

تشخیص و رهگیری سلاح گرم و سرد با استفاده از شبکه‌های یادگیری عمیق

علی فیضی^۱، کوروش داداش تبار احمدی^{۲*}، امین توحیدی فر^۳

۱- و ۳- دانشجوی کارشناسی ارشد، ۲- استادیار، دانشگاه صنعتی مالک‌اشتر

(دریافت: ۱۴۰۱/۰۳/۱۸، پذیرش: ۱۴۰۱/۱۰/۲۶)

چکیده

تأمین و افزایش امنیت در مکان‌های خاص و عمومی همیشه مورد توجه همه افراد در زمینه‌های مختلف بوده است. به‌منظور تأمین و افزایش امنیت در مکان‌هایی عمومی همانند مدارس، دانشگاه‌ها، دفاتر و... اقدامات زیادی از جمله نصب دوربین‌های نظارتی و قراردادن مأمور در محل صورت گرفته است. از آنجایی که این دوربین‌ها توسط اپراتورهای انسانی فعالیت‌های مشکوک را رصد می‌کنند، عواملی مانند حواس پرتی، عدم آگاهی، خستگی و بسیاری از عوامل دیگر می‌تواند روی کیفیت نظارت تأثیر بگذارد، سیستم تشخیص سلاح‌های گرم و سرد به‌صورت خودکار در صورت تشخیص و شناسایی سلاح‌های تعریف شده برای آن می‌تواند به‌سرعت به مأمورین نظامی و امنیتی پیام‌های خاص را اطلاع دهد. در این مقاله به‌منظور تشخیص و رهگیری سلاح گرم (اسلحه کمری) و سلاح سرد (چاقو) از الگوریتم Yolo نسخه ۵ برای تشخیص شیء و از الگوریتم Deepsort برای رهگیری استفاده شده است که عملکرد این سیستم در مقاله مورد بررسی قرار گرفته است. در نهایت سیستم پیشنهادی پس از آموزش به‌وسیله مجموعه داده اختصاصی متشکل از سلاح‌های گرم و سرد عملیات تشخیص سلاح گرم و سرد را با دقت ۹۹.۵۰٪ انجام می‌دهد همچنین سیستم پیشنهادی از نظر مقیاس، چرخش، انسداد دقت مناسبی ارائه داده است.

کلیدواژه‌ها: بینایی کامپیوتر، یادگیری عمیق، تشخیص اشیاء، شبکه‌های عصبی عمیق، تشخیص سلاح سرد و گرم.

Detection and Tracking of Firearms and Cold Weapons Using Deep Learning Networks

A. Fayzi, K. Dadashtabar*, A. Tohidi Far

Malek Ashtar University of Technology

(Received: 2022/06/08; Accepted: 2023/01/16)

Abstract

Providing and increasing security in special and public places has always been the focus of everyone in various fields. Many measures such as installing surveillance cameras and placing officers on the spot have been taken to provide and increase the security in public places such as schools, universities, offices, and so on. Because these cameras detect suspicious activity by human operators, factors such as distraction, lack of awareness, fatigue, and a variety of other factors can affect the quality of surveillance; however, the firearms detection system detects and identifies defined weapons automatically. They can send special messages to military and security officials in a timely manner. In this paper, the Yolo version 5 algorithm is used to detect and track firearms (pistols) and cold weapons (knives), and the Deepsort algorithm is used for tracking. The article investigates the system's performance. Finally, after training with a special data set of firearms and cold steel, the proposed system performs firearms and cold steel detection operations with 99.50 percent accuracy. Furthermore, the proposed system has provided adequate accuracy in terms of scale, rotation, and obstruction.

Keywords: Computer Vision, Deep Learning, Object Detection, Object Tracking, Deep Neural Networks, Firearm Detection.

