

بهبود ردیابی همراه با تشخیص اشیاء در ویدئو با استفاده از تلفیق فیلتر کالمن و یادگیری عمیق

سید مجتبی بنی هاشمی^۱، محمدرضا ذوقی^{۲*}

۱. گروه مخابرات، دانشکده مهندسی برق و کامپیوتر، دانشگاه کاشان، کاشان، ایران.

۲. استادیار، گروه مخابرات، دانشکده مهندسی برق و کامپیوتر، دانشگاه کاشان، کاشان، ایران.

تاریخ دریافت:

۲ آبان ماه ۱۴۰۴

تاریخ پذیرش:

۲۴ بهمن ماه ۱۴۰۴

چکیده

ردیابی همزمان چند هدف، یکی از مسئله‌های بنیادی بینایی ماشینی به شمار می‌رود و به دلیل کاربردهای عمومی و تجاری فراوان، طی سال‌های اخیر مورد توجه فزاینده قرار گرفته است. با وجود این، دستیابی به دقت بالا در این کار دشوار است؛ زیرا اشیاء ممکن است در تصویر فشرده و شبیه به یکدیگر ظاهر شوند یا بر اثر انسداد و تغییر زاویه دید، اطلاعات آن‌ها ناقص گردد. در این مقاله، سامانه‌ای بلادرنگ برای ردیابی چند شیء ارائه می‌شود که بر پایه نسخه‌ی اصلاح‌شده الگوریتم SORT استوار است. روند کار در دو بخش تعریف می‌شود. بخش نخست، شناسایی شیء است که با بهره‌گیری از خانواده YOLO انجام می‌پذیرد؛ هرگونه کمبود اطلاعات در این مرحله، در ادامه قابل جبران نخواهد بود. بخش دوم، ردیابی شیء است که شامل سه گام می‌باشد: ابتدا استخراج ویژگی با استفاده از آموزش انتقالی YOLOv8، سپس پیش‌بینی موقعیت به کمک فیلتر کالمن و در نهایت تخصیص و پیوند داده‌ها که با الگوریتم مجارستانی صورت می‌گیرد. نتایج با معیار MOTA ارزیابی شد و مقدار ۶۵٫۳ به دست آمد که ۷٫۲٪ برتری نسبت به روش مرجع نشان می‌دهد.

واژه‌های کلیدی: ردیابی چند شیء، ردیابی به وسیله تشخیص، ارتباط داده، یادگیری عمیق، پردازش تصویر.

Improving Multi-Object Tracking-by-Detection in Video via Fusion of Kalman Filter and Deep Learning

Seyed Mojtaba Bani Hashemi¹, Mohammadreza Zoghi^{2*}

1.2. Faculty of Electrical and Computer Engineering, University of Kashan, Tehran, Iran.

Abstract

Multi-object tracking is a fundamental computer-vision task that has drawn ever-increasing attention because of its scientific and commercial potential. Nevertheless, accurate object tracking remains highly challenging; these challenges include the high similarity and density of detected objects. Moreover, occlusion and viewpoint changes can occur as objects move. In this paper, a framework for real-time multi-object tracking is introduced that is based on a modified version of the SORT algorithm. Multi-object tracking is divided into two parts. In the first part, object detection is performed using the YOLO family; if information is lost at this stage, compensating for this lost information later is impossible. The second part concerns object tracking, which itself comprises three stages: first, feature extraction, for which transfer learning with the YOLOv8 family is used; second, position prediction using the Kalman filter; and third, data association and object matching, for which the Hungarian algorithm is employed. In the data-association stage, the use of deep-learning methods has recently expanded. Finally, the MOTA metric was adopted as the result, yielding 65.3, which is 7.2 % better than the reference paper.

Key words: Optimal UAV deployment, Signal-to-noise ratio plus interference, Urban environment, Transmission power efficiency, Particle swarm optimization.

