه همه آنها (به صورت جامع) می‌باشد. از طرف دیگر سرعت استخراج این ویژگی‌ها کاملاً به چگونگی استخراج آنها بستگی دارد. امروزه، با توسعه و گسترش مفهوم یادگیری ژرف این نوع چالش‌ها به کلی حل و به حداقل رسیده است [۲۸، ۲۹]. در نوع خاصی از یادگیری ژرف، ویژگی‌های تصویر چه سطح پایین و چه سطح بالا به صورت جامع و مهم‌تر از آن به صورت اتوماتیک استخراج شده و در طبقه‌بندی تصویر قابل استفاده می‌باشد. در مرجع [۳۰]، از یک شبکه باور عمیق^۱ به همراه یک روش مکان‌یابی هدف برای تشخیص هواپیماهای بسیار کوچک در کف باند فرودگاه استفاده شده است. روش مکان‌یابی در این پژوهش بر اساس تعریف دو پنجره لغزان مشخص و محاسبه گرادیان در قاب‌های متوالی انجام گرفته که البته روشی زمان‌بر می‌باشد. شبکه‌های عصبی پیچشی (CNNs) نوعی خاص از شبکه‌های عصبی بوده که مستقیماً توسط تصاویر دیجیتال تغذیه شده و قادر به استخراج خودکار ویژگی‌های تصاویر به صورت جامع می‌باشد. مطالعات انجام شده حاکی از آن است که رویکردهای ردیابی با استفاده از ویژگی‌های استخراج شده توسط CNNs از دقت بالاتری در برابر روش‌هایی با استخراج ویژگی به صورت کلاسیک برخوردار هستند [۲۸، ۳۱-۳۳]. در سالیان اخیر، از این نوع شبکه‌ها به صورت تنهایی و یا ترکیبی با روش‌های دیگر برای تشخیص و مکان‌یابی تصاویر ماهواره‌ای، تصاویر پهپادها و ... استفاده گردیده است [۱، ۲۸، ۳۱، ۳۴-۳۶]. مهم‌ترین چالش در استفاده از شبکه‌های ژرف، حجم زیاد داده‌های مورد استفاده، برچسب‌گذاری دستی داده‌ها (کادر محصور کننده^۳ یا چارچوب) و حجم بالای محاسباتی آنها در زمان یادگیری و تست می‌باشد [۳۵، ۳۷، ۳۸].

در این مقاله، ابتدا یک شبکه عصبی پیچشی به منظور تشخیص نوع اهداف هوایی مختلف طراحی و آموزش داده شده است. در ادامه، از شبکه پیش آموزش دیده GoogLeNet در قالب یادگیری انتقالی^۴ به منظور ردیابی هدف در قاب‌های متوالی استفاده گردیده است. در بخش دوم مقاله، شبکه‌های عصبی پیچشی و لایه‌های آن به صورت خلاصه مرور شده است. تشخیص

^۵ Artificial Neural Network

^۶ Convolution

^۷ Pooling

^۸ Rectified linear unit

^۹ Over-fitting

^{۱۰} Feature map

^{۱۱} Sub-sampling

^{۱۲} Fully connected layer

^۱ Deep belief network

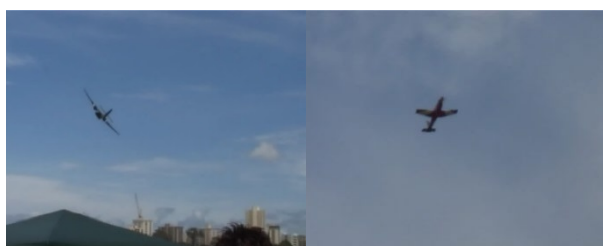
^۲ Convolutional neural networks

^۳ Bounding box

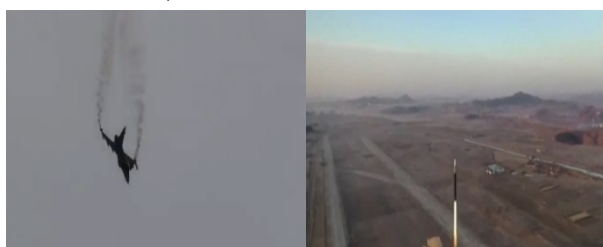
^۴ Transfer learning

جدول (۱): پایگاه داده تصویری استفاده شده برای آموزش

تعداد قاب‌ها	Label	ویدئو
۶۳۳	Cargo	هواپیمای باری
۶۴۴	Aerobatics	هواپیمای نمایشی
۱۲۴۴	Fighter	جنگنده
۱۳۶۹	Missile	موشک



(الف) (ب)



(د) (ج)

شکل (۱): تصاویر نمونه‌ای از قاب‌ها با برجسب‌های مختلف لیست شده در جدول (۱) [۱۰]

۲-۳- طراحی شبکه پیشنهادی CNN و آموزش آن

مهم‌ترین فاکتورها برای تشخیص و ردیابی اهداف هوایی دقت و سرعت الگوریتم پیشنهادی می‌باشد. به همین منظور سعی شده است، معماری CNN پیشنهادی از حداکثر سادگی ممکن برای افزایش سرعت الگوریتم برخوردار باشد. معماری CNN پیشنهادی با ۱۶ لایه متوالی در شکل (۲) نشان داده شده است. در لایه اول، اندازه تصاویر دیجیتال رنگی به منظور کاهش حجم محاسبات به اندازه 224×224 پیکسل تغییر می‌کند. در لایه بعدی از هشت فیلتر کانولوشن استفاده کرده و نقشه ویژگی استخراج می‌گردد. نقشه ویژگی به لایه نرمال‌ساز دسته‌ای^۱ و سپس به تابع فعال‌ساز یکسوسوده خطی منتقل می‌گردد. در لایه نرمال‌ساز دسته‌ای، سرعت فرآیند یادگیری افزایش و از طرف دیگر حساسیت CNN به وزن‌دهی اولیه کاهش می‌یابد. سپس فرآیند زیرنمونه‌برداری در لایه ادغام انجام شده و اندازه نقشه ویژگی برابر 112×112 می‌شود. حال تمامی این مراحل دو مرتبه دیگر با استفاده از ۱۶ و ۳۲ فیلتر کانولوشن به منظور استخراج ویژگی‌های سطح بالای تصاویر دیجیتال انجام و اندازه

تمامی نرون‌ها در لایه قبلی متصل می‌باشد. لایه کاملاً همبند از اطلاعات دریافتی برای طبقه‌بندی تصویر ورودی استفاده می‌کند [۴۱، ۴۲]. در معماری‌های مختلف بعضاً جای این لایه‌ها و یا تعداد آنها تغییر می‌کند.

به طور خلاصه، فیلترهای کانولوشن به‌کاررفته در شبکه‌های عصبی پیچشی ابعاد بزرگ تصاویر دیجیتال را بدون ازدست‌رفتن اطلاعات آنها، کاهش داده و پردازش می‌کند. علاوه بر این، طیف گسترده‌ای از ویژگی‌های تصویر در این شبکه‌ها به‌صورت خودکار و جامع استخراج شده و دیگر نیاز به طراحی الگوریتم‌های مختلف برای استخراج ویژگی‌های تصویر (به‌صورت دستی) نمی‌باشد. ضمناً، باتوجه‌به ذات لایه‌لایه این نوع از شبکه‌ها، یادگیری در آنها با عمق بیشتری رخ داده و در نتیجه از دقت بسیار بالاتری نسبت به شبکه‌های عصبی کلاسیک برخوردار هستند. در این مقاله، ساختاری برای شبکه CNN با لایه‌های مختلف به‌منظور تشخیص نوع اهداف هوایی مختلف پیشنهاد شده و در ادامه نیز برای ردیابی هدف در قاب‌های متوالی، از یادگیری انتقالی بر مبنای شبکه CNN پیش آموزش‌دیده استفاده می‌شود.

