

توسعه یک الگوریتم تشخیص اشیاء متحرک مبتنی بر یادگیری عمیق در استفاده زمان واقعی روی ربات پرنده چهارملخه

دانش آموخته کارشناسی ارشد، گروه مهندسی مکاترونیک، دانشگاه تبریز، تبریز، ایران، khanyabrahim@gmail.com

دانشیار، گروه مهندسی مکاترونیک، دانشگاه تبریز، تبریز، ایران، smrs.noorani@tabrizu.ac.ir

ابراهیم خانی

* سید محمد رضا سید نورانی

چکیده

ربات‌های پرنده چهارملخه ابزارهایی قادرمند و فناورانه برای عملیات‌های پروازی هستند و مطالعات زیادی در راستای کنترل و هوشمندسازی آنها انجام گرفته است. اما بیشتر این مطالعات بر کنترل حرکت آنها متمرکر بوده و کمتر به شناسایی اشیای متحرک توسط آنها در یک محیط پویا مانند خیابان توجه شده است. در این مقاله به توسعه‌ی روشهای ترکیبی برپایه‌ی الگوریتم ویولا-جونز و شبکه‌های عصبی عمیق می‌پردازیم تا بتوان از یک ربات پرنده چهارملخه بهمنوعی هوشمند برای شناسایی اشیای متحرک در محیط واقعی استفاده کرد. در این روش آموزش شبکه‌های عصبی عمیق با استفاده از آبشارهای طبقه‌بند ویولا-جونز، بهصورت افروزن اشیاء مانند دنده‌های ویدئویی، و با پس‌زمینه‌ی واقعی انجام می‌شود. با پیاده‌سازی تجربی الگوریتم بهینه‌سازی شده آششاری ویولا-جونز روی ربات چهارملخه نشان می‌دهیم که این الگوریتم در تشخیص فریم‌های حاوی شیء متحرک که از دوربین نصب شده روی ربات دریافت می‌شد، از نرخ عملکرد بالا، هشدار کاذب (منفی خطأ)، مثبت خطأ) کم و نرخ خطای پایین برخوردار است. عملکرد الگوریتم در این آزمون ۸۹٪ تشخیص صحیح مثبت و ۱۳٪ درصد خطأ در تشخیص بود.

همچنین دقیق‌ترین دقت الگوریتم زمانی که از فیلتر بلوری گاؤسی استفاده می‌شود افزایش می‌یابد.

واژه‌های کلیدی: چهارملخه، تشخیص اشیای متحرک، ویولا-جونز، پردازش تصویر، شبکه عصبی تلفیقی، یادگیری عمیق.

Development of an algorithm for detecting moving objects based on deep learning in real-time use on a quadcopter flying robot

E. Khani

M. R. S. Noorani

Department of Mechatronics Eng., University of Tabriz, Tabriz, Iran

Department of Mechatronics Eng., University of Tabriz, Tabriz, Iran

Abstract

Quadcopter flying robots are powerful and technological devices for flight operations, and many studies have been carried out to control and make them intelligent. But most of these studies have focused on their motion control and less attention has been paid to their application for moving object detection in a dynamic environment such as the street. In this article, we develop a hybrid method based on Viola-Jones algorithm and deep neural networks so that can be intelligently used by quadcopter to identify moving objects in the real environment. In this method, deep neural network training is done using Viola-Jones classifier cascades, by adding objects as video sequences, with a real-world background. By experimentally implementing this Viola-Jones cascade optimized algorithm on a quadcopter robot, we show that in detecting frames containing moving objects received from the camera installed on the robot, it has high performance rate, low false alarm (negative-error, positive-error) and low error rate. Algorithm performance in this test was 89% positive correct diagnosis and 13% error in diagnosis. Also, the accuracy increases when the Gaussian blur filter is used.

Keywords: quadcopter, moving object detection, Viola-Jones, image processing, convolutional neural network, deep learning.

باشد، می‌تواند به ابزاری هوشمند و قادرمند در راستای اهداف بزرگی چون مسائل امنیت شهری تبدیل شود [۵]. شناسایی اشیای متحرک نه تنها در پیداکردن افرادی که در محیطی بهخصوص در حال حرکت هستند و با در یافتن حیوانات و سایر ارکان طبیعی کمک می‌کند، بلکه در پیگیری و تعقیب خودروها مفید است و می‌تواند در توسعه‌ی خودروهای هوشمند و نظارت خودکار بر رانندگی آن‌ها نیز مؤثر باشد [۶]. پیشرفت‌های اخیر در راهکارهای خودمختاری وسایل نقلیه هوایی بی‌سنسنیشن (به پادها) پذیرش آنها را در کاربردهای حساس زمانی مانند نظارت در بالای نیز افزایش داده است [۷]. در کاربردهای حساس زمانی، پخش بلادرنگ دوربین برای اپراتورهای پهپاد به منظور درک شرایط و مانور پرنده در حین مأموریت حیاتی است [۸]. ربات‌های پرنده‌ی مجهز به تشخیص شیء هوشمند، برای تصمیم‌گیری در زمان واقعی، واپسگی خلبان به تفسیر عکس فریم‌های جریانی را برای کنترل پرنده از بین می‌برند [۹]. با این وجود، شرایط عملیاتی زمان واقعی ربات‌های پرنده

-۱ مقدمه

امروزه استفاده از وسائط نقلیه هوایی برنامه‌ریزی شده بی‌سنسنیشن، با هدف آگاهی از شرایط محیطی، نقش مهمی در فرایند تصمیم‌گیری در واکنش به شرایط اضطراری و حفاظت مدنی یافته‌اند [۱] به دلیل هزینه‌ی نگهداری کم، توانایی تحرک و مانور پذیری بالا، وسائط هوایی قابلیت زیادی در امکان‌پذیر کردن کاربردهای جدید اینترنت اشیا دارند [۲]. از این رو در بسیاری از مسینه‌ها و برنامه‌های سایبری مانند سنجش از دور، نظارت بر زمان واقعی، مدیریت بلایا، برنامه‌های نظامی، کشاورزی نوین، نظارت بر مرز و جمعیت، و تحويل کلاه بسیار کاربردی بوده‌اند [۳]. با تعبیه‌ی یک دوربین می‌توان از ابزارهای هوشمندسازی ربات‌های پرنده و از جمله چهارملخه‌ها را برپایه‌ی بینایی ماشین گسترش داد [۴]. روش است که اگر یک ربات پرنده حین گشت‌زنی بتواند از محیط تصویربرداری نیز بکند و توانایی تحلیل تصاویر را داشته