۱. مقدمه

در دهه‌های اخیر، یادگیری عمیق و کاربرد آن در بینایی کامپیوتری به دلیل عملکرد بالای نمایش داده شده در چندین تحقیق، آزمایش و برنامه‌های کاربردی، محبوبیت زیادی کسب کرده است، حتی در بیشتر روش‌های بینایی کامپیوتری، با استفاده از یادگیری عمیق به نتایج بهتری نسبت به روش‌های بینایی کامپیوتری کلاسیک رسیده است. این مقاله بر روی توسعه یک سیستم تشخیص و ره‌گیری سلاح (گرم و سرد) خودکار با استفاده از یادگیری عمیق متمرکز شده است. یک سیستم یکپارچه، متشکل از بخش تشخیص شیء و تکنیک‌های ره‌گیری ایجاد شده است که قادر به تشخیص و شناسایی سلاح‌ها در چندین منظر دوربین و ره‌گیری آن در فریم‌های ویدئو است. در نتیجه، سیستم در همان منطقی که با آن آموزش داده شده بود، دقت بالایی را گزارش می‌کند و همچنین هنگام آزمایش با دیدگاه‌های مختلف، سیستم تشخیص سلاح با تعمیم ویژگی‌های یاد گرفته شده در زمان آموزش می‌تواند دقت مناسبی را ارائه دهد.

در این تحقیق، هدف ما توسعه یک سیستم امنیتی نظارت هوشمند است که سلاح‌ها به‌ویژه اسلحه‌های کمری و چاقو را شناسایی کند. برای این منظور، ما از چند روش بینایی محاسباتی و یادگیری عمیق برای شناسایی یک سلاح از تصویر گرفته شده استفاده کرده‌ایم. کارهای اخیر در زمینه یادگیری ماشین و یادگیری عمیق به‌ویژه شبکه‌های عصبی کانولوشن پیشرفت قابل توجهی را در زمینه‌های تشخیص و شناسایی اشیاء، منحصراً در تصاویر نشان داده است. به‌عنوان اولین گام برای هر برنامه نظارت تصویری، تشخیص و طبقه‌بندی اشیاء برای کارهای بعدی همانند ره‌گیری اشیاء ضروری است. برای این منظور، ما مدل طبقه‌بندی‌کننده YOLO نسخه ۵ را آموزش دادیم، که این مدل یک طبقه‌بندی تشخیصی در زمان واقعی است و همچنین با استفاده از الگوریتم DeepSort عملیات ره‌گیری شیء تشخیص داده شده را انجام می‌دهیم.

این کار تلاشی برای طراحی و توسعه سیستمی است که می‌تواند اسلحه کمری و چاقو را در کوتاه‌ترین زمان با به‌کارگیری منابع محاسباتی کمتر شناسایی کند. در نهایت، در آینده این سیستم‌های مبتنی بر رایانه با ماشین‌های هوشمند، ربات‌ها یا ربات‌های انسان‌نما جایگزین خواهند شد. به‌منظور ارائه حس بینایی به ربات‌ها، تشخیص اشیاء و ره‌گیری آن نقش اساسی در درک اشیاء و تفسیر آن و دنبال کردن آن ایفا می‌کند. سیستم پیشنهادی ما همچنین می‌تواند در محیط‌ها و مکان‌های حساس به‌منظور حفظ امنیت برای شناسایی سلاح پیاده و راه‌اندازی شود.

۲. پیشینه تحقیق

تشخیص و شناسایی اشیاء در زمینه‌های تحقیقاتی مختلف مورد استفاده قرار می‌گیرد و به طور خاص در زمینه تشخیص و شناسایی اسلحه توسط ابزارهای شبکه‌های عصبی عمیق، تحقیقات و کارهای بسیاری انجام شده است که چندین مورد از این تحقیقات به این صورت هست: برای تشخیص تفنگ مبتنی بر تقسیم‌بندی رنگی با استفاده از الگوریتم خوشه‌بندی k-means و آشکارساز نقطه مورد نظر کار شده است که ترکیبی از آشکارساز نقطه مورد نظر Harris و توصیفگر FREAK برای تشخیص اینکه آیا یک ماسک داده شده تفنگ است یا نه استفاده می‌شود [۱].