۳- تشخیص نوع هدف

در الگوریتم تشخیص نوع هدف، ابتدا بایستی پایگاه‌داده تصویری بزرگی از تصاویر دیجیتال تشکیل و ساختاری برای CNN پیشنهادی ارائه گردد. در ادامه، فرآیند یادگیری مدل پیشنهادی توسط داده‌های آموزش انجام و عملکرد آن توسط داده‌های تست چک می‌شود.

۳-۱- پایگاه داده تصویری

به‌منظور تشخیص و ردیابی اهداف متحرک هوایی می‌بایست یک پایگاه‌داده تصویری بزرگ و قابل‌ارائه به CNN تشکیل گردد. اهداف در نظر گرفته شده در این مقاله شامل جنگنده، موشک (در لحظه شلیک و در حین پرواز)، هواپیمای باربری و نمایشی می‌باشند. این ویدئوها ترکیبی از پایگاه‌داده فراهم شده در مرجع [۱۰] و ویدئوهای موجود در اینترنت می‌باشد. در این ویدئوها چالش‌های بسیاری از قبیل تغییر مقیاس هدف، مانور سریع و چرخش‌های متعدد هدف، ایجاد شعله و دود، مخفی شدن هدف در پس‌زمینه شلوغ، بازتاب نور از روی سطح هدف و ... وجود دارد. قاب‌های موجود در هر ویدئو استخراج و تعداد آنها در جدول (۱) نشان داده شده است. به منظور آشنایی هر چه بیشتر با نوع تصاویر، تعدادی از تصاویر موجود در پایگاه داده تصویری در شکل (۱) نشان داده شده است.

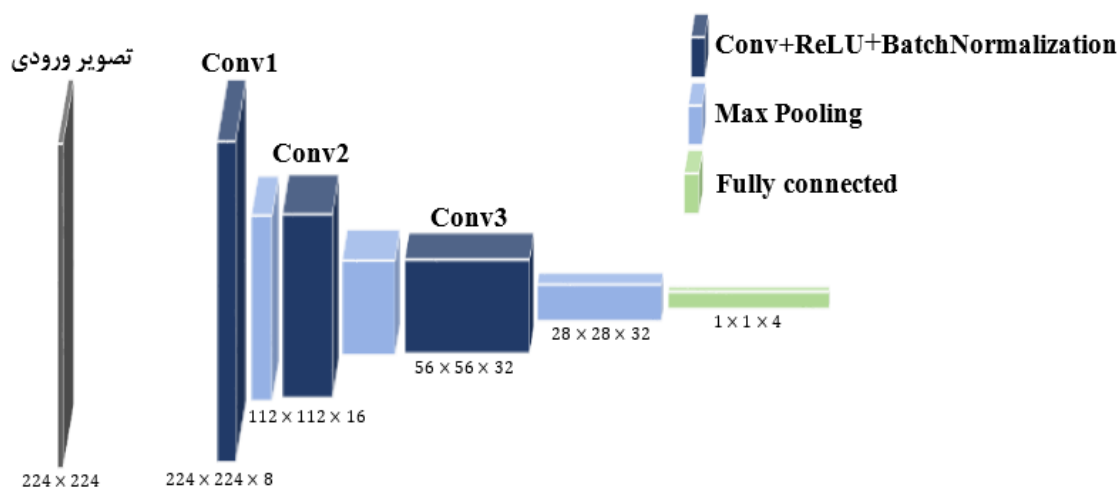
¹ Batch normalization layer

نقشه ویژگی کوچک و کوچک‌تر می‌شود. در انتها از یک لایه کاملاً همبند با ۴ نرون خروجی (منطبق بر تعداد برجسب‌های موجود در پایگاه تصویری) و لایه بیشینه هموار^۱ استفاده می‌شود. در این لایه، احتمال تعلق هر تصویر ورودی به هر یک از کلاس‌ها محاسبه گردیده و تصویر بر اساس حداکثر مقدار محاسبه شده برجسب‌گذاری می‌شود. تعداد پارامترها و اطلاعات لازم CNN پیشنهادی در جدول (۲) لیست شده است. همانگونه که در شکل ۲ مشهود است در طراحی شبکه پیشنهادی CNN سعی شده از تعدد لایه‌ها به منظور افزایش هر چه بیشتر سرعت شبکه در شناسایی هدف، پرهیز شود. علاوه بر این، ابعاد نقشه ویژگی در سه بلوک متوالی به منظور ارتقاء سرعت استخراج ویژگی توسط لایه‌های ادغام به صورت سلسله‌وار کاهش یافته است. از طرف دیگر، طیف گسترده‌ای از ویژگی‌های سطح پایین تا پیچیده با استفاده از فیلترهای کانولوشن تعریف شده استخراج می‌گردد. لایه‌های نرمال‌ساز دسته‌ای نیز به دلیل آموزش هر چه دقیق‌تر CNN پیشنهادی در این معماری گنجانده شده‌اند. از ۸۰ درصد داده‌های آموزش برای شروع فرآیند آموزش استفاده کرده و از باقی داده‌ها نیز به عنوان داده‌های اعتباریابی^۲ به منظور چک‌کردن فرآیند آموزش استفاده می‌شود. منحنی دقت آموزش در شکل (۳) نشان داده شده است. همانگونه که مشاهده می‌شود مدل آموزش‌دیده از دقت بسیار بالایی در داده‌های آموزش و اعتباریابی (بیش از ۹۹ درصد) برخوردار است. دقت داده‌های اعتباریابی پس از هر ۱۰ دوره چک شده و دقت بالای آن نشان از این دارد که CNN پیشنهادی قادر به پیش‌بینی داده‌های تست (داده‌های خارج از داده‌های آموزش^۳) نیز خواهد بود (قابلیت تعمیم‌پذیری بالایی دارد). برای تست نهایی CNN پیشنهادی از ویدئوهایی غیر از آنچه در جدول (۱) لیست شده، استفاده شده است. اطلاعات این داده‌ها در جدول (۳) نشان داده شده است. این ویدئوها در شرایطی متفاوت از داده‌های آموزش، تصویربرداری شده‌اند.