۱- مقدمه

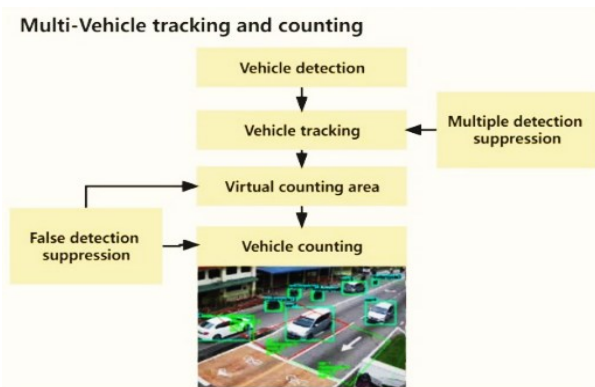
۲۰۲۳ یک الگوریتم تلفیق با ویژگی های کم عمق بر اساس SORT به نام SFFSORT⁴ پیشنهاد، و یک معماری نوآورانه برای نظارت و شمارش خودرو بر اساس ردیابی تشخیص ایجاد شد [۱۶].

در این مقاله چارچوب جدیدی برای ردیابی و شمارش معرفی شده است که بر اساس نسخه اصلاح شده الگوریتم SORT است. بر این اساس، تشخیص شیء با استفاده از سری YOLO، و ردیابی شیء نیز شامل سه مرحله استخراج ویژگی، پیش بینی موقعیت، مبتنی بر فیلتر کالمن؛ و ارتباط داده و تطبیق دادن اشیاء، با استفاده از الگوریتم مجارستانی پیشنهاد شده است. در ادامه یکسری مازول هایی برای سرکوب برخی خطاهای این مسئله ارائه گردیده است.

ساختار مقاله بدین صورت است که، ابتدا در بخش ۲ مسیر پیشنهادی در قالب پنج فاز خلاصه شده است؛ سپس بخش ۳ به چگونگی پیاده سازی آن می پردازد. ارزیابی نتایج و معیارهای سنجش در بخش ۴ تشریح شده و تحلیل داده ها در بخش ۵ ارائه می گردد. در پایان، بخش ۶ جمع بندی و نتیجه گیری را در بر می گیرد.

۲- روند الگوریتم

ردیابی براساس چارچوب پیشنهادی، نشان داده شده در شکل ۱، شامل دو بخش تقسیم می شود: تولید مدل تشخیص و ردیابی چند شیء.



شکل ۱- چارچوب ردیابی و شمارش در MOT [۱۶]

در بخش تولید مدل تشخیص، ابتدا از مجموعه بزرگ و در دسترس عموم داده های باز برای آموزش مدل اصلی، و از مدل اصلی به دست آمده، برای شناسایی نهایی استفاده می شود. در بخش ردیابی نیز برای تشخیص اشیاء از یادگیری عمیق استفاده خواهد شد. با این حال، مدل ممکن است پیش بینی های متفاوتی برای یک شیء داشته باشد. بنابراین، در نهایت یک مازول سرکوب برای بهبود دقت وجود دارد که

در دهه های اخیر، مسائل بسیاری در دنیای واقعی به واسطه الگوریتم های یادگیری عمیق در حوزه بینایی ماشین توسعه یافته، و به ثمر رسیده اند که یکی از مهم ترین کاربردهای آن، ردیابی اشیاء در ویدئو می باشد. مدل های ردیابی چندهدفه مبتنی بر شبکه های عمیق، بطور پیوسته معرفی شده و دقت را بهبود می دهند. این فناوری، چه در حوزه دفاعی و چه در بخش غیرنظامی، در سناریوهای متنوعی مانند تحلیل رفتار، پزشکی تصویری، نظارت بصری، امنیت پیرامونی، شمارش کالاهای تولیدی، آنالیز حرکات ورزشکاران، خودران سازی، سیستم های رباتیک، کشاورزی هوشمند و دیگر کاربردهای مشابه به کار گرفته می شود [۱-۴].

بمنظور ردیابی، نخست باید هدف در قاب تصویر موقعیت یابی شود؛ سپس برچسب یکتایی به آن تخصیص داده می شود. در ادامه، اشیای همانند در فریم های پی درپی، مسیریابی ایجاد می نمایند. هدف مورد نظر می تواند هر موجود یا شیء متحرکی همچون عابر پیاده و خودرو، یا بازیکن ورزشی، پرنده، خفاش، مورچه، ماهی، زنبور و اجسام دیگر باشد. ردیابی در واقع فرایندی تحلیلی است که در آن موقعیت، اندازه، جهت گیری و دیگر ویژگی های هدف در تصاویر متوالی استخراج شده و با پیش بینی رفتار حرکتی، دنبال کردن مستمر آن تضمین می گردد [۵-۸]. عواملی مانند تغییر کردن تعداد اهداف، ظاهر آنها، انسداد موقت، ناپدید شدن لحظه ای و شباهت زیاد ویژگی های اشیاء مختلف، دقت ردیابی چندهدفه را به شدت تحت تأثیر قرار می دهند.