* نویسنده مکاتبه کننده، آدرس پست الکترونیکی: smrs.noorani@tabrizu.ac.ir

تاریخ دریافت: ۲۰/۰۸/۰۰

تاریخ پذیرش: ۰۳/۰۶/۰۵

تشخیص شیء پهپادی از روش‌های یادگیری ماشین استفاده کردند. آنها طبقه‌بندی کننده‌ی K-نزدیکترین همسایه (KNN) را برای این کار استفاده نمودند [۱۸]. هریس و همکاران (۲۰۲۲) به ردبایی یک جسم متوجه توسعه کوادکوپتر با شکل ثابت براساس روش گرادیان نزولی و فیلتر حداکثر میانگین ارتفاع همبستگی (MACH) پرداختند. در این رویکرد شیء هدف براساس وجود یک اوج همبستگی و معیار تشابه میانگین تشخیص داده می‌شود هرچند دقیق‌تر بالایی را در تشخیص اشیاء متوجه نشان نمی‌دهد [۱۹]. ناگراجو و همکاران (۲۰۲۲) به تشخیص دقیق شیء متوجه پیش‌زمینه را براساس تقویم‌بندی و طبقه‌بندی بهینه پرداختند؛ رویی که در آن از ویژگی‌های استخراج شده با استفاده از شبکه‌ی نقشه‌ی خودسازمانده (SOM) بدست می‌آمد و پارامترهای وزن در شبکه نیز به کمک الگوریتم بهینه‌سازی کلونی زنبورهای مصنوعی (ABC) بهینه‌سازی می‌شد [۲۰]. چارو و همکاران (۲۰۲۲) با یک روش مبتنی بر شبکه‌های عصبی تلفیقی (CNN) که با منابع داده‌ای کارآمد آموزش دیده بود، به تشخیص وسیله نقلیه متوجه پرداختند. اگرچه این روش‌ها دقیق‌تر بالایی دارند، اما از نظر محاسباتی نیز حجم بالایی دارند. در مقابل، رویکردهای مبتنی بر تفیریق پس‌زمینه (BS) سبک وزن هستند، هرچند اطلاعات کافی را برای کارهایی مانند نظارت بر رفتار رانندگی و تشخیص نقض قوانین راهنمایی و رانندگی ارائه نمی‌دهند [۲۱]. دونگ و همکاران (۲۰۲۲) یک مدل شبکه تشخیص وسایل نقلیه سبک وزن بر اساس YOLOv5 را برای دادن که کوادکوپترها را در امر نظارت و پیگیری خودروهای در حال حرکت که اشیائی متوجه محسوب می‌شوند توانند می‌سازد [۲۲]. زانگ و همکاران (۲۰۲۲) یک نسخه‌ی بهبودیافته YOLOv5 را برای تشخیص اجسام متوجه در زیر آب با کمک بهینه‌سازی ماشین استفاده کردند. این مطالعه یک مدل تشخیص هدف تصویری سونار را به جلو را بر اساس شبکه YOLOv5 با استفاده از روش یادگیری انتقال پیشنهاد می‌داد [۲۳]. وانگ و همکاران (۲۰۲۲) از YOLOv3-Pruning برای ارائه الگوریتم تشخیص بلادرنگ شیء مبتنی بر یادگیری انتقالی استفاده کردند تا حرکت در اجسام و اجسام متوجه را تشخیص دهند. در این راهکار ترکیبی از هرس برای کاهش مقیاس مدل استفاده شد و سپس از یادگیری انتقالی برای اطمینان از دقیق‌تر شدن مدل بهره گرفته شده است [۲۴]. دومنیگو و همکاران (۲۰۲۲) برای بهبود تشخیص فعالیت انسانی در ویدئوهای زنده، از ادغام شبکه LSTM با منابع داده‌های مختلف شامل ویژگی‌ها، روش‌های تشخیص شیء بهبودیافته پیشنهاد شده است که ترکیبی از یک آشکارساز شیء معمولی و ردبایی اسکلت بهره گرفته است [۲۵]. یانگ و همکاران (۲۰۲۲) بهمنظور بهبود سیستم تشخیص خودرو بر پایه‌ی بهینه‌سازی ماشین از پیوست‌های LSTM دو لایه برای تشخیص دقیق و ردبایی مداوم اجسام متوجه بهره برداشتند. در این مطالعه، یک سیستم تشخیص شیء بهبودیافته پیشنهاد شده است که ترکیبی از یک آشکارساز شیء معمولی و پیوست‌های مبتنی بر شبکه‌ی حافظه کوتاه‌مدت ماندگار (LSTM) است. تا عملکرد تشخیص را برای رانندگی هوشمند بهبود بخشند [۲۶]. یتوسینکا و همکاران (۲۰۲۲) تشخیص و انتخاب اجسام متوجه در تصاویر ویدئویی بر اساس شبکه‌های عصبی بازگشتی را ارائه دادند که در آن روشنی برای تشخیص خودکار و انتخاب اجسام متوجه توسعه داده شده بود که بر اساس شبیه‌سازی شبکه‌های عصبی ضربه‌ای

مملو از عدم قطعیت‌ها است و این ناشی از عوامل خارجی کنترل نشده، مانند شرایط محیطی ناشناخته، شرایط جوی و باد، موانع دینامیکی، ارتفاع حرکتی ناپهینه و محدودیت مشاهده‌ی جزئی است [۹]. به علاوه، عدم قطعیت در عملیات پهپادهای حساس زمانی به عوامل خارجی محدود نمی‌شود [۱۰]. در حیطه‌ی بهینه‌سازی ماشین از جمله عوامل داخلی که باعث ایجاد عدم قطعیت می‌شوند عبارتند از: فریم‌های پر نویز در حین پخش، وضوح تصویر پایین، تنظیمات ناپهینه‌ی دوربین و خروجی‌های تشخیصی نادرست از آشکارسازهای شی مبتنی بر بهینه‌سازی [۱۱]. به همین دلیل توسعه و بهینه‌سازی الگوریتم‌های هوشمند تشخیص اشیاء (متوجه) از تصاویر هوایی موضوعی باز در حوزه‌ی ربات‌های پرنده است، و از جمله نتایج این تکامل می‌توان به بهره‌گیری از این روش‌ها در ناوبری خود این ربات‌های پرنده اشاره نمود.

از این رو در این مطالعه به توسعه‌ی یک الگوریتم ترکیبی بهینه در تشخیص اشیاء متوجه برای استفاده در یک ربات پرنده چهارملخه می‌پردازیم. طراحی و ساخت این چهارملخه بر اساس مدل‌های مرسمون انجام گرفته است و از دستاوردهای این کار پژوهشی محسب می‌شود که برای پیاده‌سازی تجربی الگوریتم ارائه شده از آن استفاده می‌شود. هدف اصلی ما ارائه یک الگوریتم تشخیص شیء متوجه هوشمند مبتنی بر شبکه‌های عصبی عمیق است که می‌تواند با نرخ تشخیص بالا (یعنی تعداد فریم بالا در ثانیه) در شرایط زمان واقعی عمل کند. سرانجام از این الگوریتم می‌توان در شناسایی و تعقیب اشیاء متوجه همچون افراد یا خودروهای تحت تعقیب استفاده نمود.

ادامه‌ی این مقاله به صورت زیر سازماندهی می‌شود. در بخش بعد پیشینه‌ی پژوهش‌های انجام‌شده در توسعه‌ی روش‌ها تشخیص اشیاء ارائه خواهد شد. در بخش ۳ ربات چهارملخه ساخته شده معروفی و مأموریت آن تشریح می‌شود. در بخش ۴ مبانی الگوریتم ویولا-جوزن در تشخیص اشیاء متوجه را بیان می‌کنیم. در بخش ۵ توسعه‌ی الگوریتم ترکیبی بر پایه‌ی شبکه عصبی عمیق تلفیقی را تشریح می‌کنیم. در بخش ۶ نتایج حاصل از پیاده‌سازی‌های تجربی تحت شرایط مختلف ارائه و مورد بحث قرار خواهد گرفت. در نهایت نتیجه‌گیری در بخش ۷ ارائه می‌گردد.

۲- پیشینه‌ی پژوهش

مطالعات متعددی وجود دارد که به بررسی و یا مقایسه‌ی الگوریتم‌های تشخیص شیء توسط یک ربات پرنده تصویربردار در محیط، به ویژه مقایسه‌ی الگوریتم‌های یادگیری ماشین و یادگیری عمیق متتمرکز شده‌اند [۱۶-۱۲].

ژانو و همکاران (۲۰۲۳) با استفاده از رویکرد YOLOv7-sea به تشخیص اشیاء در تصاویر دریابی گرفته شده UAV بر پایه‌ی YOLOv7 پرداختند [۱۷]. گالو و همکاران (۲۰۲۳) از تشخیص شیء عمیق تحت YOLOv7 برای شناسایی علف‌های هرز در محصولات کشاورزی استفاده کردند. عملکرد YOLOv7 بر روی مجموعه داده‌های واقعی که از تصاویر پهپاد گرفته می‌شد اجرا شد. این امر در سرعت‌های بالای حرکت پهپاد تقریباً غیرممکن بهنظر می‌رسد [۱۰]. وانگ و همکاران (۲۰۲۳) به تشخیص شیء پهپادی تعمیم‌یافته از طریق تفکیک دامنه فرکانس پرداختند [۱۱]. هرچند رویکرد آنها دقیق‌تر بالایی داشت اما در پردازش‌های بلادرنگ دچار خطای شد. زابوب و همکاران (۲۰۲۳) در

این اشیا که در پژوهش حاضر بیشتر خودروهای ساکن و در حال حرکت در خیابان‌ها و بزرگراه‌ها هستند با دقت خاصی تشخیص داده می‌شوند که نتایج کار در ادامه مورد بحث و ارزیابی قرار می‌گیرد. با توجه به مأموریت مذکور، پرواز خیلی بلند و دور ضرورتی ندارد و تنها پرواز در فاصله ۸ الی ۱۰ متری از سطح زمین با سرعت متوسطی حدود ۲۰ متر در ثانیه کفایت می‌کند. در این ربات چهارملخه پرنده از چهار موتور EMAX استفاده شده است. بنابراین، سرعت پیش‌رانی ربات در دامنه‌ی صفر تا شش کیلومتر بر ساعت خواهد بود و دامنه‌ی دید دوربین تصویربرداری نصب شده بر روی ربات ۱ تا ۱۰۰ متر می‌باشد. بسته به فاصله‌ی تصویربرداری و نیز سرعت حرکت ربات چهارملخه، رزو لوشن ووضوح دوربین تغییر می‌کند و در برخی موارد ممکن است کم می‌شود. این دوربین قادر است تصویربرداری ۱۵ فریم در ثانیه را دارد که بهترین عملکرد آن در یک بازه فاصله از زمین و دامنه تصویربرداری مشخص خواهد بود. بنابراین این تنظیمات باید به گونه‌ای صورت گیرد که بهترین کیفیت تصویربرداری حاصل شود.