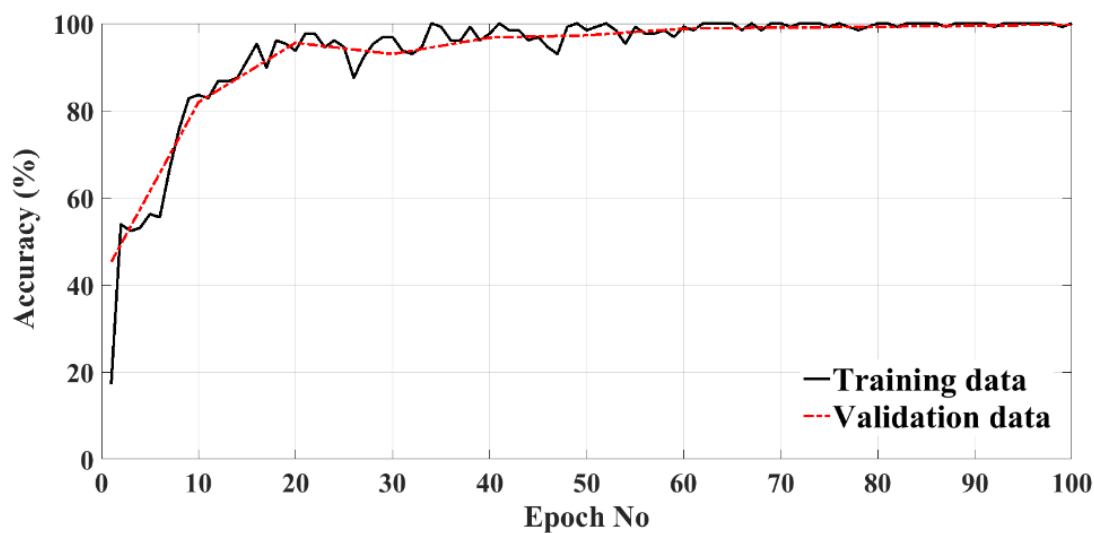
¹ Softmax layer

² Validation data

³ Training data



شکل (۲): معماری پیشنهادی CNN



شکل (۳) فرآیند آموزش CNN پیشنهادی

۹	ادغام	روش ماکزیمم 2 × 2	56 × 56 × 16
۱۰	کانولوشن	32(3 × 3 × 16)	56 × 56 × 32
۱۱	نرمال ساز دسته‌ای	۳۲ کانال	56 × 56 × 32
۱۲	فعال ساز یک‌سو شده خطی	-	56 × 56 × 32
۱۳	ادغام	روش ماکزیمم 2 × 2	28 × 28 × 32
۱۴	کاملاً هم‌بند	۵	1 × 1 × 5
۱۵	بیشینه هموار	-	1 × 1 × 5
۱۶	طبقه‌بندی کننده	-	1 × 1 × 5

جدول (۲): اطلاعات لایه‌های CNN پیشنهادی

ردیف	نام لایه	ویژگی لایه	اندازه
۱	ورودی تصویر	-	224 × 224
۲	کانولوشن	8(3 × 3 × 3)	224 × 224 × 8
۳	نرمال ساز دسته‌ای	۸ کانال	224 × 224 × 8
۴	فعال ساز یک‌سو شده خطی	-	224 × 224 × 8
۵	ادغام	روش ماکزیمم 2 × 2	112 × 112 × 8
۶	کانولوشن	16(3 × 3 × 8)	112 × 112 × 16
۷	نرمال ساز دسته‌ای	۱۶ کانال	112 × 112 × 16
۸	فعال ساز یک‌سو شده خطی	-	112 × 112 × 16

شکل (۴): ماتریس درهم‌ریختگی داده‌های تست برای شناسایی چهار کلاس از اهداف مختلف

در شکل (۴)، صحت و بازیابی در آخرین ردیف و ستون ماتریس برای هر کلاس لیست شده‌اند. صحت و بازیابی برای هر کلاس در یک طبقه‌بندی‌کننده با عملکرد مطلوب می‌بایست حداکثر باشند. در ماتریس درهم‌ریختگی نشان داده شده، از کل تعداد هواپیماهای باربری در داده‌های تست ۲٪، ۵٪ و ۳٪ به اشتباه جنگنده، موشک و هواپیمای نمایشی تشخیص داده شده است. از طرف دیگر ۹٪ از جنگنده‌ها نیز به اشتباه به عنوان هواپیمای نمایشی تشخیص داده شده‌اند. برای سیستم‌های تصمیم‌گیر خودکار در پهپادها، این حالت‌ها بسیار حائز اهمیت می‌باشند. مقدار این نوع از خطاها که از آنها به عنوان خطاهای نوع اول و دوم یاد می‌شود، در طبقه‌بندی‌کننده پیشنهادی برای سه هدف هواپیمای نمایشی، موشک و جنگنده مطلوب و برای هواپیمای باربری قابل قبول می‌باشد.

منحنی مشخصه عملکرد (ROC)^۴ ابزاری دیگر برای ارزیابی کمی عملکرد طبقه‌بندی‌کننده پیشنهادی می‌باشد. در این منحنی نرخ مثبت‌های صحیح برای هر کلاس بر اساس نرخ مثبت کاذب در مقادیر آستانه مختلف رسم می‌شود. هر چه منحنی‌های نشان‌داده‌شده برای هر کلاس در شکل (۵) به سمت راست و بالای شکل متمایل باشند (بدون مثبت و منفی کاذب)، عملکرد طبقه‌بندی‌کننده در آن کلاس مطلوب‌تر می‌باشد. همان‌گونه که مشخص است طبقه‌بندی‌کننده پیشنهادی فقط در تشخیص هواپیمای باربری عملکرد چندان مناسبی ندارد. در شکل (۵)، مقدار AUC^۵ (سطح زیر منحنی ROC) نیز برای هر کلاس نشان‌داده‌شده است. هر چه این مقدار به یک نزدیک‌تر باشد عملکرد طبقه‌بندی‌کننده پیشنهادی برای آن کلاس مناسب‌تر است.

۴- ردیابی هدف

به‌منظور ردیابی هدف در قاب‌های متوالی، ابتدا چارچوب هدف در داده‌های آموزش توسط روش‌های ترکیبی خودکار و دستی برای هر قاب تعیین می‌شود. سپس از ساختار یک نوع CNN پیش آموزش‌دیده^۶ و یادگیری انتقالی در ساختار شبکه‌های پیچشی

به‌منظور ارزیابی طبقه‌بندی‌کننده آموزش‌دیده، از پارامترهای دقت^۱، صحت^۲ و بازیابی^۳ برای هر کلاس استفاده شده است.

جدول (۳): پایگاه داده تصویری استفاده شده برای داده‌های تست

برچسب ویدئو	Label	تعداد قاب‌ها
هواپیمای باری	Cargo	۲۰۱
هواپیمای نمایشی	Aerobatics	۴۱۵
جنگنده	Fighter	۴۹۶
موشک	Missile	۱۲۳۴

دقت طبقه‌بندی‌کننده نشان‌دهنده تعداد برچسب‌های پیش‌بینی شده صحیح به کل داده‌ها می‌باشد [۴۳]:

$$(۱) \quad \text{دقت} = \frac{t_p + t_n}{P + N}$$

که در آن P و N به ترتیب تعداد واقعی موردهای مثبت و منفی می‌باشند. t_p و t_n نیز تعداد مثبت و منفی‌هایی است که به صورت صحیح پیش‌بینی شده‌اند (مثب و منفی صحیح). دقت کلی طبقه‌بندی‌کننده در آخرین درایه ماتریس درهم‌ریختگی (پایین سمت راست) در شکل (۴) نشان داده شده و برابر ۹۶٪/۳ است. معادلات صحت و بازیابی نیز به صورت [۴۳]:

$$(۲) \quad \text{صحت} = \frac{t_p}{t_p + f_p}$$

$$(۳) \quad \text{بازیابی} = \frac{t_p}{t_p + f_n}$$

بوده که در آن f_p و f_n مثبت‌ها و منفی‌هایی هستند که به اشتباه برچسب‌گذاری شده‌اند (مثبت و منفی کاذب).