روش های سنتی که عمدتاً در حوزه شناسایی خودرو [۹]، پیشنهاد شده اند بر هیستوگرام گرادیان جهت دار [۱۰] و ویژگی های هار [۱۱] مبتنی هستند و مشکل عمده آنها این است که اغلب به نرخ های مثبت کاذب بالا منجر می گردند.

در سال ۲۰۲۰، FairMOT به عنوان رویکردی ساده و کارآمد معرفی شد [۱۲]، و در آن برخلاف ایده ی مرسوم «ابتدا آشکارسازی، و سپس شناسایی»، با دو شاخه ی تشخیص، و استخراج ویژگی مجدد بطور یکسان و برابر رفتار می نماید. اساس این الگوریتم مشابه CenterTrack است تا تأثیر جعبه های مرزی تشخیص را بر روی مدل ReID¹ کاهش دهد. با تخمین نقطه ی مرکزی هدف در نقشه ی ویژگی با وضوح بالا، ویژگی ReID می تواند بهتر با مرکز شیء هماهنگ شود. در ادامه، مدل JDE بر اساس الگوریتم تشخیص YOLOv3، با استفاده از شبکه هرمی ویژگی^۲ برای پیش بینی اشیاء از مقیاس های چندگانه، و ترکیب ویژگی های عمیق استخراج شده از شبکه تشخیص و ویژگی های ظاهری به دست آمده از قبل، پیشنهاد گردید [۱۳]. در سال ۲۰۲۱ الگوریتمی که برگرفته از الگوریتم SORT³ است به نام SORT-YM ارائه گردید که از یک استراتژی پیش بینی حرکت برای پیش بینی مکان اجسام مسدود شده استفاده شده است [۱۵]. در سال

³ Simple, Online and Realtime Tracking

⁴ Shallow feature fusion SORT

¹ Re-identification

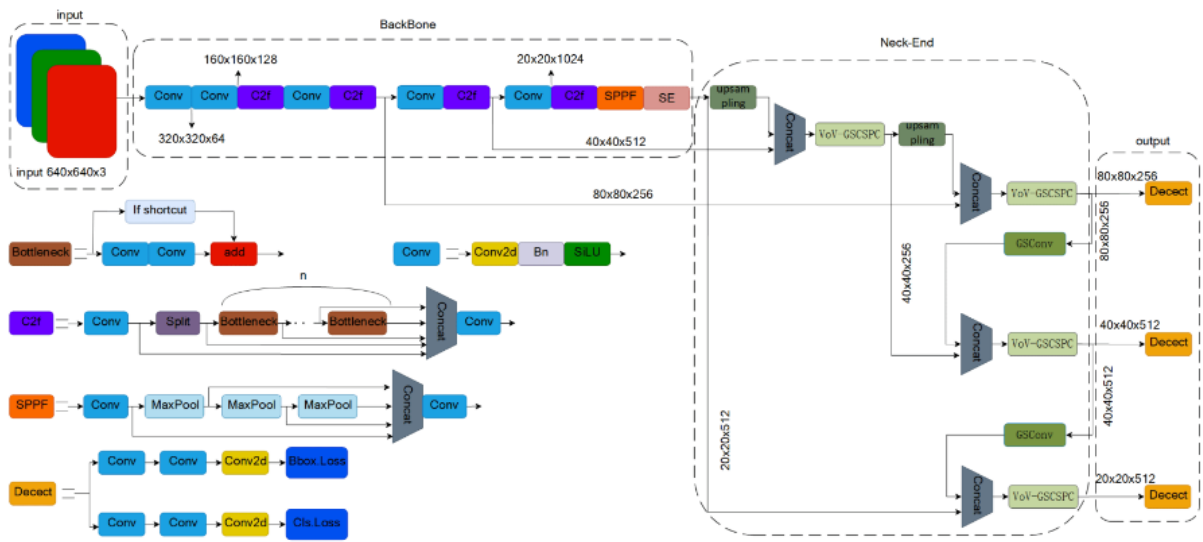
² Feature pyramid network (FPN)

در حین ردیابی خطا را سرکوب میکند. این بخش مفهوم طراحی و فرآیند اجرای هر جزء در چارچوب پیشنهادی را شرح میدهد.