شکل ۱- ربات پرنده چهارملخه ساخته شده

۴- الگوریتم تشخیص شیء آبشاری تقویت شده از ویژگی‌های ساده

در اینجا از مدل آشیاری تقویت شده متنی بر ویژگی‌های ساده‌ی ویولا-جونز (2001) برای تشخیص شی سریع، استفاده می‌شود [۲۵]. تشخیص شی در راه‌کار مبتكراهه ویولا-جونز سه گام دارد. گام نخست، انتگرال‌گیری از تصویر ورودی است؛ که محاسبه‌ی ویژگی‌ها توسط توصیف‌گر را سرعت می‌بخشد. پس از آن از ویژگی‌های استخراج شده آموزش صورت می‌گیرد. این آموزش بر اساس روش آدابویست که روشی برای طبقه‌بندی در یادگیری ماشین محسوب می‌شود صورت می‌گیرد و طبقه‌بندی ویژگی‌ها را تنها با تکیه بر تعداد کمی از ویژگی‌های اصلی از یک گروه ویژگی بزرگ‌تر انجام می‌دهد. در نهایت، طبقه‌بندی‌های پیچیده‌تری از طبقه‌بندی‌های اولیه ایجاد می‌گردد که تصمیم‌سازی نهایی از وجود شیء هدف در تصویر ورودی را انجام می‌دهد. این یک راهکار بهمنه‌ی پردازش تصویر است که آن نواحی از تصویر که شیء هدف در آنجا وجود دارد تقویت می‌شود و آن نواحی که احتمال کمتری از حضور شیء هدف در آنجا می‌رود از کل تصویر پاک می‌شود. بنابراین با تفیری و حذف پس‌زمینه سرعت شناسایی شیء افزایش می‌یابد. در این مقاله، مطابق شکل ۲، از چهار نوع ویژگی مستطیلی موسوم به ویژگی‌های شبه هار بر روی تصویر استفاده می‌شود. ویژگی‌های شبه هار عبارتند از نواحی روشن و تیره مستطیل شکلی که بر روی زیر پنجه در ۲۴ پیکسل مورد بررسی (نماید احتمالی در برگیرنده

تشخیص و جداسازی اجسام متحرک را انجام می‌داد [۲۷]. پوشا و همکاران (۲۰۲۰) از استخراج ویژگی‌های هیستوگرام گرادیان‌گرا (HOG) و طبقه‌بند KNN برای شناسایی وسایل نقلیه در بزرگراه‌ها استفاده کردند. اندازه جعبه مرزی مورد استفاده در این تحقیق ثابت بود [۲۸]. ساندینو و همکاران (۲۰۲۰) برای ناویری یک پهباخ خودگردان و درک فعال آن از هدف در یک محیط درهم‌ریخته و دارای عدم قطعیت، یک چارچوب فرمول‌بندی شده در قالب فرآیند تصمیم‌گیری مارکف قابل مشاهده جزئی (POMDP) را برای یک برنامه امداد و نجات (SAR) با شیوه‌سازی در نرم‌افزار GAZEBO توصیف کردند که در آن تشخیص شیء هدف بر پایه‌ی بینایی ماشین صورت می‌گرفت [۲۹].

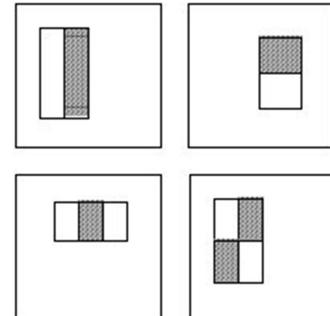
اگرچه این مرور مقالات تمام پیشینه‌ی غنی پژوهش‌های گذشته را در برنمی‌گیرد، لیکن توجهات پژوهشگران در دو سال اخیر را منعکس می‌کند. این راهکارها عمدهاً بر استفاده از الگوریتم‌های هوش مصنوعی و بهره‌گیری از دوربین و بینایی ماشین در «تشخیص اشیاء متحرک» به جای روش‌های قدیمی‌تر مانند استفاده از حسگرهای رادیویی، صوتی و لیزیزی تمرکز دارند. از سویی، استفاده از ربات‌های سیار مانند چهارملخه‌ها که امکان جابجایی در مکان‌های طبیعی مختلف بدون نیاز به حضور انسان را دارند اهمیت زیادی پیدا کرده است. این بدین جهت است که در بسیاری از سوانح و محیط‌های طبیعی انسان حضور انسان به طور مستقیم وجود ندارد و یا با خطرات جدی همراه است. همچنین توانایی انسان محدود است و همین محدودیت موجب تصمیم‌گیری‌ها و انتخاب راهکارهای نامناسب در کنترل آسیب‌ها می‌گردد. همان‌طور که پیش از این ذکر شد، تجهیزشدن به امکاناتی همچون ربات‌های پرنده (مثلًا کوادکوپتر) می‌تواند تا حدود زیادی از این محدودیت‌ها بکاهد.

۳- ساخت و مأموریت ربات چهارملخه

اجزای اصلی استفاده شده در ساخت ربات پرنده چهارملخه شامل شاسی منعطف، ۴ موتور DC بدون جاروبک اندازه متوسط با ملخ، «کنترل پرواز» APM 2.8 Mega با نام تجاری آردوکوپتر برای ناویری و حفظ تعادل، مجذبه GPS UBLOX M8N، به همراه مازول منبع تقدیمی، «کنترل گر سرعت» Hobbywing X-rotor ۳۰ آمپر برای کنترل دور موتورها و نهایتاً تنظیم نیروی پیشرانی، مازول فرستنده صدا و تصویر ۸ TS5823 کاتالوگ دارای باند ۵/۸ GHz به منظور ارسال تصاویر به زمین همین مأموریت، گیرنده بی‌سیم تصویر، و نهایتاً دوربین می‌باشد. ابعاد ربات چهارملخه مورد استفاده در پژوهش حاضر ۳۲×۳۲ سانتیمتر و بدون گارد محافظ است. تصویر این ربات در شکل ۱ نشان داده شده است.

مأموریت پروازی این ربات چهارملخه، گشتزنی در یک محیط باز و فیلم‌برداری از آن جا به منظور کشف و دریابی اشیاء متحرک احتمالی تشخیص داده شده در توالی‌های ویدئویی است. این ربات باید قادر باشد تا از یک ارتفاع مشخص و با یک سرعت محدوده‌بندی شده تصاویر ویدئویی باکیفیت مناسب برای تحلیل در پردازشگر اصلی ارسال کند. بنابراین، مراحل زیر برای پیاده‌سازی این هدف اجرا می‌شود. اجمالاً می‌توان گفت، پیاده‌سازی الگوریتم تشخیص شیء متحرک بر اساس تصاویر دریافتی از ربات چهارملخه یک طبقه‌بند باینری را ایجاد می‌کند که توانایی تفکیک اشیاء متحرک را از اشیاء ساکن مجزا خواهد داشت.

شیء هدف که در تشخیص چهره همان صورت انسان است) تعریف شده‌اند. ارزش هر ویژگی مشخص شده برابر با مجموع تمام مقادیر پیکسل‌های داخل مستطیل‌های روشن است که از مجموع تمام پیکسل‌های داخل مستطیل‌های تیره کاسته می‌شوند. سایر نقاط عکس که خارج از این مستطیل‌ها قرار دارند در محاسبه ویژگی در نظر گرفته نمی‌شوند. به این ترتیب برای هر پنجره ۲۴×۲۴ در پیکسل متاخر با یک ویژگی یک عدد به عنوان ارزش این ویژگی محاسبه می‌شود. این ویژگی‌ها برای شناسایی انسان و اشیاء متخرک در تصاویر بسیار موثر عمل می‌کنند.



شکل ۲- ویژگی‌های ساده‌ی ویولا-جونز یا ویژگی‌های شبه هار که به نسبت پنجره‌ی تشخیص محوطه نشان داده شده است

برای محاسبه تصویر انتگرالی از معادله (۱) استفاده می‌شود؛ در واقع در یک تصویر انتگرالی مقدار هر پیکسل برابر با مجموع تمام پیکسل‌های بالا و سمت راست این پیکسل روی تصویر اصلی به علاوه مقدار خود پیکسل روی تصویر اصلی است:

$$(1) \quad ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

که در اینجا (x, y) نماینده‌ی مکان پیکسل‌ها می‌باشد. همچین با کمک دو رابطه‌ی بازگشتی زیر می‌توان تصویر انتگرالی را تنها با یک بار عبور از روی تصویر اصلی ساخت:

$$(2) \quad s(x, y) = s(x, y - 1) + i(x, y)$$

$$(3) \quad ii(x, y) = ii(x - 1, y) + s(x, y)$$

که $s(x, y)$ جمع سطحی پیکسل‌ها، و $ii(-1, y) = 0$ و $s(x, -1) = 0$ هستند.