Output Class	Fighter	Missile	Aerobatics	Cargo	
	20.2%	0.0%	0.0%	2.0%	91.0%
	0.0%	52.6%	0.0%	0.5%	99.1%
	0.9%	0.0%	17.7%	0.3%	93.7%
	0.0%	0.0%	0.0%	5.8%	99.3%
Target Class	Fighter	Missile	Aerobatics	Cargo	
	95.8%	99.9%	100%	67.7%	96.3%

^۴ Receiver operating characteristic

^۵ Area under curve

^۶ Pre-trained CNNs

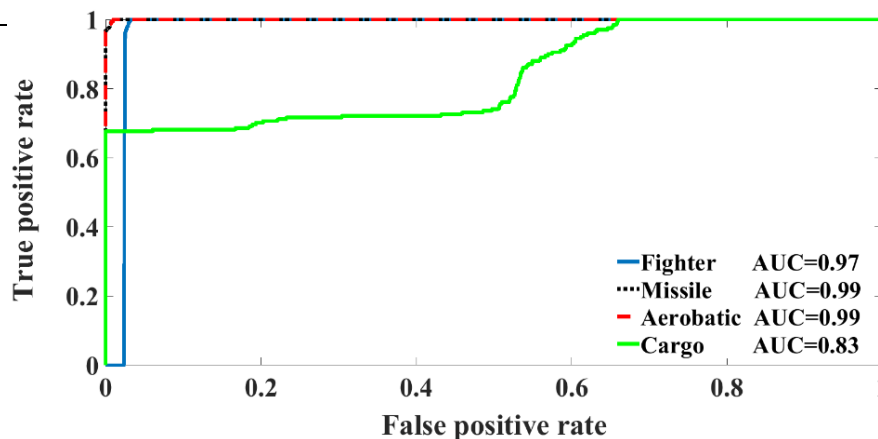
^۱ Accuracy

^۲ Precision

^۳ Recall

مبتنی بر ناحیه (RCNN)^۱ استفاده کرده و چارچوب هدف در داده‌های تست پیش‌بینی می‌شود.

¹ Regions with convolutional neural network



شکل (۵) منحنی ROC داده‌های تست برای شناسایی چهار کلاس مختلف

نسبت به الگوریتم استفاده شده در تصاویر جنگنده می‌باشد. در ویدئوهایی با مقیاس بسیار کوچک، با رنگی شبیه به پس‌زمینه و دارای پس‌زمینه شلوغ نیز راهی جز برچسب‌گذاری دستی نمی‌باشد (تصاویر موشک در حین پرواز).

۴-۲- شبکه‌های عصبی پیچشی مبتنی بر ناحیه (RCNN)

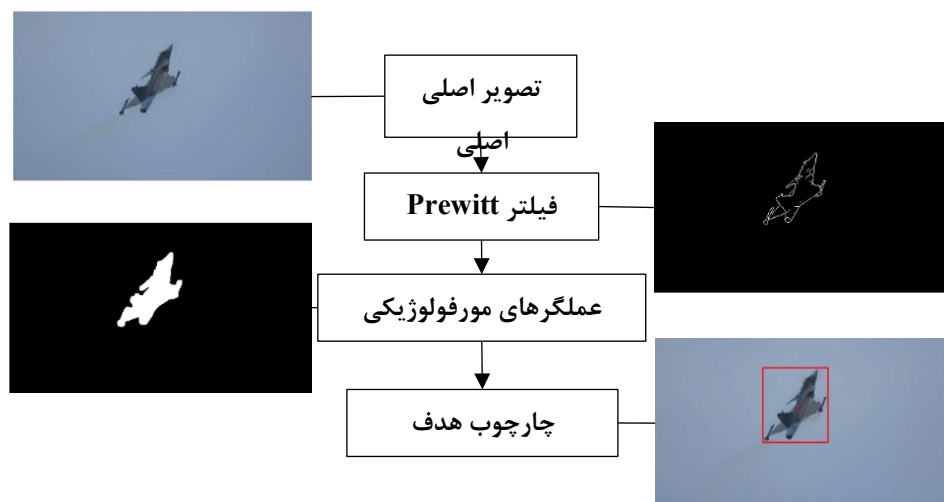
همان‌گونه که قبلاً توضیح داده شد از شبکه‌های عصبی پیچشی برای استخراج اتوماتیک ویژگی‌های تصاویر استفاده می‌شود. در بخش قبلی ویژگی‌های استخراج شده در لایه کاملاً هم‌بند گردآوری شده و سپس توسط لایه طبقه‌بندی‌کننده برای طبقه‌بندی تصاویر استفاده گردید. در این بخش لایه آخر را حذف کرده و به نحوی دیگر از ویژگی‌های استخراج شده استفاده می‌گردد. ردیابی بر اساس RCNN شامل سه مرحله می‌باشد. همان‌گونه که در شکل (۸) نشان داده شده است، در مرحله اول، تصاویر برای یافتن نواحی از تصویر با احتمال حضور شیء در آن، اسکن می‌شود. از این ناحیه‌ها به‌عنوان نواحی پیشنهادی (RP)^۲ یاد می‌شود. الگوریتم‌های متنوع و بی‌شماری برای یافتن RP های تصویر ارائه شده است. در این مقاله از روش حیطه‌های لبه^۳ برای این منظور استفاده شده است.

۴-۱- تعیین چارچوب واقعی هدف در داده‌های آموزش

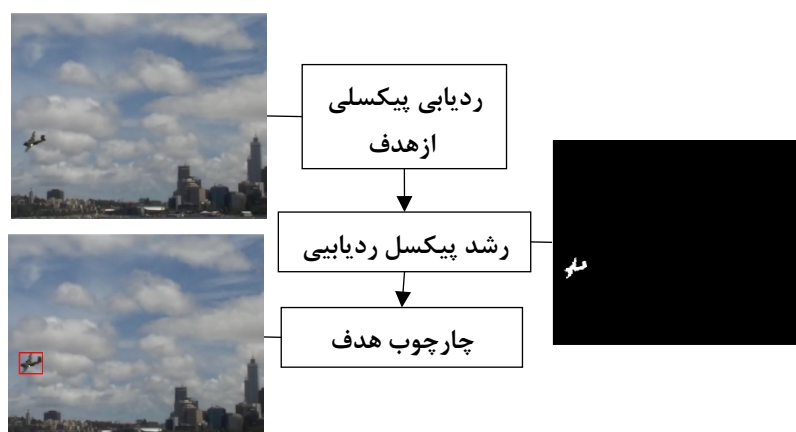
در مرحله نخست، چارچوب واقعی برای هر هدف مشخص در قاب‌های متوالی تصویر می‌بایست تعیین گردد. تعیین چارچوب به‌صورت دستی برای حجم زیادی از داده‌ها زمان‌بر می‌باشد. به همین منظور در این مقاله از روش‌های ترکیبی خودکار و دستی برای تعیین چارچوب متناسب مبتنی بر نوع ویدئو استفاده گردیده است. در ویدئوهایی که در آنها پس‌زمینه خلوت بوده، هدف انفرادی بوده و کاملاً از پس‌زمینه متمایز است (مانند تصاویر جنگنده) از روش‌های مبتنی بر لبه‌یابی کلاسیک استفاده شده است. در این روش تصاویر دیجیتال پیش‌پردازش شده، سپس هدف با استفاده از الگوریتم لبه‌یابی و فیلترهای مورفولوژیکی کشف می‌گردد. در شکل (۶) فلوچارت الگوریتم به‌صورت خلاصه نشان داده شده است. این الگوریتم بسیار سریع بوده و به‌صورت خودکار چارچوب هر هدف را در تصویر مشخص می‌کند. اما در بعضی حالت‌ها نیز (خروج دود یا بازتاب نور) دچار خطا می‌گردد. در این حالت‌ها، چارچوب هر هدف به‌صورت دستی تصحیح گردیده است.

در دیگر ویدئوها، پس‌زمینه کاملاً پیچیده بوده و مقیاس هدف در قاب‌های متوالی به‌شدت تغییر می‌کند (مانند تصاویر هواپیمای باری و نمایشی). در این نوع از ویدئوها از روش انباشتن سیلابی^۱ استفاده شده است. در این روش، تعدادی پیکسل از هدف موردنظر به‌صورت دستی انتخاب شده و پیکسل‌های متصل و پیکسل‌های مشابه به پیکسل انتخابی بر اساس فاصله اقلیدسی تا مرز ناحیه هدف به‌صورت خودکار رشد و هدف کشف می‌گردد (شکل (۷)). این الگوریتم بسیار کندتر، اما دارای دقت بالاتری

^۲ Region proposals^۳ Edge boxes^۱ Flood fill



شکل (۶): تعیین چارچوب واقعی هدف با استفاده از لبه‌یابی و فیلترهای مورفولوژیکی



شکل (۷): تعیین چارچوب واقعی هدف با استفاده از روش انباشتن سیلابی

الگوریتم ماشین بردار پشتیبان^۱ پالایش و چارچوب شئی یافته شده در RP مشخص می‌شود. در انتهای این روش، RPها با چارچوب‌های تعیین شده دوباره در مکان اصلی خود در تصویر مستقر شده و دوباره آن را بازسازی می‌کنند.