الف) آشکارساز شبکه عصبی عمیق

با محبوبیت شبکه‌های عصبی کانولوشنی و توسعه سخت افزار، استفاده از آنها به رویکرد اصلی برای تشخیص اشیاء تبدیل شده است. شناسایی اشیاء^۱، به معنای تشخیص اتوماتیک اشیاء موجود در تصاویر به صورت عکس یا فیلم توسط الگوریتم‌های مختلف است. این الگوریتم‌ها در

سال‌های اخیر با توجه به افزایش سرعت و دقت، کاربردهای فراوانی در زمینه‌های مختلف مانند خودروهای هوشمند، امنیت، رباتیک، پزشکی، بازی‌های رایانه‌ای، مدیریت ایمنی کارگاه‌ها و بسیاری موارد دیگر دارند. YOLO^۲ یکی از کارآمدترین الگوریتم‌های حوزه تشخیص اشیاء است که می‌تواند اشیای موجود در تصاویر را با سرعت و دقت بالا شناسایی کند. در این مقاله از نسخه پیشرفته YOLOv8 [۱۷] استفاده خواهد شد. در شکل ۲ مدل ساده‌شده و اولیه YOLOv8 آورده شده است.



شکل ۲ - فلوجارت ساده‌شده مدل YOLOv8

استفاده از YOLOv8 به عنوان چارچوب پایه و بارگذاری فایل وزن YOLOv8s.pt از پیش آموزش دیده ساخته شده است. این مدل با استفاده از مجموعه داده MS COCO آموزش داده، و اعتبارسنجی شده است.

ج) پیش‌بینی وضعیت

در این بخش وضعیت و موقعیت جسم در فریم بعدی توسط فیلتر کالمن پیش‌بینی می‌شود تا کادر یا جعبه پیش‌بینی شده بدست آید. این الگوریتم در دو گام اجرا می‌شود. در گام پیش‌بینی، فیلتر کالمن تخمینی از وضعیت فعلی متغیرها را در شرایط عدم قطعیت ارائه می‌کند. زمانی که نتیجه اندازه‌گیری بعدی بدست آید، تخمین قبلی با میانگین وزن‌دار آپدیت می‌شود. به این ترتیب که وزن اطلاعاتی که دارای قطعیت بیشتری هستند، بیشتر خواهد بود. الگوریتم بازگشتی می‌باشد و با استفاده از ورودی‌های جدید و حالات محاسبه شده قبلی به صورت بی‌درنگ اجرا می‌شود.

د) ارتباط داده

در این بخش میزان ارتباط جعبه‌های مرزی در مرحله تشخیص و جعبه‌های مرزی در مرحله پیش‌بینی وضعیت محاسبه می‌شود. سپس

ب) استخراج ویژگی با استفاده از یادگیری انتقالی

شبکه عصبی پیشنهاد شده مبتنی بر یادگیری انتقالی مجموعه‌ای از توصیفگرها را از یک تصویر استخراج می‌کند، و توصیفگرهای استخراج‌شده نسبت به انتقال، چرخش و مقیاس-پذیری تصویر تغییرناپذیر است. یادگیری انتقالی یک تکنیک رایج برای تطبیق مدل‌های CNN از پیش آموزش دیده با وظایف MOT است. در نتیجه، به جای آموزش کل مدل از ابتدا، یادگیری انتقال بر استفاده مجدد از نمایش ویژگی‌های آموخته شده از یک مدل از پیش آموزش دیده و تنظیم دقیق آنها بر روی داده‌های خاص MOT استوار است. این به سیستم‌های MOT اجازه می‌دهد تا از قابلیت‌های تعمیم مدل‌های از پیش آموزش دیده بهره ببرند و در عین حال با ویژگی‌های بصری خاص و چالش‌های سناریوهای ردیابی سازگار شوند.