چالش اصلی در روش ویولا-جونز یافتن تعداد محدودی ویژگی میان تعداد بسیار زیاد آنها است طوری که بتوان با ترکیب خطی آنها یک طبقه‌بند قوی ایجاد کرد. به عبارت دیگر در این الگوریتم طبقه‌بند «قوی» به عنوان ترکیبی خطی از طبقه‌بندی‌های ساده‌ی «وزن داده شده‌ی ضعیف» ایجاد می‌شود، یعنی داریم:

$$(4) \quad h(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

که اینجا هر $h_t(x)$ یک طبقه‌بند ضعیف است که شامل یک ویژگی j_t ، یک آستانه θ_j و یک معین جهت نابرابری p_j است:

$$(5) \quad h_t(x) = \begin{cases} 1 & \text{if } p_j f_j(x) < p_j \theta_j \\ 0 & \text{otherwise} \end{cases}$$

در معادله بالا، x زیرپنجره‌ی 24×24 بعدی از تصویر ورودی است.

برای هر ویژگی تصویر زیر یک طبقه‌بند h آموزش داده می‌شود که از یک ویژگی واحد استفاده می‌کند. خطای هر طبقه‌بند ضعیف به صورت زیر

$$\epsilon_j = \sum_i w_i [h_j(x_i) - y_i] \quad (6)$$

به ازای تمامی ویژگی‌های استخراج شده، الگوریتم ضعیفتر عملکرد طبقه‌بند نهایی را تعیین می‌کند، به طوری که تعداد نمونه‌هایی که نادرست طبقه‌بندی شوند به حداقل برسد. بنابراین، طبقه‌بند h_j که کمترین خطای ϵ_j را بدهد انتخاب می‌شود و آنگاه وزن‌ها به صورت زیر به روزرسانی می‌شوند:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-\epsilon_i} \quad (7)$$

که در اینجا اگر طبقه‌بندی داده‌ی x_i به درستی انجام شده باشد e_i صفر است و عملکرد طبقه‌بند مناسب است و اگر مقدار e_i یک باشد خلاف این است. همچنین مقدار β_t از رابطه‌ی زیر محاسبه می‌شود:

$$\beta_t = \frac{\epsilon_t}{1 - \epsilon_t} \quad (8)$$

و ضرائب وزن α_t در ترکیب خطی معادله (۴) عبارت خواهند شد از:

$$\alpha_t = \log \frac{1}{\beta_t} \quad (9)$$

۵- توسعه‌ی الگوریتم تشخیص شیء ویولا-جونز

برپایه‌ی شبکه‌ی عصبی تلفیقی عمیق

تشخیص اجسام متخرک کوچک از فاصله دور هنگام فیلم‌برداری توسط واسطه نقلیه پرنده بدون سرشیون همچون رات چهارملخه پرنده موضوعی است که در این پژوهش مورد توجه واقع می‌شود. فقدان شکل، متمايز و جزئیات بافتی در اجسام کوچک (بهدلیل فاصله فیلم‌برداری)، استفاده مستقیم از روش‌های تشخیص شیء برپایه‌ی شبکه‌های عصبی تلفیقی (CNN) عملکرد بسیار پایین‌تری را نشان می‌دهد. از این رو در پژوهش حاضر روشی برای تشخیص اجسام متخرک کوچک در فیلم‌های ضبط شده با استفاده از یک پرنده‌ی مجھزه دوربین ویدئویی پیشنهاد می‌شود. مراحل اصلی روش پیشنهادی تثبیت ویدئو، تخمین و تفیریق پس‌زمینه، تقسیم‌بندی فریم با استفاده از CNN و آستانه‌گذاری فریم قطعه‌بندی شده است.

البته در آموزش یک CNN لازم است که یک مجموعه داده بزرگ همراه با برچسب در دسترس باشد. برچسب‌زنن دستی اجسام متخرک کوچک در ویدئوها بسیار دشوار و زمان بر است و به علاوه چنین مجموعه داده‌های برچسب‌گذاری شده‌ای در حال حاضر وجود ندارد. برای رفع این مشکل، در این پژوهش آموزش یک CNN با استفاده از راهکار آشماره‌ای طبقه‌بند ویولا-جونز را پیشنهاد می‌کنیم که در آن افزودن اشیاء مانند دباله‌های ویدئویی با پس‌زمینه واقعی است. بنابراین ما برای آموزش آشکارساز، داده‌های آموزشی مصنوعی را با پس‌زمینه‌های ویدئویی دنیای واقعی که با استفاده از دوربین پهلو گرفته شده‌اند تولید می‌کنیم. این داده‌ها اشیای حباب‌مانندی هستند که به طور مصنوعی تولید می‌شوند. اجسام متخرک مصنوعی تولید شده تحت نظرارت یک متخصص، از اجسام متخرک منتخب در فیلم‌های پهلو و با ردیابی دستی آنها، و با افزودن تغییرات تصادفی کوچک استخراج می‌شوند. اشیاء به دست آمده به نام «اجسام متخرک مصنوعی» به گونه‌ای ساخته شدند که از ظاهر و الگوهای حرکت اجسام متخرک در محل مورد مطالعه را تقلید کنند. لازم به ذکر است که این روش برای تولید داده‌های آموزشی مصنوعی را می‌توان به راحتی برای هر نوع شیء هدف و با پس‌زمینه‌های متفاوت در دنیای واقعی تطبیق داد.

محاسبه می‌شود.

اگر موقعیت جدید شیء متحرک خارج از ناحیه قابل مشاهده صحنه باشد، آن را به کادر اضافه نمی‌کنیم. برای شبیه‌سازی پرواز و بازگشت اشیاء متحرک به ناحیه قابل مشاهده صحنه، موقعیت اشیاء را پیچیری می‌کنیم اما آنها را به کادر اضافه نمی‌کنیم. هنگامی که اشیاء متحرک مصنوعی را به یک قاب اضافه می‌کنیم، از مقادیر پیکسل آنها به عنوان کانال آلفا برای ترکیب بین پس زمینه و پیکسل‌های سیاه استفاده می‌کنیم. بنابراین، اشیاء متحرک مصنوعی که میانگین بافت کمتری دارند سبک‌تر به نظر می‌رسند، یعنی نسبت کنتراست پایینی نسبت به پس زمینه دارند و بالعکس.

ما شبکه عصبی تلفیقی خود را با دنباله‌هایی از ۵ فریم متولی 256×256 پیکسلی که به طور تصادفی از دنباله‌های آموزشی مصنوعی برش داده شده‌اند آموزش می‌دهیم. شبکه به از دنباله‌ها آموزش داده می‌شود تا قاب واقعی زمینی مربوط به فریم میانی دنباله را پیش‌بینی کند. از آن جا که تعداد اشیاء متحرک در یک توالی منفرد نسبتاً کم است، اکثر قاب‌های بریده شده حاوی تعداد کمی شیء متحرک هستند یا اصلاً شیء متحرکی ندارند. بنابراین، گنجاندن دنباله‌های برش داده شده در مجموعه آموزشی با توزیع یکنواخت منجر به عدم تعادل تعداد اشیاء متحرک موجود در هر دنباله می‌شود. با در نظر گرفتن مشکلات آموزش شبکه با مجموعه آموزشی نامتعادل، تصمیم گرفتیم دنباله‌های برش خود را با یک توزیع احتمالی متناسب با تعداد اشیاء متحرک موجود در ناحیه قاب، وارد مجموعه آموزشی کنیم. به این ترتیب، دنباله‌هایی با تعداد زیادی شیء متحرک، اکرچه کمتر نمونه‌برداری شده‌اند، ولی به بیشتر در مجموعه آموزشی گنجانده می‌شوند. در مقابل، توالی‌هایی با تعداد کم یا بدون شیء متحرک، کمتر در مجموعه آموزشی قرار داده می‌شوند. با نمونه‌برداری از توالی‌هایی که به این روش برش می‌خورند، یک مجموعه آموزشی با 53760 ، یک مجموعه اعتبارسنجی با 12800 و یک مجموعه آزمایشی با 12800 نمونه بدست می‌وریم. در نهایت، برای ارزیابی عملکرد الگوریتم تشخیص اشیاء متحرک در تصویر، نتیجه تشخیص شبکه آموزش دیده را با موقعیت‌های واقعی شیء متحرک مقایسه می‌کنیم. اگر فاصله بین موقعیت شیءی که متحرک تشخیص داده می‌شود با موقعیت صحیح برچسب‌گذاری شده‌ی آن کمتر از 10 پیکسل باشد، در نظر گرفته می‌شود که شیء متحرک به درستی تشخیص داده شده است. معیار 10 پیکسل را بر اساس اندازه متوسط یک شیء متحرک در تصویر و برای معرفی درجه‌ای از تحمل به برچسب‌گذاری‌های نادقيق انسانی انتخاب کردیم. عملکرد آشکارساز بر حسب فراخوانی آن بیان می‌شود.