۳-۴- یادگیری انتقالی با استفاده از CNN پیش آموزش‌دیده

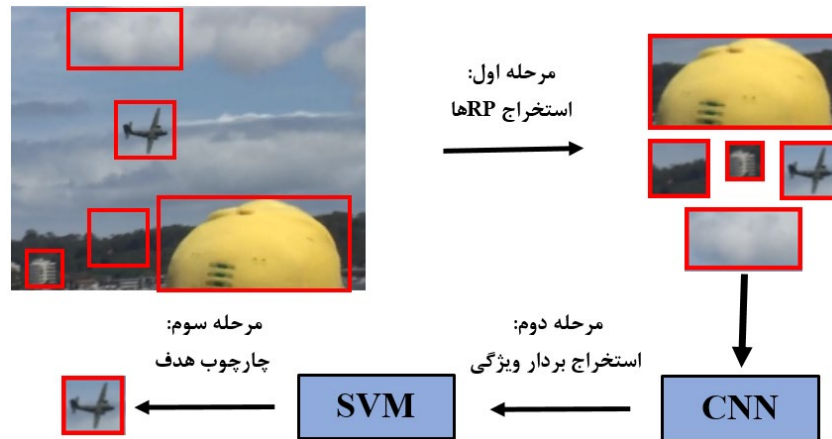
چگونگی استخراج ویژگی‌های RPهای تصویر حساس‌ترین مرحله در الگوریتم RCNN می‌باشد. برای اینکه ردیابی اهداف پرنده بادقت بالایی انجام شود، می‌بایست ویژگی‌های سطح بالای RPها استخراج گردد. به عبارت دیگر هر چه CNN پیشنهادی از عمق

همان‌گونه که می‌دانیم لبه‌های تصویر یک نمای کلی از تصویر و اشیاء داخل آن ارائه می‌دهد. در واقع احتمال حضور شئی در هر ناحیه‌ای از تصویر که در آن لبه باشد، وجود دارد، پس به سادگی با شناسایی لبه‌های تصویر، RPهای تصویر استخراج می‌گردند [۴۴]. در مرحله دوم، اندازه RPها برای ورود به CNN تغییر کرده و ویژگی‌های هر RP توسط شبکه عصبی پیچشی و با عبور از لایه‌های متوالی آن استخراج گردیده و در یک بردار ویژگی قرار می‌گیرد. در مرحله سوم، بردار ویژگی محاسبه شده توسط

^۱ Support vector machine

محاسبات بالا، محوشدگی و یا انفجار گرادبان را دارد. برای حل این مشکل می‌توان از یادگیری انتقالی استفاده نمود.

بیشتری برخوردار باشد، دقت پیش‌بینی چارچوب هدف افزایش می‌یابد. از طرف دیگر طراحی CNN با عمق زیاد مشکلات و چالش‌های خاص خود از قبیل عدم همگرایی، بیش‌برازش، حجم



شکل (۸) ک: الگوریتم RCNN

پپیچیدگی آن توسط ماژول آغازین کنترل می‌شود. این ماژول خود شامل لایه‌های کانولوشن، ادغام، واحد یکسوشده خطی و لایه الحاق^۲ می‌باشد. در این ماژول، تصویر ورودی توسط لایه‌های کانولوشن با اندازه‌های مختلف و به صورت موازی با یکدیگر پردازش گردیده و با عبور از واحدهای یکسوشده خطی و ادغام توسط لایه الحاق به ماژول آغازین بعدی تحویل داده می‌شود [۴۶، ۴۲]. با استفاده از معماری GoogLeNet تصویر با استفاده از لایه‌های کانولوشن با اندازه‌های مختلف و به صورت موازی اسکن شده و ویژگی‌های تصویر شامل جزئی‌ترین تا کلی‌ترین اطلاعات تصویر استخراج می‌گردد. در این مقاله برای ردیابی اهداف هوایی، لایه کاملاً همبند GoogLeNet به نحوی اصلاح گردیده تا خود را با پایگاه داده تصویری تهیه شده در این مقاله وفق داده و قادر به تمیز دادن هدف از پس‌زمینه تصویر باشد.

۴-۴- آموزش RCNN با استفاده از GoogLeNet

برای ردیابی هدف، قاب‌های استخراج شده در جدول (۱) را (به عنوان ورودی) به همراه چارچوب‌های واقعی مشخص شده برای آنها در بخش ۴-۱ (به عنوان خروجی) به RCNN داده و فرآیند آموزش انجام شده است. روند آموزش با استفاده از داده‌های آموزش در شکل (۹) نشان داده شده است. همانگونه که مشاهده می‌شود دقت در داده‌های آموزش در ۱۰۰۰ دوره به بیش از ۹۵٪ می‌رسد.

از یادگیری انتقالی معمولاً در یادگیری ژرف استفاده می‌شود. در این نوع از یادگیری، می‌توان از ساختار و پارامترهای CNN های توسعه‌یافته و تست شده در داده‌های بزرگ (مانند ImageNet [45]) و با یک سری تنظیمات اضافی جزئی یا کلی برای تطبیق هرچه بهتر با کاربرد جدید، استفاده نمود. در واقع، CNN پیش آموزش‌دیده یک نقطه شروع مناسب برای یادگیری بوده که می‌تواند با انتقال ساختار و پارامترهای خود، از طراحی و آموزش دوباره یک شبکه ژرف با انواع و اقسام چالش‌ها و مشکلات جلوگیری نماید. در این مقاله از GoogLeNet به عنوان یک CNN پیش آموزش‌دیده برای استخراج هر چه بهتر بردار ویژگی‌ها استفاده شده است.

ساختار GoogLeNet

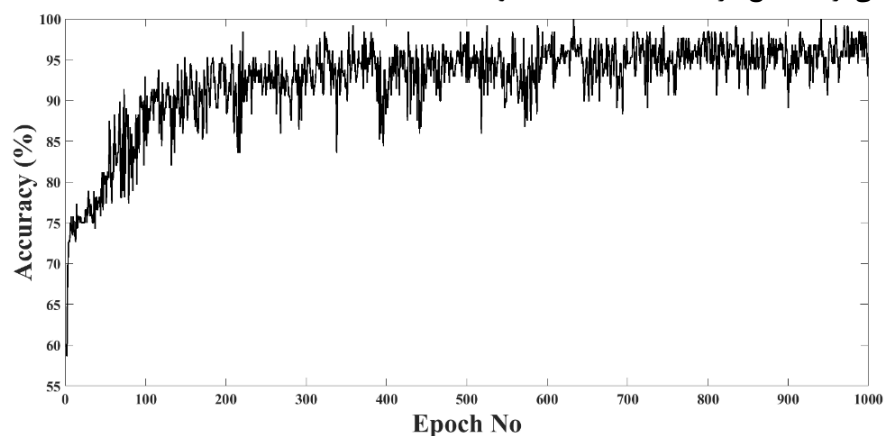
GoogLeNet نوعی CNN ژرف با ۲۷ لایه شامل ۲ لایه کانولوشن، ۵ لایه ادغام، ۹ ماژول آغازین^۱ و دارای بیش از ۷ میلیون پارامتر بوده که توسط پژوهشگران شرکت گوگل با ارائه نسخه جدیدی از شبکه آغازین در سال ۲۰۱۴ طراحی و با استفاده از پایگاه داده‌ای بسیار بزرگ (برای طبقه‌بندی ۱۰۰۰ شیء مختلف) تست گردیده است. معرفی ماژول آغازین در این ساختار حجم سنگین محاسبات را کاهش داده و مسئله بیش‌برازش را در ساختار GoogLeNet رفع نموده است. به عبارت دیگر، در این ساختار، عمق شبکه افزایش یافته اما