در طی یک تحقیق مجموعه داده‌ای را طراحی و توسعه دادند که در آن مجموعه داده تفنگ‌ها برای انجام عملیات یادگیری در یک محیط بدون محدودیت قرار می‌گرفتند. این مجموعه داده، مجموعه‌ای از ۲۵۰ فیلم ضبط شده است که از حدود ۲۵۰۰ فریم غیرمشابه برچسب‌گذاری شده تشکیل شده است. در این کار M2Det باهدف آموزش مجموعه داده توسعه یافته به کار گرفته شد. علاوه بر این، ویژگی‌های پایه بهبود یافتند که برای بهبود عملکرد این آشکارساز M2Det، ویژگی‌های کم‌عمق، متوسط و عمیق از ستون فقرات مطابق با ناحیه پذیرش آن ادغام شدند. نتایج تجربی نشان داد که مجموعه داده توسعه یافته دقتی در حدود ۸۷،۴۲ درصد دارد [۲].

برای شناسایی آتش و تفنگ‌های دستی در مناطقی که از طریق دوربین‌ها نظارت می‌شدند، یک الگوریتم یادگیری عمیق بر اساس الگوریتم YOLO ورژن ۳ فرموله شد که به پردازش یک ویدئوی فریم به فریم برای شناسایی این نوع ناهنجاری‌ها در زمان واقعی و ایجاد هشدار برای مقامات مربوطه کمک می‌کند. الگوریتم نهایی خطای اعتبارسنجی را در حدود ۰،۲۸۶۴ ارائه کرد. دقت به‌دست آمده در مجموعه داده IMFDB حدود ۸۹،۳٪، ۸۲،۶٪ در UGR و ۸۶،۵٪ در FireNet ارائه شد [۳].

همچنین روشی معرفی شد [۴] که در آن الگوریتم YOLO-v3 [۵] پیاده‌سازی و تعداد FPها و FNها با الگوریتم شبکه عصبی کانولوشن مبتنی بر منطقه سریع‌تر (RCNN) [۶] مقایسه شد. نتایج با توسعه مجموعه داده‌ای متشکل از تفنگ‌های دستی با زوایای احتمالی و همراه با مجموعه داده ImageNet [۷] افزایش یافت. الگوریتم معرفی شده به‌منظور آموزش داده‌های ترکیبی به کار گرفته شد و از چهار ویدیوی مختلف برای تأیید نتایج این الگوریتم استفاده شد. تکنیک معرفی شده قادر به تشخیص تفنگ دستی در صحنه‌های مختلف همراه با چرخش‌ها، مقیاس‌ها و اشکال مختلف بود. نتایج نشان داد که الگوریتم معرفی شده به دلیل سرعت بیشتر و دقت بالاتر به راحتی به می‌توان عنوان جایگزینی برای RCNN سریع‌تر مورد استفاده قرار بگیرد. علاوه بر

حالت‌ها و انواع تشخیص اشیاء: به صورت کلی می‌توان تشخیص اشیاء را به رویکردهای مبتنی بر یادگیری ماشین و رویکردهای مبتنی بر یادگیری عمیق تقسیم کرد که در رویکردهای سنتی‌تر مبتنی بر یادگیری ماشین، تکنیک‌های بینایی کامپیوتری برای مشاهده ویژگی‌های مختلف یک تصویر، مانند هیستوگرام یا لبه‌ها، برای شناسایی گروه‌هایی از پیکسل‌هایی که ممکن است به یک شیء تعلق داشته باشند، استفاده می‌شود. پس از آن این ویژگی‌ها به یک مدل رگرسیونی برای این‌که مکان شیء را همراه با برچسب آن پیش‌بینی کند وارد می‌شدند.