۶- نتایج و بحث

دقت الگوریتم‌های پیاده‌سازی شده در دو حالت آفلاین و آنلاین بررسی خواهیم نمود. در حالت آفلاین پس از اتمام تصویربرداری ربات از محیط، فیلم‌های ضبط و ذخیره می‌شوند و پس از رسیدن 5-i -CORE زمین در پردازشگر اصلی که در اینجا یک لپ‌تاپ با پردازنده 5-i -CORE است، پردازش‌های تصویری جهت تشخیص اشیاء متحرک در فیلم‌ها انجام می‌گیرد. پردازش‌های آفلاین بهمنظور بررسی عملکرد کلی الگوریتم توسعه‌داده شده در حالتی که در آن ربات چهارملخه و

برای آزمایش میدانی روش پیشنهادی، از تصاویر ربات چهارملخه ساخته شده خود که در ارتفاع بین 7 تا 15 متری پرواز داده شد استفاده می‌کنیم. محدودیت این سیستم نیاز آن به یک ایستگاه پایه در که همان کامپیوتر شخصی ما با ریزپردازه 5-i -CORE بود می‌باشد. این ربات چهار ملخه پرنده مجهز به مژول تصویربردار است که امکان ضبط ویدئوهای نافرشده $4K$ را فراهم می‌کند. به علاوه به طور جدآنگاه از یک دوربین برای فیلم‌برداری مکان‌های مورد مطالعه استفاده شد تا تفاوت عملکرد هر دو دوربین در خبط فیلم مشاهده شود و نیز تأثیر سوگیری تجهیزات را از بین ببرد و این‌گونه شرایط متفاوت برای تکرارپذیری آزمایش فراهم شده باشد. تمامی ویدئوهای ضبط شده دارای وضوح $4k$ با نرخ فریم 25 فریم بر ثانیه هستند. شرایط نوری مختلف با استفاده از تنظیمات خودکار دوربین حذف شد تا مشکلی در پردازش خودکار فیلم‌های ضبط شده ایجاد نکند.

از میان ضطهای موجود با انتخاب بخش‌هایی که در آنجا شیء متحرکی مشاهده نمی‌شود نمونه‌هایی از تصویر پس زمینه به دست آورده‌یم. با این حال، از آن جایی که تشخیص اجسام متحرک کوچک در فیلم‌ها دشوار است، ممکن بود برخی از اجسام متحرک در چندین فریم از تکه فیلم‌های انتخاب شده باقی بمانند. با این وجود، ما انتظار نداریم که تعداد کمی از اجسام باقیمانده بر عملکرد آشکارساز تأثیر منفی بگذارد، زیرا CNN‌ها می‌توانند مقدار مشخصی از نویز برچسب‌زنی را تحمل کنند.

برای حذف تأثیر حرکت کلی دوربین، یعنی ثبت توالی ویدئو، تبدیل افین را بین هر فریم در دنباله و فریم اول قرار می‌دهیم و سپس همه فریم‌ها را در یک قاب مرجع مشترک تاب می‌دهیم. برای تخمین تبدیل وابسته بین دو فریم، ابتدا نقاط کلیدی را در هر دو فریم شناسایی کرده و توصیفگرهای آنها را با استفاده از آشکارساز و توصیفگر ORB محسوبه می‌کنیم. سپس، با مقایسه توصیفگرهای با استفاده از فاصله همینگ، جفت‌های منطبق را از نقاط کلیدی پیدا می‌کنیم. نهایتاً منطبق‌ها برای تخمین تبدیل قوی با استفاده از الگوریتم RANSAC استفاده می‌شوند. فریم‌های ویدئوهای تشییت شده را در بلوک‌های 1024×1024 پیکسل هماندازه می‌کنیم و از هر طرف تصویر 200 پیکسل را حذف می‌کنیم تا اثرات حاشیه‌ای ناشی از تشییت از بین برود. پس از این روش، 96 دنباله ویدئوی پس زمینه با فریم‌هایی به اندازه 1024×1024 پیکسل باقی می‌ماند و هر دنباله 3 ثانیه طول دارد.

با بررسی ظاهر و رفتار اشیاء متحرک دنیای واقعی در ویدئوهایی که دنباله‌های پس زمینه از آنها استخراج شده بودند، تصمیم گرفتیم اجسام متحرک مصنوعی را به عنوان حباب‌های بیضوی مدل‌سازی شده با استفاده از مستهنهای گاؤسی دو بعدی با انحراف از معیارهای تصادفی از فواصل زمانی به ترتیب برای محورهای x و y نشان دهیم.

در هر فریم سرعت جدید v و جهت جدید θ هر جسم متحرک از فریم قبلی به صورت محاسبه می‌شود:

$$v_t = v_{t-1} + \Delta v \quad (10)$$

$$\theta_t = \theta_{t-1} + \Delta \theta \quad (11)$$

که در آن Δv و $\Delta \theta$ از توزیع‌های طبیعی با میانگین صفر و انحراف از معیار به ترتیب 2 و 30 نمونه‌برداری می‌شوند و v_{t-1} و θ_{t-1} سرعت و جهت شیء متحرک در قاب قبلی هستند. سپس موقعیت‌های جدید اشیاء متحرک با استفاده از پیش‌بینی سرعت آنها در دو راستای x و y



شکل ۴- تصاویری از یک فیلم کیفیت بالای برداشته شده از وبگاه و عملکرد آفلاین الگوریتم پیشنهادی

۶-۱- پرسی تأثیر فیلتر بلوری گاؤسی بر عملکرد

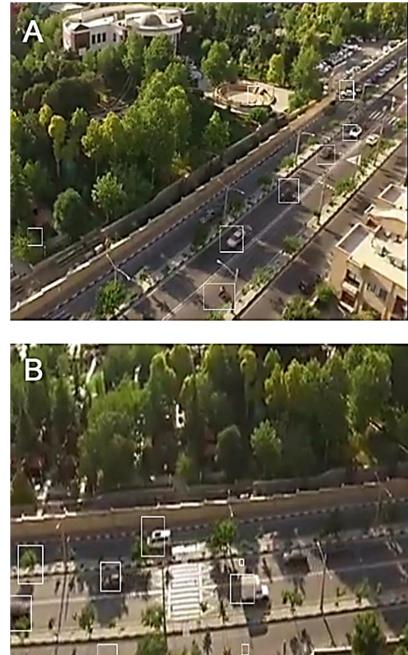
مانند هر داده‌ی حسگری دیگری، تصاویر دوربین نیز می‌تواند انواع نویز را پیدا نماید. روش‌های صاف کردن تصویر به کاهش نویز کمک می‌کند. جهت ارزیابی عملکرد الگوریتم پردازش گر و تشخیص دهنده‌ی اشیاء متاخر ابتدا از یک فیلتر بلور (که به آن تاری نیز گفته می‌شود) از کتابخانه OpenCV که بر بستر زبان برنامه‌نویسی پایتون نوشته شده است، برای صاف کردن تصاویر در ویدئوهای دریافتی استفاده می‌کنیم. در این پژوهش، از فیلتر گاؤسی استفاده خواهیم کرد. فیلترهای گاؤسی دارای ویژگی عدم خروج بیش از حد نسبت به ورودی تابع مرحله‌ای و در عین حال زمان برخاست و نشست حداقل هستند. از نظر پردازشی، فیلتر گاؤسی هر لبه‌ی تیز در تصویر را صاف می‌کند و در عین حال تاری بیش از حد را به حداقل می‌رساند. پارامترهای تنظیمی در این فیلتر شامل اندازه‌ی کرنل گاؤسی σ انحراف استانداردهای کرنل در امتداد محورهای X و Y و نوع حاشیه‌ی که معین مزه‌های تصویر در زمانی که هسته روی حاشیه‌های تصویر می‌افتد است، می‌باشد.

در این پژوهش، عملکرد الگوریتم تشخیص اشیاء متاخر در حالت بدون فیلتر، و سه حالت فیلترشده با طول کرنل‌ها صفر، ۱۰ و ۲۰ بررسی می‌شود. هدف این بررسی تأثیر نقش تغییرات فیلتر بلور گاؤسی بر دقت نهایی الگوریتم است. به این منظور ۱۰ فیلم ضبط شده‌ی ریات و ۱۰ فیلم بارگیری شده از وبگاه‌های آزاد مورد ارزیابی قرار گرفتند. طول هر یک از این فیلم‌ها ۱ الی ۳ دقیقه است که در شرایط مختلف گرفته شده‌اند. سپس هر یک از حالت‌های فیلتری تنظیم و روی تمامی فیلم‌های آزمون اعمال می‌شود که به عنوان یک پیش‌پردازش قبل از اجرای الگوریتم تشخیص اشیاء متاخر محسوب می‌گردد. برای ارزیابی هرچه بهتر الگوریتم فیلم‌های مورد آزمون در شرایط مختلف مانند شدت نور متفاوت، سرعت و ارتفاع پرواز متفاوت، و با وضوح متفاوت گفته شده است.