² Concatenation layer

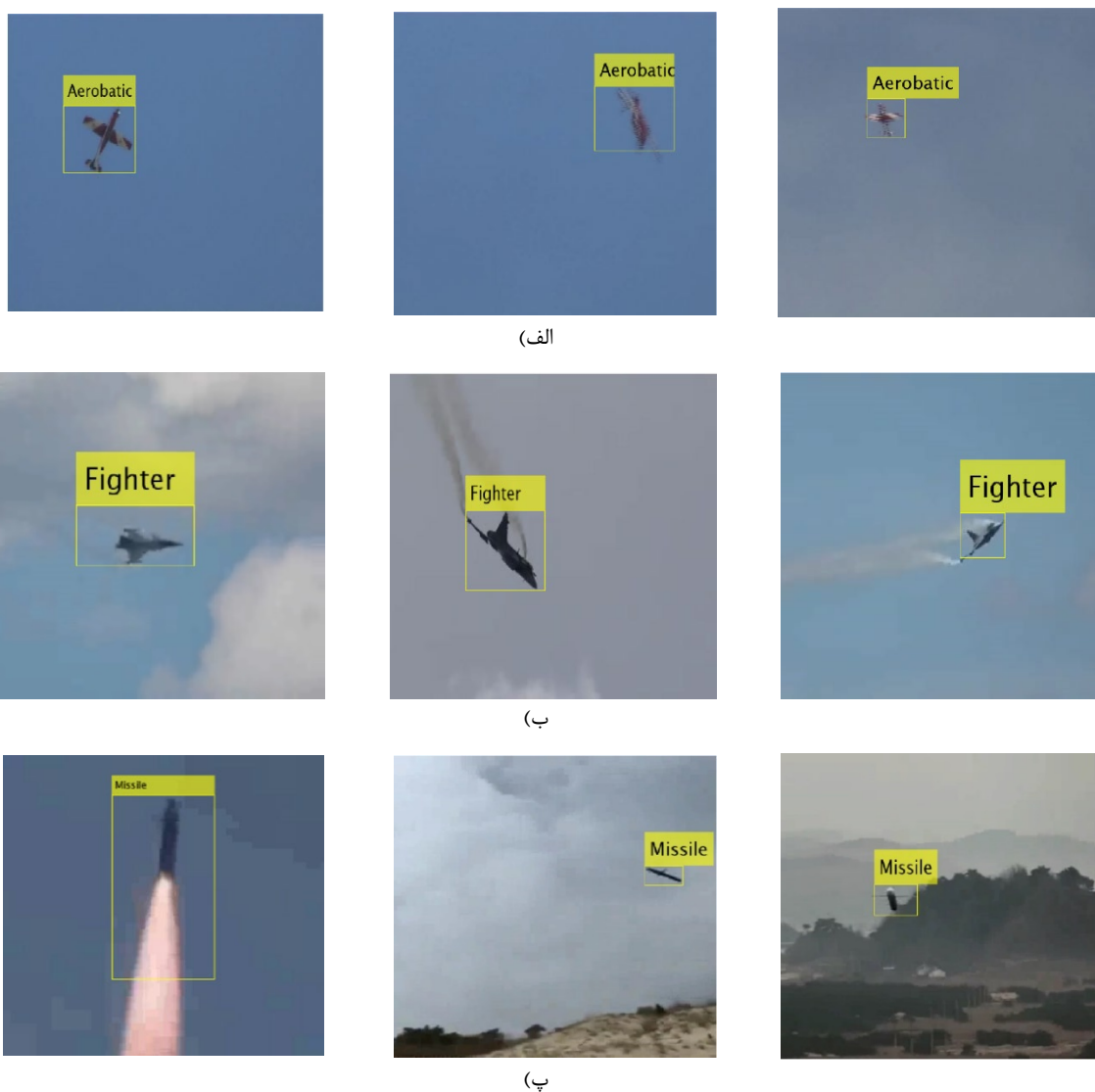
¹ Inception

لیست شده در جدول (۳) (داده‌های تست) در شکل (۱۰) برای قاب‌هایی بافاصله زمانی ۱۵ ثانیه از یکدیگر نشان داده شده است.

در قاب‌های تست، ابتدا نوع هر هدف توسط CNN پیشنهادی تشخیص داده شده و سپس موقعیت آن با استفاده از RCNN آموزش دیده ردیابی می‌گردد. نتایج برای هدف‌های مختلف و



شکل (۹): فرآیند آموزش RCNN





(ت)

$$t = i$$

$$t = i + 15$$

$$t = i + 30$$

شکل (۱۰): نتایج تشخیص و ردیابی برای قاب‌هایی با فاصله زمانی ۱۵ ثانیه از یکدیگر در داده‌های تست توسط مدل‌های CNN و RCNN توسعه یافته

ردیابی می‌باشند. سرعت دیافراگم سنسورهای بصری موجود از سرعت مطلوبی برخوردار بوده و قابلیت تشخیص و ردیابی اهداف پرنده با سرعت بالا را دارا می‌باشند. الگوریتم تشخیص و ردیابی توسعه‌یافته بر روی سیستمی با مشخصات Intel Core i7 با قدرت پردازش ۲/۹ GH و با حافظه اصلی ۱۶ GB پیاده‌سازی گردید. میانگین زمان تشخیص و ردیابی هدف در هر قاب به ترتیب حدود ۰/۵۳ و ۴/۸۵ ثانیه می‌باشد. زمان ردیابی هدف وابسته به تعداد RP های استخراج شده از تصویر بوده که این پارامتر نیز خود تابعی از پیچیدگی پس‌زمینه تصویر است. زمان محاسبه تشخیص نوع هدف مطلوب، اما ردیابی آن برای سیستم‌های برخط کمی زمان‌بر بوده، با این‌وجود از روش‌های مبتنی بر استخراج ویژگی‌های کلاسیک سریع‌تر می‌باشد. این مشکل را می‌توان با استفاده از روش‌های جدید ردیابی اشیاء و همچنین با به‌کارگیری سیستم‌هایی با پردازنده قدرتمند که در حوزه‌های نظامی به‌وفور یافت می‌شود، حل نمود؛ بنابراین باتوجه به سخت‌افزارهای قدرتمند موجود در این حوزه، دقت و سرعت الگوریتم طراحی شده وابسته به سرعت اهداف پرنده نمی‌باشد. مهم‌ترین مزیت این روش در برابر روش‌های استخراج ویژگی کلاسیک، خودکار بودن آن و عدم نیاز به انتخاب دستی هدف در قاب ابتدایی می‌باشد.