از طرف دیگر، رویکردهای مبتنی بر یادگیری عمیق از شبکه‌های عصبی کانولوشنال (CNN) [۹] برای انجام تشخیص شیء استفاده می‌کنند که در آن ویژگی‌ها نیازی به تعریف و استخراج جداگانه ندارند و این کار به صورت خودکار توسط شبکه‌های عصبی کانولوشنال انجام می‌پذیرد. در زمینه تشخیص اشیاء این مفاهیم بسیار کاربردی و مهم هستند که به توضیح آن‌ها می‌پردازیم:

جعبه‌های مرزی یا Bounding Box: یک چهارضلعی تا حد ممکن کوچک که برای احاطه کردن شیء مورد نظر استفاده می‌شود. این کادر عموماً با چهار مقدار توصیف می‌شود: (bx, by, bh, bw). که در آن عبارت‌های (bx, by) نشان دهند مختصات مرکز کادر است و bh, bw به ترتیب نشان‌دهنده ارتفاع و عرض کادر هستند که در شکل (۲) مثالی از جعبه مرزی با مشخص کردن نقطه مرکزی و عرض و ارتفاع آن در تصویر نشان داده شده است.



شکل ۲. کادر مشخص‌کننده موقعیت شیء با مختصات (bx, by, bh, bw)

لنگرها، کادر محصورکننده (Anchor Boxes): این کادرها مجموعه‌ای از کادرهای محصورکننده از پیش تعریف‌شده با ارتفاع و عرض مشخص هستند. این کادرها برای بررسی مقیاس و نسبت دسته‌های مشخص اشیاء تعریف‌شده و معمولاً بر اساس اندازه اشیاء در مجموعه داده‌های آموزشی انتخاب می‌شوند. در طول فرآیند تشخیص، کادرهای محصورکننده از پیش تعریف‌شده در سراسر عکس به کار برده می‌شوند. شبکه به پیش‌بینی احتمال و سایر ویژگی‌ها از جمله پس‌زمینه، اشتراک نسبت به اجتماع (IoU)

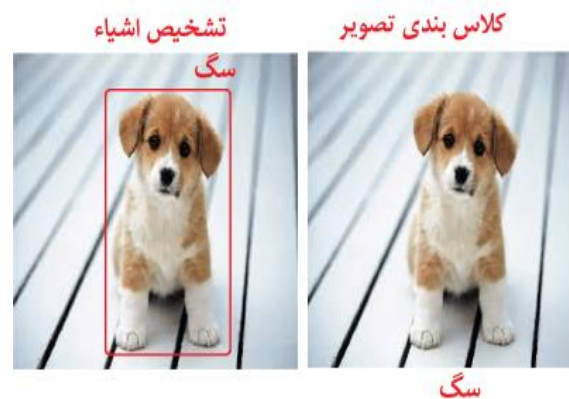
این، این الگوریتم به طور مؤثر در یک محیط زمان واقعی کار می‌کرد.

یک سیستم تشخیص سلاح را بر اساس مجموعه‌ای از لایه‌های کانولوشنی معنایی بررسی کردند که در آن مشکل شناسایی و مکان‌یابی یک سلاح به مجموعه‌ای از مشکلات کوچک‌تر مربوط به اجزای جداگانه یک سلاح تجزیه می‌شد. خروجی‌های شبکه‌های مجزا برای به دست آوردن خروجی کلی سیستم جمع شدند. نتایج نشان داد که سیستم مورد بررسی در مقایسه با رویکرد یکپارچه بر اساس یک DCNN سودمندتر بود [۸]. یک سیستم تشخیص سلاح را بر اساس مجموعه داده اختصاصی به نام Pishtaw معرفی کرده‌اند که در آن با استفاده از الگوریتم Yolo به‌دقت ۹۹٫۰۱ درصد دست یافتند [۲۲].

۳. تشخیص و ره‌گیری اشیاء

تشخیص اشیاء یکی از روش‌های بینایی کامپیوتری است که برای شناسایی و پیدا کردن مکان اشیاء درون یک تصویر یا ویدئو استفاده می‌شود. به صورت خاص، با کشیدن چهارضلعی‌هایی در اطراف این اشیاء به ما امکان تشخیص اشیاء مذکور در صحنه را می‌دهد.