نتایج تأثیر فیلتر بلوری گاؤسی بر الگوریتم تشخیص پیشنهادی ما در جدول ۱ ارائه شده است. برای این کار در یک فایل ویدئویی، منتخب

پردازشگر در گیر انتقال داده نیستد صورت می‌گیرد. در این بخش از فیلم گرفته شده توسط دوربین ربات چهارمخلخه (هر کدام ۲ تا ۳ دقیقه)، برای این بررسی استفاده می‌شود. این فیلم‌ها تصاویری از خیابان‌های شهری هستند که اشیاء متحرك در آنها عمدتاً وسایل نقلیه می‌باشند. علاوه بر فیلم‌های گرفته شده توسط ربات از ۱۰ قطعه فیلم دیگر که با کیفیت بالاتر در ویگاه‌های پژوهشی قرار داده شده بودند نیز برای بررسی عملکرد الگوریتم‌ها در حالت آفلاین بهره گرفتیم. برای نمونه در شکل‌های ۳ و ۴ تصاویری از یک فیلم گرفته شده‌ی ربات از خیابانی در شهر تبریز و نیز از یک فیلم برداشت‌شده از ویگاه نشان داده شده‌اند.

تمامی ویدئوهای دریافتی پس از انتقال به پردازنده مرکزی، بدون هیچ‌گونه ویرایش و تغییر ابعاد یا کنتراست مورد آزمون قرار گرفتند تا ابتداً عملکرد اصلی الگوریتم مورد بررسی قرار گرفته و ایرادهای آن یافته و رفع شود. در این مرحله هدف، کاهش حداکثری ایرادات در تنظیم پارامترهای شبکه عصی مدل شده است؛ به نحوی که ویدئو خروجی به خوبی بتواند اشیاء متحرك را با نشانه‌گذاری قاب سفیدرنگ در توالی‌های ویدئویی مشخص و دنبال کند. با نگاه اجمالی به شکل‌های ۳ و ۴ می‌توان مشاهده کرد که عملکرد الگوریتم ترکیبی پیشنهادی ما بر مبنای آموزش شبکه‌ی تلفیقی به طرز چشم‌گیری مطلوب است.



شکل ۳- تصاویری از یک فیلم ضبط شده توسط ربات پرنده چهارم لخه از خیابانی در شهر تبریز و عملکرد آفلاین الگوریتم پیشنهادی

یکی از فیلم‌های آزمون که بهطور تصادفی انتخاب شد و مدت زمان آن ۱/۵ دقیقه بود استفاده کردیم تا اثر تغییرات شدت نور را بررسی کنیم. برای اعمال تغییر شدت نور از برنامه فوتوشاپ بهره گرفتیم و بهطور مصنوعی فریم‌ها را در چهار حالت ۱۰۰٪ درصد نور محیط، ۷۵٪، ۵۰٪ و ۲۵٪ ذخیره کردیم. در ادامه هر یک از این چهار فیلم مورد ارزیابی مجدد الگوریتم قرار گرفت. با آگاهی از تعداد اشیاء متحرک در این فیلم و شمارش تعداد تشخیص‌های الگوریتم شامل تعداد تشخیص‌های

مثبت صحیح و مثبت ناصحیح نتایج مطابق جدول ۲ به دست آمد.

مطابق این نتایج، الگوریتم ابشاری ویولا-جونز که مبتنی بر تفريق پس زمینه کار می‌کند و با یک شبکه عصبی عمیق تلفیقی بهبود یافته است، در مقابل تغییرات نور محیط بهشت تأثیرپذیر است. عملکرد این الگوریتم تشخیص که در پژوهش حاضر به کار گرفته شد در تنظیمات نور ۱۰۰ درصدی که کمی بیشتر از نور اولیه و عادی محیط است، در یک فیلم ویدئویی که در طول روز ضبط شده بود، تاحدودی پایین است؛ زیرا تابش پیش از حد نور باعث محوشدن لبه‌های مشخص کننده اشیاء در تصویر می‌شود و دقت الگوریتم تشخیص را کاهش می‌دهد. اگرچه دقت تشخیص صحیح در این حالت ۸۲٪ است ولی بالاترین مقدار دقت ۸۸٪ است و زمانی که شدت نور ۷۵٪ تنظیم شده است (کمی پایین‌تر از حالت اولیه و عادی فیلم) حاصل می‌شود. در شدت‌های نوری ۵٪ و ۲۵٪ نیز تصویر به حدی تاریک می‌شود که دقت تشخیص افت قابل توجه می‌کند.

باید اشاره کنیم که در عمل زمانی که ربات چهارملخه بطور آنلاین و بلادرنگ به تشخیص اشیاء در توالی ویدئوی خودش می‌پردازد، تنظیم شدت نور محیط یک امر خارج از اراده است. تنظیم تغییرات شدت تهیه در حالت آفلاین امکان‌پذیر است تا مشکلات مربوط به زمان بر بودن پردازش‌ها خلی در امر همزمان فیلمبرداری و تشخیص وارد نکند. با این حال می‌توان پیشنهاد کرد که برای کاستن تأثیر نامطلوب شدت نور بهتر است از یک دوربین باکیفیت و پوشخ خوب استفاده شود تا تغییرات شدت نور و همچنین لرزش ربات پرندۀ در پردازش فیلم تأثیر چشم‌گیر نداشته باشد.

جدول ۲- ارزیابی الگوریتم پیشنهادی در تشخیص شیء متحرک در ویدئوی آفلاین بر حسب شدت نور داده شده به تصویر

شدت نور (%)	تعداد اشیاء (عدد)	تعداد اشیاء (٪)					
۲۵	۹	۸۳	۲۹	۳۵	۱/۵	۱۰۰	
۲۰	۷	۸۹	۳۱	۳۵	۱/۵	۷۵	
۳۲	۱۴	۶۹	۲۴	۳۵	۱/۵	۵۰	
۳۱	۱۱	۶۳	۲۲	۳۵	۱/۵	۲۵	

۳-۶- تفريقي پس زمینه

هدف اصلی اين مطالعه، دستیابی به بهترین پارامترهای الگوریتم تفريقي پس زمینه، برای بهترین تأثیرگذاری در تشخیص اشیاء متحرک، در توالی‌های ویدئویی است. اين موضوع را با تنظیمات پارامترهای کلاس تفريقي پس زمینه در الگوریتم پیشنهادی در قالب چهار مدل بررسی می‌کنیم. پارامترهای تنظیمي الگوریتم تقسيم‌بندی دودویي

بهطور تصادفی در يك بازه‌ی زمانی معين تعداد اشیاء متحرک واقعی شمرده می‌شود و سپس با اعمال يك حالت فیلتری و پس از آن اجرای الگوریتم مبتنی بر يادگیری عمیق بر روی همان ویدئوی منتخب، مجددًا تعداد تشخیص‌های صحیحی که توسط الگوریتم داده شده است، شمارش می‌شود. بدین صورت همان‌طور که در جدول ۱ آمده است، اعمال يك فیلتر بلوری گاؤسی می‌تواند اثرات مخرب نویز و تابش‌های گاه و بی‌گاه خورشید به لنز دوربین را بکاهد. اما گاهی استفاده‌ی زیادی از آن می‌تواند تأثیر معکوس در تشخیص نهایی داشته باشد. آن طور که نتایج آزمون‌ها نشان می‌دهد، الگوریتم تشخیص در فیلم‌های ویدئویی بدون استفاده از فیلتر، از ۲۳ شیء متحرک ۱۶ مورد را صحیح تشخیص داده است (دقت تشخیص صحیح مثبت ۶۹٪). با اعمال فیلتر گاؤسی با کرنل صفر به بهترین عملکرد الگوریتم دست می‌یابیم که از ۲۳ مورد شیء متحرک در همان فیلم، ۱۹ شیء صحیح تشخیص داده شده‌اند (۸۳٪). باید توجه داشت که اعمال فیلتر تا آنجا می‌تواند در بهبود دقت خروجی الگوریتم مؤثر باشد که لبه‌های تیز اشیاء را کلامو نکند. از این رو می‌بینیم که دقت الگوریتم با افزایش طول کرنل به ۱۰ و ۲۰ به ترتیب به ۴۷٪ و ۳۰٪ کاهش یافته است. این موضوع همچنین در خطای مثبت ناصحیح تشخیص (یعنی تعداد اشیاء متحرک تشخیص داده شده‌اند نبوده‌اند اما به اشتباه متحرک تشخیص داده شده‌اند) و نیز خطای منفی ناصحیح (یعنی تعداد اشیاء غیرمتحرک در تصویر که به اشتباه متحرک تشخیص داده شده‌اند) هم تأثیر منفی می‌گذارد؛ با تقویت فیلتر بلوری هرچه لبه‌های تیز اشیاء صاف‌تر شود احتمال آنکه الگوریتم به اشتباه يك جسم ساکن را متعارک تشخیص دهد بالاتر می‌رود.