مجموعه داده MOTChallenge_Train16 شامل ۱۱۰۴۰۷ تصویر است، که بر پنج نوع شیء «انسان»، «ماشین»، «موتور سیکلت»، «کامیون» و «اتوبوس» تمرکز شده است. مدل اولیه با

² You Only Look Once

¹ Object detection

از زبان پایتون از طریق مرورگر در collab بسیار سریع است، زیرا می تواند مدل را با استفاده از GPU گران قیمت ارائه شده توسط collab آموزش دهد. پردازنده گرافیکی موجود توسط collab شامل Nvidia K80s، T4s، P4s و P100s است. پردازنده گرافیکی ارائه شده به عنوان مثال Nvidia Tesla T4 دارای ۱۶ گیگابایت حافظه و ۲۵۶۰ هسته CUDA است که می تواند برای محاسبه الگوریتم موازی پیچیده برای انواع آموزش مدل YOLO استفاده شود. با مقایسه زمان ها با استفاده از google collab و کامپیوتر شخصی که دارای GPU GTX 1660ti است، google collab می تواند سرعت ۲.۵ برابری ارائه دهد.

۴- نتایج و ارزیابی بر اساس چالش MOT

در این بخش از مجموعه داده چالش MOT برای ارزیابی استفاده خواهد شد. معیار مهم کنونی در زمینه ردیابی چند هدفه MOTChallenge است که به عنوان یک سکوی عمومی برای بارگذاری و انتشار نتایج تحقیقات در مورد روش های ردیابی چندهدفه عمل می کند و دارای بزرگترین مجموعه داده ردیابی عابر پیاده در دسترس عموم است که شامل MOT16 می شود. در این آزمایش از تشخیص عمومی YOLOv8 استفاده شد. داده های مقایسه ای از وبسایت رسمی MOTChallenge به دست آمده است. به منظور مقایسه مزایا و معایب الگوریتم های مختلف، لازم است شاخص های ارزیابی معقولی برای ارزیابی عملکرد الگوریتم ها تنظیم شود. شاخص های ارزیابی رایج در ردیابی چند شیء عبارتند از: دقت ردیابی چند شیء (MOTA) [۱۸]، امتیاز (IDF1) [۱۹]، عمدتاً ردیابی شده (MT)، اغلب گمشده (ML)، سوئیچ های شناسه (ID).

MOTA پرکاربردترین شاخص در ارزیابی عملکرد ردیابی است. این معیار به طور جامع پدیده تشخیص نادرست، شناسایی از دست رفته و سوئیچ های ID مسیر را در نظر می گیرد، پس می تواند عملکرد کلی ردیاب را منعکس کند. محاسبه به شرح زیر است:

$$MOTA = 1 - \frac{\sum (FN + FP + IDs)}{\sum GT} \quad (2)$$

که در آن:

FN: تعداد منفی های غلط^۱ در تمام ویدئو

FP: تعداد مثبت های غلط^۲ در تمام ویدئو

IDs: تعداد کل تغییرهای شناسه

GT: تعداد اشیاء واقعی در کل فریم های ویدئو

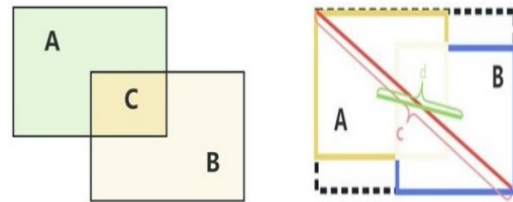
IDF1 به طور جامع میزان دقت شناسه ID و نرخ فراخوان ID را در نظر می گیرد. به دقت تخصیص شناسه ID در مسیر حساس است. فرمول محاسبه IDF1 به شرح زیر است:

$$IDF1 = \frac{2IDTP}{2IDTP + IDFP + IDFN} \quad (3)$$

² False Positive

با استفاده از معیاری به نام IOU که فرمول آن در زیر آورده شده میزان همپوشانی میان جعبه ها بدست می آید (شکل ۲). سپس با استفاده از الگوریتم مجارستانی انتساب میان اشیاء در فریم فعلی و قبلی بدست آمده و شناسه به فریم بعدی منتقل می شود.