به طور معمول تشخیص اشیاء با کلاس بندی تصویر به جای هم مورد استفاده قرار می‌گیرد ولی در واقعیت به این صورت نیست، در کلاس بندی به یک تصویر یک برچسب اختصاص می‌یابد. به عنوان مثال عکس یک سگ برچسب "سگ" را دریافت می‌کند. تصویری از دو سگ هم برچسب "سگ" را دریافت می‌کند. از سوی دیگر، تشخیص اشیاء، چهارضلعی را در اطراف هر سگ ترسیم می‌کند و روی جعبه برچسب "سگ" می‌گذارد. مدل پیش‌بینی می‌کند که هر شیء کجاست و چه برچسبی باید اعمال شود. به این ترتیب، تشخیص شیء اطلاعات بیشتری در مورد یک تصویر نسبت به کلاس بندی تصویر ارائه می‌دهد که شکل (۱) به صورت کاملاً واضح تفاوت بین کلاس بندی تصویر و تشخیص اشیاء را نشان می‌دهد.



شکل ۱. تفاوت کلاس بندی تصویر و تشخیص اشیاء

تضعیف غیر بیشینه (non maximum suppression): اگر چندین کادر برای یک شیء تخمین زده شده باشد، این روش همه کادرها را کنار می گذارد، به جز کادری که دارای بیشینه است که شکل (۶) نشان دهنده اعمال الگوریتم تضعیف غیر بیشینه روی خروجی الگوریتم تشخیص شیء هست.



شکل ۶. اعمال الگوریتم تضعیف غیر بیشینه

معیارهای ارزیابی: mAP معیاری است که برای تعیین دقت ابزارهای تشخیص اشیاء مورداستفاده قرار می گیرد. mAP به صورت درصد بیان می شود.

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (1)$$

معیار فوق بیانگر میانگین دقت تشخیص تمام نمونه های اشیاء در یک عکس است.

به طور کلی سه معیار برای ارزیابی عملکرد الگوریتم های تشخیص اشیاء وجود دارد: ۱- سرعت تشخیص در فریم ها بر حسب فریم بر ثانیه (FPS)^۱ ۲- precision - recall^۳ است. رایج ترین معیار مورداستفاده، AP^۲ است که از precision و recall مشتق شده است. AP معمولاً به شیوه ای خاص دسته ارزیابی می شود، یعنی برای هر دسته شیء به طور جداگانه محاسبه می شود. برای مقایسه عملکرد بر روی تمام دسته های شیء، میانگین AP (mAP^۳) بر روی تمام دسته های شیء به عنوان معیار نهایی عملکرد به کار گرفته می شود.

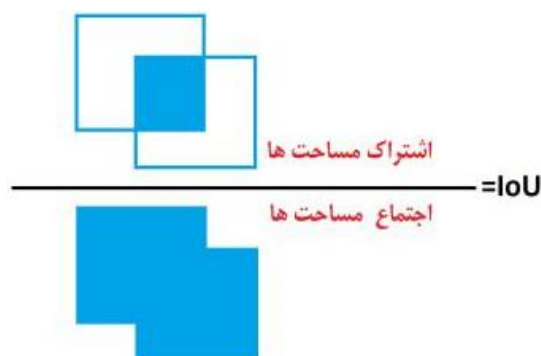
معیارهای دقت (Precision) و صحت (Recall): معیار دقت، درصد پیش بینی های مثبتی که به درستی به عنوان مثبت پیش بینی شده اند را محاسبه می کند. معیار صحت نیز نسبت «مثبت های صحیح» (True Positives) به تمامی خروجی های ممکن را محاسبه می کند. این دو معیار، ارتباط

و غیره می پردازد. پیش بینی ها برای ارتقای تک تک لنگرها مورداستفاده قرار می گیرند. همچنین امکان تعریف چندین لنگر برای اندازه های مختلف اشیاء وجود دارد که شکل (۳) مراحل کلی اعم از تعریف لنگر تا تشخیص شیء را نشان می دهد.