شایان ذکر است در پژوهش حاضر هدف کاستن خطای مثبت ناصحیح و کم کردن تشخیص‌های اشتباه اشیاء متحرک نیست بلکه هدف افزایش تشخیص‌های صحیح است. بدین معنا که هرچه تعداد اشیاء متحرک بیشتری در يك قاب تصویر تشخیص داده شوند، الگوریتم عملکرد بهتری را نشان داده است. از این رو در تنظیم نهایی از يك فیلتر بلوری گاؤسی با پارامترهای طولی و عرضی و کرنل ۲ استفاده می‌شود، زیرا بهترین عملکرد را در مقایسه با سایر تنظیمات نشان داد.

جدول ۱- ارزیابی الگوریتم پیشنهادی در تشخیص شیء متحرک در ویدئوی آفلاین بر حسب طول کرنل فیلتر گاؤسی [gray, (31, 31), λ]

بدون فیلتر	صفرا	۱۰	۲۰
۱۷	۴	۶۹	۱۶
۳۰	۷	۸۳	۱۹
۴۰	۹	۴۷	۱۱
۴۰	۹	۳۰	۷

۴- بررسی تأثیر شدت نور بر عملکرد

یکی دیگر از مواردی که در کدنویسی الگوریتم بدان توجه شد تغییرات نور محیطی و تأثیر آن بر دقت نهایی الگوریتم بوده است. تأثیر کم و زیادی نور محیطی بر روی عملکرد تشخیص لبه‌ها سیار مهم است و نتایج این بررسی در جدول ۲ آورده شده است. برای این منظور از

پس زمینه/پیش زمینه مبتنی بر مخلوط گاؤسی شامل ماسک fgmask فریم ویدئویی بعدی و نرخ یادگیری می شود. fgmask را یک تصویر بازتری ۸ بیتی درنظر می گیریم. در مورد فریم ویدئویی بعدی قاب شیء متحرک بدون مقیاس بندی استفاده می شود و باید در بازه‌ی [۰،۰۵۵] باشد. نرخ یادگیری نیز مقداری بین ۰ و ۱ خواهد بود که نشان می دهد مدل با چه سرعتی عمل تفرقی پس زمینه را آموزش می بیند. مقدار صفر به معنی آن است که مدل پس زمینه اصلاً پرورشانی نمی شود، و مقدار ۱ به معنی اینکه مدل پس زمینه کاملاً از آخرین فریم مقداردهی اولیه شده است. مقدار پارامتر منفی نیز برای آن گرفته می شود که الگوریتم از نرخ یادگیری انتخابی خودکار استفاده کند. در این پژوهش بهمنظور بررسی عملکرد کلاس تغیریق‌ساز، چهار نرخ یادگیری مختلف را مورد ارزیابی قرار دادیم.

همان‌طور که در جدول ۳ آورده شده است، نرخ یادگیری شیکه در دقت خروجی الگوریتم تأثیرگذار است. برای نمونه، وجود نرخ یادگیری با مقدار خیلی کم مطلوب است، طوری که بهترین دقت الگوریتم زمانی حاصل می شود که نرخ یادگیری ۰/۰۱ گرفته شود که در این حالت دقت الگوریتم به ۸۶٪ می رسد. اما با افزایش بیش از این مقدار دیگر دقت افزایشی نشان نمی دهد. البته صفر گرفتن نرخ یادگیری نیز در هر شکل کاری کمترین دقت الگوریتم یعنی ۴۵٪ را همراه داشت. از سوی دیگر بزرگ گرفتن آن نیز هم موجب کاهش دقت در تشخیص صحیح و هم موجب افزایش خطای مثبت ناصحیح می شود که برای نمونه این مقدار بهزادی نرخ یادگیری ۱، بهترتبیب ۰/۸۱٪ (کمتر از ۸۶٪) و (بیشترین خطای مثبت ناصحیح) بوده است. با این توضیحات ما الگوریتم تشخیص شیء متحرک را با نرخ یادگیری ۰/۰۱ آموزش دادیم و نتیجه‌ی آن روی یک فیلم منتخب از تصاویر گرفته شده از چهارملخه و فیلم دیگری که از ویگاه اپارات گرفته شده، به ترتیب در شکل‌های ۳ و ۴ نشان داده شده است. همان‌طور که مشاهده می شود دقت الگوریتم در تشخیص بلادرنگ اشیاء متحرک (خودروها) در توالی‌های ویدئویی که با دوربین‌های با رزوشن‌های مختلف فیلمبرداری شده‌اند، به دلیل تنظیم صحیح پارامترهای مدل، خیلی خوب است.

جدول ۳- ارزیابی الگوریتم پیشنهادی در تشخیص شیء متحرک در ویدئویی آفلاین بر نرخ یادگیری

نرخ یادگیری	تعداد اشیاء مثبت	تعداد اشیاء منفی						
صفر	۱۲	۴۵	۱۷	۳۷				
۰/۰۱	۷	۸۶	۳۲	۳۷				
۰/۱	۹	۵۹	۲۲	۳۷				
۱	۱۱	۸۱	۳۰	۳۷				

۴-۶- ارزیابی الگوریتم پیشنهادی در شرایط آنلاین

در این گام به پیاده‌سازی الگوریتم در حالت آنلاین و ارزیابی عملکرد آن در تشخیص اشیاء متحرک محیط در توالی ویدئویی که در زمان واقعی توسط ربات چهارملخه ضبط می شود می پردازیم. بنابراین در ابتدا لازم است تا تصاویر گرفته شده کمی پس از ضبط و ذخیره از



شکل ۵- تصاویری از یک فیلم ضبط شده توسط ربات پرنده چهارملخه و تشخیص اشیاء متحرک در آن توسط الگوریتم پیشنهادی به طور آنلاین

جدول ۴ نتیجه‌ی ارزیابی الگوریتم را روی یک فایل ویدئویی زنده که به طور تصادفی از فیلم‌های ارسالی چهارملخه انتخاب شده است، نشان می دهد. مثل قبل، تعداد تشخیص‌های صحیح و ناصحیح مثبت (مثبت به معنی شیء متحرک می باشد) مورد نظر قرار گرفت. هدف ما بررسی دقت و سرعت پردازش الگوریتم تشخیص شیء متحرک در زمان واقعی می باشد و انتظار می رود که تشخیص شیء متحرک همچنان با دقت بالایی انجام شود. مطابق انتظار در حالت آنلاین بسته به انتخاب طول کرنل فیلتر بلوری گاؤسی، نرخ خطای الگوریتم تغییر می کند. پس از آنکه ویژگی‌های استخراجی از الگوریتم پیشنهادی ما برای تشخیص شیء متحرک در فیلم ویدئویی آنلاین بکار گرفته شد، دقت آزمون دوم که با طول کرنل ۰/۰۱ دقت، بالاترین رتبه را دریافت کرد. در مقابل افزایش طول کرنل موجب کاهش دقت شد که کمترین آن به ازای طول کرنل ۰/۰۲ با دقت ۵۱٪ حاصل شد. از این رو در آزمون‌های بعدی طول کرنل برابر با ۰/۰۱ انتخاب گردید.

باکیفیت‌تر رفع نمود. ولی در این پژوهش سختافزار موجود ما ظرفیت بیش از این نداشت. با وجود عملکرد قوی الگوریتم در تشخیص صحیح اشیاء متحرک، اما در بعضی از فیلم‌ها ضعف‌هایی نیز مشاهده شد. در شرایطی که شیء نامتحرک تشبیه زیادی به اشیاء متحرک در فیلم داشته باشد الگوریتم به اشتباه آن را متحرک تشخیص می‌داد و موجب کاهش دقت می‌شد. برای مثال گاهی برگ درختان و یا جداول کنار خیابان به عنوان شیء متحرک شناسایی شد، شکل ۷. این موضوع با فاصله‌ی دوربین از سطح زمین و همچنین نور محیط ارتباط دارد. از آن مهم‌تر، به دلیل آنکه ربات پرنده به طور مداوم در حال حرکت همراه با لرزش است می‌تواند تأثیر بسزایی در کیفیت فیلم‌برداری و تشخیص‌های ناصحیح داشته باشد. بنابراین پایدارسازی حرکت ربات چهارملخه موضوع مهمی در گرفتن دقت بالا در تشخیص اشیاء متحرک آنلاین دارد. همچنین تیره و روشن شدن گاه‌به‌گاه اشیاء ساکن در اثر سایه ناشی از سایر اجسام می‌تواند در تشخیص اشتباه تأثیر گذار باشد.