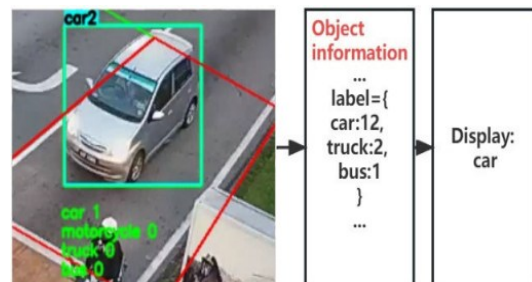
$$IOU = \frac{|A \cap B|}{|A \cup B|} = \frac{|C|}{|A \cup B|} \quad (1)$$



شکل ۳ - شکل سمت چپ نمودار شماتیک IOU را نشان می دهد، A و B دو جعبه هستند و c تقاطع آنهاست. شکل سمت راست A و B دو جعبه تشخیص هستند، c طول مورب جعبه مرزی بیرونی، و d فاصله مرکز است.

۵) ماژول سرکوب

یکی از مشکلاتی که در زمینه ردیابی چند شیء موجود است تغییرات مکرر شیء به دلیل تغییر در طبقه بندی های مختلف صورت گرفته به وسیله ماژول تشخیص در مرحله اولیه است. برای این مشکل، راه حلی که اتخاذ شده است بدین صورت است که برای هر شیء یک بافر ذخیره طبقه بندی در نظر گرفته می شود و برچسب طبقه بند ماکزیمم به عنوان برچسب اصلی در نظر گرفته می شود این راه حل در شکل ۳ نمایش داده شده است. مشکل دیگری که وجود دارد هنگامی است که جسم مسدود می شود. برای حل این مشکل توسط رگرسیون چندگانه موقعیت جسم در فریم مسدود شده پیش بینی می شود.



شکل ۴ - برای شیء شناسایی شده در فریم های مختلف برچسب های مختلفی خورده است که با استفاده از بافر کردن اطلاعات شیء غالب به عنوان برچسب شناخته می شود.

۳- پیاده سازی

برای پیاده سازی، از محیط رایگان google collab استفاده می گردد که برای آموزش مدل آسانتر و سریعتر است. همچنین، نوشتن و استفاده

¹ False Negative

که در آن:

IDTP: تعداد شناسه واقعی پیش‌بینی شده به عنوان شناسه واقعی
 IDFP: تعداد شناسه نادرست پیش‌بینی شده به عنوان شناسه واقعی
 IDFN: تعداد شناسه واقعی پیش‌بینی شده به عنوان شناسه نادرست
 MT منعکس کننده حد بالای درجه ردیابی است. هنگامی که یک مسیر تا بیش از ۸۰٪ ردیابی می‌شود، می‌توان آن را به عنوان MT در نظر گرفت. سوئیچ‌های ID روی نتایج MT تأثیری نمی‌گذارند و فقط بر روی خود مسیر تمرکز می‌کنند.

ML نشان دهنده درجه تلفات ردیابی یک مسیر است که معمولاً همراه با MT برای منعکس کردن حد پایین درجه ردیابی استفاده می‌شود. هنگامی که یک مسیر تنها کمتر از ۲۰٪ ردیابی می‌شود، می‌توان آن را به عنوان ML در نظر گرفت.
 IDs تعداد دفعاتی را نشان می‌دهد که شناسه اختصاص داده شده توسط مرجع، تغییر کرده است.
 نکته: توجه داشته باشید که علامت ↑ در کنار هر پارامتر بدین معناست که هرچه پارامتر مقادیر بزرگتری داشته باشد مناسب‌تر است و علامت ↓ نشان دهنده این است که هرچه پارامتر مقادیر کوچکتری داشته باشد مناسب‌تر است.

جدول ۱- مقایسه عملکرد سیستم‌های ردیابی

MT ↑	ML ↓	FN ↓	FP ↓	IDs ↓	IDF1 ↑	MOTA ↑	روش رهگیری
11.7%	30.9%	32615	7318	1001	-	33.4	روش SORT [۲۰]
19%	36.9%	79149	3280	682	52.5	54.4	Tracktor++ [۲۱]
20.7%	35.8%	76844	2394	617	54.9	56.2	Tracktor++V2 [۲۲]
27.3%	35%	68432	6167	535	67.6	58.8	LPC_MOT [۲۳]
24.1%	30.8%	68923	3660	616	58.7	59.9	Mfi_tst [۲۴]
17%	38.7%	83707	9792	735	53.5	48.3	BLSTM_MTP_O [۲۵]
26.1%	36.5%	71026	4074	370	62.4	58.6	IQHAT [۲۶]
33.3%	28.2%	57269	8328	428	67.1	63.8	UTM [۲۷]
-	-	38039	4442	672	65.5	60.9	SFFSORT [۱۶]
37.9%	19.6%	50956	10876	571	64.4	65.3	روش پیشنهادی