۵- نتیجه‌گیری:

مدت زمانی نسبتاً مدیدی از معرفی شبکه‌های CNN گذشته و در این مدت پیشرفت‌های بی‌شماری با طراحی ساختارهای متنوعی از آنها و منطبق بر کاربردهای متفاوت توسط پژوهشگران حاصل گردیده است. توسعه چشمگیر این شبکه‌ها در سه حوزه طراحی شبکه‌های جدید با لایه‌های مختلف، یادگیری انتقالی و استخراج ویژگی شکل‌گرفته است. باتوجه به نیازسنجی

به‌منظور تجزیه و تحلیل کارایی الگوریتم توسعه داده شده، میزان همپوشانی (^{51}IoU) چارچوب واقعی و پیش‌بینی شده برای هر قاب طبق معادله [۴۷]:

$$IoU = \frac{\text{Area of overlap}}{\text{Area of union}} \quad (۴)$$

محاسبه می‌شود. در این معادله، میزان همپوشانی بر اساس همپوشانی چارچوب واقعی و پیش‌بینی شده بر اجتماع آنها، در داده‌های آموزش و داده‌های تست محاسبه شده و در جدول (۴) لیست شده است.

جدول (۴): میزان همپوشانی چارچوب واقعی و پیش‌بینی شده هدف

برای داده‌های آموزش و تست

نوع داده	Label	میزان همپوشانی	میانگین همپوشانی
داده‌های آموزش	Cargo	۰/۶۵	۰/۶۶
	Aerobatics	۰/۷۴	
	Fighter	۰/۷۰	
	Missile	۰/۵۸	
داده‌های تست	Cargo	۰/۶۱	۰/۶۰
	Aerobatics	۰/۶۶	
	Fighter	۰/۶۴	
	Missile	۰/۵۱	

مطابق انتظار، دقت در داده‌های تست از داده‌های آموزش کمتر اما مطلوب می‌باشد. تنها دقت ردیابی موشک از بقیه اهداف کمتر می‌باشد. علت این امر نیز خروجی موشک در حال آتش، اندازه بسیار کوچک موشک در حال حرکت و تغییر مقیاس شدید آن در حین پرواز می‌باشد. در هنگام حرکت سریع اهداف پرنده، سرعت دیافراگم دوربین (تعداد قاب بر ثانیه) و سرعت پردازش الگوریتم طراحی شده دو فاکتور مهم و تأثیرگذار در تشخیص و

⁵¹ Intersection over union

شده از دقت و سرعت مطلوبی در ردیابی اهداف هوایی برخوردار می‌باشد. برتری روش پیشنهادی ردیابی هدف ارائه شده در برابر روش‌های کلاسیک، خودکار بودن آن است. در واقع در روش‌های کلاسیک، ابتدا می‌بایست هدف به صورت دستی انتخاب و سپس ویژگی‌های آن هم نه به صورت جامع استخراج گردد. اما در این روش، چارچوب هدف به شبکه آموزش داده شده و سپس مدل آموزش یافته به صورت خودکار هدف را ردیابی می‌نماید. میانگین میزان همپوشانی چارچوب واقعی و پیش‌بینی شده هدف برای داده‌های آموزش و تست به ترتیب برابر با ۰/۶۶ و ۰/۶۰ بوده و میانگین زمان ردیابی هدف برابر ۴/۸۵ می‌باشد. حال می‌توان از این دو الگوریتم در یک بسته جامع به منظور تشخیص و ردیابی خودکار اهداف هوایی استفاده کرده و برتری هوایی را در صحنه نبرد به دست آورد.

در پژوهش‌های آتی می‌توان ساختارهای دیگری با استفاده از دیگر CNN های پیش آموزش دیده و نوین برای تشخیص و ردیابی هر چه سریع‌تر و دقیق‌تر اهداف مختلف هوایی طراحی نمود. همچنین باتوجه به کاهش عملکرد سنسورهای بصری در شب و در شرایط بد آب‌وهوایی، می‌توان شبکه‌ای ژرف با استفاده از سیستم‌های مادون قرمز طراحی و تست نمود.

صورت گرفته در صحنه نبرد، نوع اهداف هوایی می‌بایست بادقت و سرعت بالایی تشخیص داده شود. در بخش اول مقاله، شبکه CNN به گونه‌ای طراحی شده که قابلیت مطلوبی به منظور برآورده سازی نیازسنجی صورت گرفته در صحنه نبرد را دارا می‌باشد. در واقع در شبکه توسعه داده شده تعداد لایه‌ها به نحوی کنترل گردیده که فرآیند تشخیص نوع هدف با سرعت مطلوبی انجام شود. از طرف دیگر اندازه و نحوه چینش لایه‌های گنجانده شده در CNN توسعه داده شده (پیش‌بینی، ادغام و ...) به نحوی انجام شده تا شبکه از دقت بالایی در تشخیص نوع هدف برخوردار بوده و قابلیت کاربرد در سامانه‌های مختلف نظامی را دارا باشد. دقت کلی طبقه‌بندی برای داده‌های آموزش و تست برای چهار نوع هدف هوایی به ترتیب برابر ۰/۹۹/۸ و ۰/۹۶/۳ بوده و میانگین زمان تشخیص برابر ۰/۵۳ ثانیه می‌باشد. در بخش بعدی، الگوریتمی با استفاده از یادگیری انتقالی و شبکه RCNN به منظور ردیابی اهداف هوایی در قاب‌های متوالی توسعه داده شده است. یادگیری انتقالی توسط شبکه پیش آموزش دیده GoogLeNet و بر اساس پایگاه داده تصویری ارائه شده در این مقاله انجام شده است. در این حالت شبکه GoogLeNet با پایگاه داده تطبیق داده شده و از آن برای استخراج ویژگی‌های نواری پیشنهادی استفاده می‌شود. باتوجه به عمق شبکه GoogLeNet تطبیق داده شده، الگوریتم RCNN توسعه داده

۶- مراجع

- [9] M. Anvaripour, R. S. Kandovan, and S. Soltanpour, "Low distance airplanes detection and tracking visually using spectral residual and KLT composition," *Research Institute for ICT, ACECR, Tehran, Iran*, 2014.
- [10] A. S. Mian, "Realtime visual tracking of aircrafts," in *2008 Digital Image Computing: Techniques and Applications*, 2008, pp. 351-356: IEEE.
- [11] S. Yeom, "Long Distance Ground Target Tracking with Aerial Image-to-Position Conversion and Improved Track Association," *Drones*, vol. 6, no. 3, p. 55, 2022.
- [12] S. Yeom, "Moving people tracking and false track removing with infrared thermal imaging by a multirotor," *Drones*, vol. 5, no. 3, p. 65, 2021.
- [13] S. Yeom, "Long distance moving vehicle tracking with a multirotor based on IMM-directional track association," *Applied Sciences*, vol. 11, no. 23, p. 11234, 2021.
- [14] S. R. Vantaram and E. Saber, "Survey of contemporary trends in color image segmentation," *Journal of Electronic Imaging*, vol. 21, no. 4, pp. 040901-040901, 2012.
- [15] A. G. Oskouei, M. Hashemzadeh, B. Asheghi, and M. A. Balafar, "CGFFCM: Cluster-weight and Group-local Feature-weight learning in Fuzzy C-Means clustering algorithm for color image segmentation," *Applied Soft Computing*, vol. 113, p. 108005, 2021.
- [16] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, 2006, vol. 2, pp. 2169-2178: IEEE.
- [17] F. Hu, G.-S. Xia, Z. Wang, X. Huang, L. Zhang, and H. Sun, "Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification," *IEEE Journal of Selected Topics in Applied*
- [1] L.-Y. Lo, C. H. Yiu, Y. Tang, A.-S. Yang, B. Li, and C.-Y. Wen, "Dynamic Object Tracking on Autonomous UAV System for Surveillance Applications," *Sensors*, vol. 21, no. 23, p. 7888, 2021.
- [2] Y. Wu, Y. Sui, and G. Wang, "Vision-based real-time aerial object localization and tracking for UAV sensing system," *IEEE Access*, vol. 5, pp. 23969-23978, 2017.
- [3] C. Fu, A. Carrio, M. A. Olivares-Mendez, R. Suarez-Fernandez, and P. Campoy, "Robust real-time vision-based aircraft tracking from unmanned aerial vehicles," in *2014 IEEE international conference on robotics and automation (ICRA)*, 2014, pp. 5441-5446: IEEE.
- [4] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE international conference on image processing (ICIP)*, 2016, pp. 3464-3468: IEEE.
- [5] V. AZADZADEH, "A New Framework for Tracking Aerial Moving Targets using AFFINE-SIFT Algorithm," *Journal Of Electronical & Cyber Defence*. 2017, Vol. 5, No.1, pp. 47-59. (in persian).
- [6] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang, "Fast visual tracking via dense spatio-temporal context learning," in *Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 2014, pp. 127-141: Springer.
- [7] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm computing surveys (CSUR)*, vol. 38, no. 4, pp. 13-es, 2006.
- [8] E. N. Barmounakis, E. I. Vlahogianni, and J. C. Golias, "Unmanned Aerial Aircraft Systems for transportation engineering: Current practice and future challenges," *International Journal of Transportation Science and Technology*, vol. 5, no. 3, pp. 111-122, 2016.