شکل ۷- در پردازش آنلاین این تصویر مواردی مانند علام کف خیابان، برگ درختان و جداول کنار خیابان به اشتباه جسم متحرک تشخیص داده شده است.

۶-۵- ارزیابی سرعت و ماندگاری تشخیص

با توجه به ماهیت فیلم‌های جمع‌آوری شده که بهشت تحت تأثیر شرایط محیط هستند، مقادیر پیکسل‌های تصویر مرتباً در حال تغییرند و ثابت‌ماندن پنجه‌ی تشخیص بر روی یک شیء متحرک (ماندگاری) خود یک چالش است. این موضوع نیز در این پژوهش مورد ارزیابی قرار گرفته است. برای این منظور فیلم‌ها را مجدداً مورد بازبینی قراردادیم تا زمان ماندگاری پنجه‌ی تشخیص در فیلم‌برداری بلادرگ از محیط را بستجیم. مطابق نتایج جدول ۵، متوسط زمان تشخیص شیء بین ۶ تا ۸ ثانیه است که تقریباً می‌توان گفت مستقل از فیلتر اعمالی است. این معیار می‌تواند در طراحی کنترل‌گرهای تشییت‌ساز ربات چهارملخه مفید واقع شود. اما مهم‌تر از سرعت تشخیص، ماندگاری تشخیص در ویدئویی دریافتی از دوربین می‌باشد.

جدول ۵- مقایسه‌ی سرعت و ماندگاری تشخیص بر حسب ثانیه

بدون فیلتر	۲۰	۱۰	۵	۱۳	۶	۹	۴	۲	۸	۱۳	۱۰	۱۶	۱۰
بدون فیلتر													
صفر													
۱۰													
۱۳													
۹													
۶													

جدول ۴- ارزیابی الگوریتم پیشنهادی در تشخیص شیء متحرک در ویدئوی آنلاین بر حسب طول کرنل فیلتر گاوی [gray, (31, 31), λ]

بدون فیلتر	۲۰	۴۷	۳۳	۷۰	۱۸	۳۸
صفر						
۱۰						
۱۱						
۴۷						
۴۲						
۶۵						
۳۱						
۲۴						
۱۱						
۲۳						
۱۸						

شکل ۶ تصویری از فیلم ضبط شده‌ی ربات چهارملخه از ترافیک خیابانی در شهر تبریز را نشان می‌دهد که قاب‌های تشخیص سفید دور اشیاء متحرک شناسایی شده رسم شده‌اند. وجود ترافیک سنگین منجر به توقف نایابی برخی از خودروها می‌شود و برخی دیگر در حال حرکت هستند. می‌توان گفت که یکی از پرچالش‌ترین وظایف تعیین شده برای الگوریتم پیشنهادی، تشخیص هرچه صحیح‌تر اشیاء متحرک در چنین وضعیت‌هایی است؛ زمانی که به دلیل بازتاب نور از سطح صیقلی اشیاء دقت تشخیص دستخوش تغییر می‌شود؛ اشیاء مرتباً در حال حرکت و توقف هستند؛ گاه سایه‌ی اشیاء نیز به عنوان دنباله‌ای از شیء متحرک در نظر گرفته می‌شود. از سویی، با طیف وسیعی از اشیاء متحرک همچون عابران پیاده، حیوانات، حرکت درختها بر اثر وزش باد، موتورسیکلت‌ها و ماشین‌های بزرگ مواجه هستیم که باید تمامی این اشیاء در صورت تحرک شناسایی شوند. اگر بتوان از تعداد تشخیص‌های اشتباه (مثبت ناصحیح) چشم‌پوشی نمود، راهکار بهینه‌سازی شده‌ی آبشاری ویلاجونز در مطالعه حاضر با ۸۹٪ تشخیص صحیح مثبت و ۱۳٪ درصد خطای تشخیص تا حد بالای قابل اعتماد است.

در برخی از موارد قاب تشخیص در کنار سایه خودرو نیز ظاهر شده است و ممکن است تصور شود که پردازش‌گر سایه‌ی خودرو را نیز یک شیء متحرک گرفته است. اما با پرش سریع قاب تشخیص از آن طی توالی ویدئویی، می‌توان این طور گفت که به دلیل آنکه الگوریتم سرعت واقعی فیلم را در هنگام پردازش کاهش می‌دهد (یعنی کاهش تعداد فریم‌های فایل ویدیویی جهت جلوگیری از هنگ‌کردن پردازنده) این انفاق گاه‌به‌گاه رخ می‌دهد و سریع اصلاح می‌شود.



شکل ۶- تشخیص اشیاء متحرک در ترافیک شهری توسط الگوریتم پیشنهادی

اگرچه این یک نقطه‌ی ضعف در پردازش‌های زمان واقعی می‌باشد ولی می‌توان آن را با به کارگیری پردازنده‌ی قوی‌تر و دوربین با تصاویر

۷- نتیجه‌گیری

- در این پژوهش به توسعه‌ی یک الگوریتم تشخیص اشیاء متحرك همراه با پیاده‌سازی آن روی یک ربات پرنده‌ی چهارملخه پرداختیم. الگوریتم ویولا-جونز بایه‌ی این کار بود که یک شبکه‌ی عصبی تلفیقی عمیق را آموزش می‌داد. این ترکیب می‌توانست تشخیص اشیاء متتحرك را تسربی و دقت آن را افزایش بخشد. به علاوه روش حاضر در مقایسه با روش‌هایی که تنها متکی بر شبکه‌ی عصبی تلفیقی هستند حساسیت به تغییرات ناگهانی در حرکت دوربین و یا شدت تابش نور را کاهش می‌داد. روش پیشنهادی در مقایسه با کارهای قبلی به طور همزمان شناسایی و طبقه‌بندی اجسام را منجر به کاهش زمان اجرای الگوریتم می‌شد. این امر امکان پیاده‌سازی آنلاین الگوریتم پیشنهادی را در زمان کوتاه‌تر و با دقت بالاتر فراهم می‌آورد. به علاوه این الگوریتم در مقایسه با شبکه‌هایی مانند LSTM سریع‌تر و به مقداردهی تصادفی اولیه‌ی وزن‌ها کمتر حساس است. این الگوریتم تحت فیلترهای مختلف در آزمون‌های آفلاین و آنلاین بررسی گردید.

۸- مراجع

- Zhang Y, Wu C, Zhang T, Liu Y, Zheng Y. Self-attention guidance and multiscale feature fusion-based UAV image object detection. *IEEE Geosci Remote Sens Lett*. 2023;20:1–5.
- Akhshatha KR, Karunakar AK, Shenoy S, Dhareshwar C V, Johnson DG. Manipal-UAV person detection dataset: A step towards benchmarking dataset and algorithms for small object detection. *ISPRS J Photogramm Remote Sens*. 2023;195:77–89.
- Zhang H, Shao F, He X, Zhang Z, Cai Y, Bi S. Research on object detection and recognition method for UAV aerial images based on improved YOLOv5. *Drones*. 2023;7(6):402.
- Li X, Diao W, Mao Y, Gao P, Mao X, Li X, et al. OGNN: Occlusion-guided multi-task network for object detection in UAV images. *ISPRS J Photogramm Remote Sens*. 2023;199:242–57.
- Lu W, Lan C, Niu C, Liu W, Lyu L, Shi Q, et al. A cnn-transformer hybrid model based on cswin transformer for uav image object detection. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2023;16:1211–31.
- Wang K, Fu X, Huang Y, Cao C, Shi G, Zha ZJ. Generalized uav object detection via frequency domain disentanglement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023. 1064–73.
- Yu M, Leung H. Small-object detection for UAV-based images. In: 2023 IEEE International Systems Conference (SysCon). IEEE; 2023. 1–6.
- Lu S, Lu H, Dong J, Wu S. Object detection for UAV aerial scenarios based on vectorized IOU. *Sensors*. 2023;23(6):3061.
- Ye T, Qin W, Zhao Z, Gao X, Deng X, Ouyang Y. Real-time object detection network in UAV-vision based on CNN and transformer. *IEEE Trans Instrum Meas*. 2023;72:1–13.
- Gallo I, Rehman AU, Debkordi RH, Landro N, La Grassa R, Boschetti M. Deep object detection of crop weeds: Performance of YOLOv7 on a real case dataset from UAV images. *Remote Sens*. 2023;15(2):539.
- Tian G, Liu J, Yang W. A dual neural network for object detection in UAV images. *Neurocomputing*. 2021;443:292–301.
- Zou Z, Chen K, Shi Z, Guo Y, Ye J. Object detection in 20 years: A survey. *Proc IEEE*. 2023;111(3):257–76.
- Kaur J, Singh W. Tools, techniques, datasets and application areas for object detection in an image: a review. *Multimed Tools Appl*. 2022;81(27):38297–351.