تشخیص اهداف، همراه با توسعه روش‌های جدید و هموار کردن راه برای پیاده‌سازی‌های عملی در حوزه‌های مختلف پیشنهاداتی ارائه گردیده است. براین اساس، روش پیشنهادی به دو بخش تقسیم می‌شود: اول تشخیص شیء است که اگر در این مرحله اطلاعاتی از بین برود در ادامه جبران این اطلاعات از دست رفته امکان‌پذیر نیست. برای این کار از سری YOLO استفاده شد. دوم، ردیابی شیء است که این بخش نیز شامل سه مرحله، استخراج ویژگی، پیش‌بینی موقعیت، و پیوند داده و تطبیق اشیاء است که بترتیب، آموزش انتقالی سری YOLOv8، فیلتر کالمن، و الگوریتم مجارستانی مورد استفاده قرار گرفت.

با بررسی نتایج درج‌شده در جدول ۱ مشاهده می‌شود که معیار MOTA و در نتیجه دقت ردیابی اشیاء افزایش پیدا کرده‌است که دلیل آن می‌تواند استفاده از فیلتر کالمن به جای فیلتر ذرات و همچنین استفاده از YOLO انتقالی به جای الگوریتم‌های سبک باشد. همچنین معیار IDF1 نیز مقدار نسبتاً قابل قبولی را ثبت کرده‌است و نشان‌دهنده این است که دقت شناسه نیز خوب است که این نیز به دلیل بافر اطلاعات است. اما برتری را می‌توان در معیارهای MT و ML به وضوح درک کرد که وقتی یک شیء شناسایی می‌شود آن شیء و مسیر را به راحتی گم نمی‌کند که این امر نیز به سبب استفاده از رگرسیون چندجمله‌ای است. البته در مقابل این بهبود، بالا رفتن سرعت و سربرار محاسباتی پیش می‌آید.

۶- منابع

- [1] Y. Cui, B. Hou, Q. Wu, B. Ren, S. Wang, and L. Jiao, "Remote sensing object tracking with deep reinforcement learning under occlusion," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1-13, 2021.
- [2] C. Kumar and S. Singh, "Security standards for real time video surveillance and moving object tracking challenges, limitations, and future: A case study," Multimed. Tools Appl., pp. 1-32, 2023.
- [3] S. Guo et al., "A review of deep learning-based visual multi-object tracking algorithms for autonomous

۵- نتیجه‌گیری و بحث

ردیابی اشیاء از موضوعات پرکاربرد در دنیای واقعی، از نظارت و امنیت گرفته تا ناوبری مستقل و تعامل انسان و کامپیوتر بوده است، و در حوزه هوش مصنوعی و بینایی ماشین راه‌حلی برای غلبه بر چالش‌های آن ارائه می‌گردد. در این مقاله نیز با هدف بهبود عملکرد ردیابی و