- algorithm," *J. Geomat. Sci. Technol*, vol. 36, pp. 614-619, 2019.
- [34] 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, pp. 4210-4214: IEEE.
- [35] 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.
- [36] H. Zhang, G. Wang, Z. Lei, and J.-N. Hwang, "Eye in the sky: Drone-based object tracking and 3d localization," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 899-907.
- [37] S. Zhang, L. Zhuo, H. Zhang, and J. Li, "Object tracking in unmanned aerial vehicle videos via multifeature discrimination and instance-aware attention network," *Remote Sensing*, vol. 12, no. 16, p. 2646, 2020.
- [38] A. Kouris, C. Kyrkou, and C.-S. Bouganis, "Informed region selection for efficient uav-based object detectors: Altitude-aware vehicle detection with cycar dataset," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 51-58: IEEE.
- [39] [H. Mohseni, M. Najafzadeh, M. Zarie, A. Jahed Saravania, and S. Zare, "Calculation of the Combined Threat of Air Targets Using Neuro-Fuzzy Systems," \(in Persian\), 2022, vol. 9, no. 2, pp. 69-78](#)
- [40] J. Mazloun and H. Bigdeli, "An Optimized Compound Deep Neural Network Integrating With Feature Selection for Intrusion Detection System in Cyber Attacks Electronic and Cyber Defense," vol. 10, no. 4, pp. 41-51, 2023. (in persian).<https://dor.isc.ac/dor/20.1001.1.23224347.1401.10.4.5.5>
- [41] M. Zarie, A. Jahedsaravani, and M. Massinaei, "Flotation froth image classification using convolutional neural networks," *Minerals Engineering*, vol. 155, p. 106443, 2020.
- [42] A. Jahedsaravani, M. Massinaei, and M. Zarie, "Prediction of Froth Flotation Performance Using Convolutional Neural Networks," *Mining, Metallurgy & Exploration*, pp. 1-15, 2023.
- [43] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.
- [44] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 2014, pp. 391-405: Springer.
- [45] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211-252, 2015.
- [46] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
- [47] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 658-666.
- Earth Observations and Remote Sensing*, vol. 8, no. 5, 2015.
- [18] L. Liu and Z. Shi, "Airplane detection based on rotation invariant and sparse coding in remote sensing images," *Optik*, vol. 125, no. 18, pp. 5327-5333, 2014.
- [19] X. Yao, J. Han, L. Guo, S. Bu, and Z. Liu, "A coarse-to-fine model for airport detection from remote sensing images using target-oriented visual saliency and CRF," *Neurocomputing*, vol. 164, pp. 162-172, 2015.
- [20] B. Zhao, Y. Zhong, G.-S. Xia, and L. Zhang, "Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 2108-2123, 2015.
- [21] H. Shuo, W. Na, and S. Huajun, "Object tracking method based on SURF," *AASRI Procedia*, vol. 3, pp. 351-356, 2012.
- [22] L. Jianfang, Z. Hao, and G. Jingli, "A novel fast target tracking method for UAV aerial image," *Open Physics*, vol. 15, no. 1, pp. 420-426, 2017.
- [23] S. Omachi and M. Omachi, "Fast template matching with polynomials," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2139-2149, 2007.
- [24] X. Zhang, H. Liu, and X. Li, "Target tracking for mobile robot platforms via object matching and background anti-matching," *Robotics and Autonomous Systems*, vol. 58, no. 11, pp. 1197-1206, 2010.
- [25] H. Zhou, M. Taj, and A. Cavallaro, "Target detection and tracking with heterogeneous sensors," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 4, pp. 503-513, 2008.
- [26] P. Fang, J. Lu, Y. Tian, and Z. Miao, "An improved object tracking method in UAV videos," *Procedia Engineering*, vol. 15, pp. 634-638, 2011.
- [27] J. Sun, "A fast MEANSHIFT algorithm-based target tracking system," *Sensors*, vol. 12, no. 6, pp. 8218-8235, 2012.
- [28] X. Liu and Z. Zhang, "A vision-based target detection, tracking, and positioning algorithm for unmanned aerial vehicle," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1-12, 2021.
- [29] S. V. Alavi Panah Zo, N. Zarei, and M. Molazadeh Golmahaleh, "discrimination of sea chaff in search mode based on deep learning algorithm," in *Radar*, 2023, vol. 10, no. 1, pp. 75-85. (in persian)<https://dor.isc.ac/dor/20.1001.1.23454024.1401.10.1.10.9>
- [30] Q. V. Le, "Building high-level features using large scale unsupervised learning," in *2013 IEEE international conference on acoustics, speech and signal processing*, 2013, pp. 8595-8598: IEEE.
- [31] Q. Wu *et al.*, "Improved mask R-CNN for aircraft detection in remote sensing images," *Sensors*, vol. 21, no. 8, p. 2618, 2021.
- [32] P. Zhao, H. Gao, Y. Zhang, H. Li, and R. Yang, "An Aircraft Detection Method Based on Improved Mask R-CNN in Remotely Sensed Imagery," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 1370-1373: IEEE.
- [33] M. Yuan, T. Jiang, and X. Wang, "Aircraft target detection in remote sensing image based on improved YOLOv3

Recognition and Tracking of Aerial Targets Using Convolutional Neural Network

A. Jahed Saravani* 

*Assistant Professor, Khatam Al-Anbiya Air Defense University, Tehran, Iran

(Received: 2022/09/08, Revised: 2022/12/17, Accepted: 2023/01/01, Published: 2023/01/21)

DOR: <https://dor.isc.ac/dor/20.1001.1.23454024.1401.10.2.2.3>

Abstract

Automatic recognition and tracking systems of aerial targets are of particular importance in the battle field. These types of systems use visual sensors, have the ability to be installed on various military systems, and are suitable for discovering and tracking low-altitude targets. In this manuscript, a convolutional neural network was designed to recognize the type of aerial targets (cargo, aerobatics, fighter and missile) and then target tracking using a pre-trained network (GoogLeNet) and transfer learning in the form of a region with convolutional neural network was done. The recognition accuracy of aerial targets in the test data set is 96.3%. On the other hand, the overlap value between the actual and predicted bounding box of target in the test data set for cargo and aerobatics plane, fighter and missile is 0.61, 0.66, 0.64 and 0.51, respectively, which shows the desirable accuracy of the developed model for targets tracking in consecutive frames.

Keywords: : Aerial targets, Target recognition, Target tracking, CNN, RCNN.

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Publisher: Imam Hussein University

Authors



*Corresponding Author Email: A_jahed@khadu.ac.ir