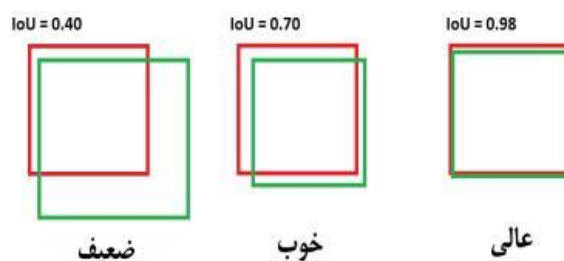
شکل ۳. تعریف لنگر برای شیء خاص و تشخیص شیء بعد از اعمال عملیات متناسب تعریف شده

اشتراک پیرامون اجتماع (IOU): IOU یک معیار ارزیابی کننده است که برای بررسی دقت کادر محصورکننده پیش بینی شده با توجه به کادر محصورکننده واقعی، استفاده می شود. که در شکل (۴) فرمول مربوطه و تصویر آن نشان دهنده نحوه محاسبه IoU هست.



شکل ۴. معیار ارزیابی IoU [۱۳]

IOU برای مقادیر بزرگتر از ۰/۵ پیش بینی خوب به حساب می آید و برای تکامل بیشتر در نظر گرفته می شود که شکل ۵ مقایسه ای برای انواع مقادیر به دست آمده با محاسبه توسط فرمول IoU نشان داده شده است.



شکل ۵. مقایسه انواع مقادیر IoU [۱۳]

¹ Frames Per Second

² Average Precision

³ Mean Average Precision

تقسیم می‌کند و اگر مقدار آن بیشتر از حد آستانه باشد که برای آن تعیین شده، به منزله وجود شیء است و اگر کمتر باشد این نتیجه گرفته می‌شود که در آن محل شیء وجود ندارد و یا Bounding Box پیشنهادی الگوریتم شیء موردنظر را پوشش نمی‌دهد.

- فرمول محاسبه IOU:

$$IOU(b, b^g) = \frac{Area(b \cap b^g)}{Area(b \cup b^g)} \quad (۶)$$

- که اگر IOU کمتر مساوی حد آستانه باشد آنگاه یک FP داریم.

- و یا اگر IOU بزرگ‌تر از حد آستانه تعریف شده باشد آنگاه یک TP داریم.

در فرمول (۶) عبارت b ، Bounding Box یا جعبه پیشنهادی و b^g نیز مکان واقعی شیء است.

یادگیری انتقالی (Transfer Learning): یادگیری انتقالی [۱۰]، مفهومی است که در آن از یک شبکه عصبی از پیش آموزش داده شده (جهت دسته‌بندی تصاویر)، برای مقاصد دیگر (نظیر دسته‌بندی، تشخیص اشیاء و ...) استفاده می‌شود. روش کار بدین صورت است که جهت یادگیری انتقالی، مدل از پیش آموزش داده شده توسط داده‌های جدید و به‌وسیله پارامترهای متناسب با این داده‌ها، آموزش دوباره (Re-Train) داده می‌شود (تعدادی از پارامترهای وزن شبکه عصبی در نتیجه چنین فرآیندی، دوباره تنظیم می‌شوند). از آنجایی که در یادگیری انتقالی نیاز به یادگیری و تنظیم تعداد زیادی پارامتر وزن وجود ندارد، چنین کاری باعث صرفه‌جویی بسیار زیاد در زمان و قدرت محاسباتی لازم برای آموزش شبکه عصبی می‌شود یادگیری انتقالی می‌تواند در بسیاری از کارهای یادگیری عمیق همانند پردازش زبان طبیعی، برچسب‌گذاری تصویر، شناسایی تصویر، کلاس‌بندی تصویر و ... مورد استفاده قرار بگیرد.