- Int. Conf. Image Process. (ICIP), 2016, pp. 3464–3468.
- [22] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, “Tracking without bells and whistles,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 941–951.
- [23] P. Dai, R. Weng, W. Choi, C. Zhang, Z. He, and W. Ding, “Learning a proposal classifier for multiple object tracking,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 2443–2452.
- [24] J. Yang, H. Ge, J. Yang, Y. Tong, and S. Su, “Online multi-object tracking using multi-function integration and tracking simulation training,” *Appl. Intell.*, vol. 52, no. 2, pp. 1268–1288, 2022.
- [25] C. Kim, L. Fuxin, M. Alotaibi, and J. M. Rehg, “Discriminative appearance modeling with multi-track pooling for real-time multi-object tracking,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 9553–9562.
- [26] Y. He, X. Wei, X. Hong, W. Ke, and Y. Gong, “Identity-quantity harmonic multi-object tracking,” *IEEE Trans. Image Process.*, vol. 31, pp. 2201–2215, 2022.
- [27] S. You, H. Yao, B. Bao, and C. Xu, “UTM: A unified multiple object tracking model with identity-aware feature enhancement,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2023, pp. 21876–21886.
- driving,” *Appl. Sci. (Switzerland)*, vol. 12, no. 21, p. 10741, Nov. 2022.
- [4] N. Hu et al., “LettuceTrack: Detection and tracking of lettuce for robotic precision spray in agriculture,” *Front. Plant Sci.*, vol. 13, p. 1003243, 2022.
- [5] A. Rangesh and M. M. Trivedi, “No blind spots: Full-surround multi-object tracking for autonomous vehicles using cameras and lidars,” *IEEE Trans. Intell. Veh.*, vol. 4, no. 4, pp. 588–599, 2019.
- [6] P. Ehmann et al., “360-Multiple object tracking in team sport athletes: Reliability and relationship to visuospatial cognitive functions,” *Psychol. Sport Exerc.*, vol. 55, p. 101952, 2021.
- [7] C. Abeysinghe, C. Reid, H. Rezatofighi, and B. Meyer, “Tracking different ant species: An unsupervised domain adaptation framework and a dataset for multi-object tracking,” in Proc. Int. Joint Conf. Artif. Intell. (IJCAI), 2023, pp. 546–554.
- [8] R. R. V. Leocádio, A. K. R. Segundo, and G. Pessin, “Multiple object tracking in native bee hives: A case study with Jataí in the field,” in *Brazilian Conf. Intell. Syst.*, Springer, 2023, pp. 176–191.
- [9] M. A. B. Zuraimi and F. H. K. Z. Kamaru Zaman, “Vehicle detection and tracking using YOLO and DeepSORT,” in Proc. IEEE 11th Symp. Comput. Appl. Ind. Electron. (ISCAIE), 2021, pp. 23–29.
- [10] S. Zhang and X. Wang, “Human detection and object tracking based on histograms of oriented gradients,” in Proc. 2013 9th Int. Conf. Nat. Comput. (ICNC), IEEE, 2013, pp. 1349–1353.
- [11] B. F. Momin and T. M. Mujawar, “Vehicle detection and attribute based search of vehicles in video surveillance system,” in Proc. 2015 Int. Conf. Circuits Power Comput. Technol. (ICCPCT), IEEE, 2015, pp. 1–4.
- [12] Y. Zhang et al., “FairMOT: On the fairness of detection and re-identification in multiple object tracking,” *Int. J. Comput. Vis.*, vol. 129, no. 11, pp. 3069–3087, 2021.
- [13] R. Stone, “{CenterTrack}: An {IP} overlay network for tracking {DOS} floods,” in Proc. 9th USENIX Security Symp., 2000.
- [14] Z. Wang et al., “Towards real-time multi-object tracking,” in *Lect. Notes Comput. Sci.*, vol. 12356, Springer, 2020, pp. 107–122.
- [15] E. Yu, Z. Li, S. Han, and H. Wang, “RelationTrack: Relation-aware multiple object tracking with decoupled representation,” *IEEE Trans. Multimed.*, vol. 25, pp. 2686–2697, 2023.
- [16] H. Wu et al., “Sort-YM: An algorithm of multi-object tracking with YOLOv4-Tiny and motion prediction,” *Electronics (Switzerland)*, vol. 10, no. 18, 2021.
- [17] T. Zhonglin et al., “SFFSORT multi-object tracking by shallow feature fusion for vehicle counting,” *IEEE Access*, vol. 11, pp. 76827–76841, 2023.
- [18] Z. Huangfu and S. Li, “Lightweight You Only Look Once v8: An upgraded YOLO v8 algorithm for small object identification in unmanned aerial vehicle images,” *Appl. Sci.*, vol. 13, no. 22, p. 12369, 2023.
- [19] K. Bernardin and R. Stiefelhagen, “Evaluating multiple object tracking performance: The CLEAR MOT metrics,” *EURASIP J. Image Video Process.*, vol. 2008, pp. 1–10, 2008.
- [20] E. Ristani, F. Solera, R. S. Zou, R. Cucchiara, and C. Tomasi, “Performance measures and a data set for multi-object multi-camera tracking,” in Proc. ECCV Workshops, 2016.
- [21] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple online and realtime tracking,” in Proc. IEEE