ره‌گیری اشیاء: ره‌گیری اشیاء [۱۱] یکی از کاربردهای بینایی کامپیوتر است که با استفاده از یادگیری عمیق به نتایج خوبی دست‌یافته شده است که در آن سیستم یک مجموعه اولیه از تشخیص اشیاء را به‌عنوان ورودی می‌گیرد و یک شناسه منحصر‌به‌فرد برای هر یک از تشخیص‌های اولیه ایجاد می‌کند و سپس اشیاء شناسایی شده را در حین حرکت در فریم‌های یک ویدیو ردیابی می‌کند.

در ره‌گیری اشیاء معمولاً در اطراف شیء مورد ردیابی کادر مستطیلی (مستطیل مرزی) وجود دارد که این مستطیل موجود

معکوسی با یکدیگر دارند. شکل (۷) نشانگر ماتریس درهم‌ریختگی است.

کلاس پیش‌بینی شده

	مثبت	منفی
مثبت	مثبت صحیح (True Positive)	منفی غلط (False Negative)
منفی	مثبت غلط (False Positive)	منفی صحیح (True Negative)
کلاس واقعی		

شکل ۷. ماتریس درهم‌ریختگی

- فرمول محاسبه معیار precision به‌صورت زیر است:

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

- فرمول محاسبه معیار recall به‌صورت زیر محاسبه می‌شود:

$$Recall = \frac{TP}{TP + FN} \quad (۳)$$

- همچنین معیار AP نیز به این صورت محاسبه می‌شود:

$$AP = \sum_{k=0}^{k=n-1} [\text{recall}(k) - \text{recall}(k+1)] \times \text{precision}(k) \quad (۴)$$

در فرمول (۴)، n برابر با حد آستانه^۱ است.

- معیار mAP نیز به این شکل محاسبه می‌شود:

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k \quad (۵)$$

در فرمول بالا AP_k نشان‌دهنده معیار AP برای کلاس k است، n نیز نشان‌دهنده تعداد کلاس اشیاء است.

از معیارهایی دیگر که به‌منظور محاسبه و اندازه‌گیری کارایی منطقه شناسایی شده به‌وسیله Bounding Box مورد استفاده قرار می‌گیرد، معیار IOU^۲ است. این معیار میزان صحیح بودن Bounding Box تولیدشده توسط الگوریتم را اندازه‌گیری می‌کند که نشان‌دهنده این است که منطقه پیشنهادی شامل شیء است یا نه.

معیار IOU به این صورت عمل می‌کند که اشتراک Bounding Box پیشنهادی با Box واقعی اطراف شیء را بر اجتماع آن‌ها

^۳ Ground Truth

^۱ Threshold

^۲ Intersection Over Union

که در آن y_i و S_i نشان دهنده میانگین و مشاهدات جعبه محدودکننده ماتریس کوواریانس برای مسیر i -ام، و z نشان دهنده z -امین جعبه محصورکننده شناسایی شده است. فاصله بصری $d^{(2)}$ به توصیفگرهای ویژگی ظاهری متکی است و با عبارت:

$$d^{(2)}(i, j) = \min\{1 - r_j^T r_{jk}^{(i)} | r_{jk}^{(i)} \in \mathcal{R}_i\} \quad (9)$$

نشان داده می شود که در آن \mathcal{R}_i توصیفگر فرم استخراج شده از بخشی از تصویر در z -امین کادر محصورکننده شناسایی شده است، و \mathcal{R}_i مجموعه ای از ۱۰۰ توصیف کننده ظاهر \mathcal{R}_i^A مرتبط با ره گیر i است.

در این مقاله عملکرد الگوریتم DeepSort به سه مرحله تقسیم می شود:

۱. **شناسایی:** در این مرحله برای محاسبه و شناسایی اشیاء از یک متد شناسایی اشیاء مبتنی بر شبکه عصبی پیچشی استفاده می شود (در این مقاله از الگوریتم YOLO [۵] استفاده شده است).

۲. **برآورد:** مرحله میانی این رویکرد که پیش از مرحله مطابقت قرار می گیرد، برآورد مدل است. در این مرحله وضعیت مسیر و ردپای هر یک از اشیاء به عنوان یک بردار با ۸ مقدار تعریف می شود که عبارتند از: مرکز کادر (x, y) ، مقیاس کادر (s) ، نسبت ابعاد کادر (a) و مشتقات آن ها در طول زمان که همان نرخ زمانی تغییر موقعیت است. سپس این وضعیت ها با استفاده از فیلتر کالمن در قالب یک سیستم پویا مدل سازی می شوند. اگر در چندین فریم متوالی ردی از شیء شناسایی نشود، آن شیء خارج از فریم یا گمشده در نظر گرفته می شود؛ و ردپای کادر جدیدی که شناسایی شده، آغاز خواهد شد.

۳. **مطابقت:** در مرحله آخر، با داشتن وضعیت پیش بینی شده توسط اطلاعات پیشین در فیلتر کالمن و کادر جدید شناسایی شده در فریم کنونی، شیء جدید با شیء که در فریم های پیشین ردیابی شده بود، مقایسه و مطابقت داده می شود. این کار با اعمال الگوریتم هانگاربانیر تطبیق نمودار دوبخشی انجام می شود. همچنین می توان با فرمول بندی فاصله تطبیق وزن داد و این فرآیند را تقویت کرد.

* به صورت عمومی ره گیری اشیاء به دودسته تقسیم می شود:

۱. **ره گیری تک شیء:** ره گیری تک شیء جز ساده ترین بخش الگوریتم ره گیری است. در این کار، هدف الگوریتم به این صورت است که به سادگی روی یک شیء در تصویر تمرکز می کند و آن را تا زمانی که از کادر خارج می شود، ره گیری

در اطراف شیء به کاربر نشان می دهد که شیء مورد ره گیری در کجای صفحه قرار دارد.

روش های کلاسیک ره گیری چند شیء به دو بخش تقسیم می شوند:

شناسایی: ابتدا تمامی اشیاء مورد نظر شناسایی می شوند.

مطابقت: سپس اشیاء مشابه با توجه به فریم پیشین، با یکدیگر تطابق داده می شوند. پس از آن این فریم های تطبیق یافته دنبال می شوند تا ردپای یک شیء به دست آید.

الگوریتم مورد استفاده در این تحقیق نیز ابتدا با استفاده از یک الگوریتم تشخیص اشیاء، شیء های موجود را پیدا می کند و سپس با استفاده از مکانیزم موجود برای ردیابی اشیاء به ردیابی اشیاء می پردازد.

الگوریتم DeepSort: الگوریتم deepsort [۱۲] یک الگوریتم ره گیری است که روشی برای ره گیری با تشخیص است که هم نتایج تشخیص پارامترهای جعبه محصورکننده و هم اطلاعات مربوط به اشیاء ردیابی شده را در نظر می گیرد تا تشخیص ها را در یک فریم جدید با اشیاء ردیابی شده قبلی مرتبط کند؛ بنابراین الگوریتم DeepSort فقط اطلاعات مربوط به فریم های فعلی و قبلی را برای پیش بینی در مورد فریم فعلی بدون نیاز به پردازش یک باره کل ویدیو در نظر می گیرد. در ابتدای ویدیو، یعنی در اولین فریم، یک شناسه یکتا به هر جعبه محصورکننده اختصاص داده می شود که یک شیء را نشان می دهد و دارای ارزش اطمینان بالاتر از آستانه تعیین شده است. الگوریتم YOLO برای تخصیص تشخیص ها در یک فریم جدید به مسیرهای موجود استفاده می شود تا تابع هزینه به مینیمم سراسری برسد. تابع هزینه شامل فاصله فضایی (Mahalanobis) $d^{(1)}$ جعبه محصورکننده شناسایی شده از موقعیت پیش بینی شده با توجه به موقعیت شناخته شده قبلی آن شیء، و یک فاصله بصری $d^{(2)}$ است که شیء شناسایی شده و تاریخچه شیء ردیابی شده را در نظر می گیرد. تابع هزینه ذکر شده یک شیء شناسایی شده z به یک الگوریتم ره گیری i با عبارت:

$$c_{i,j} = \lambda d^{(1)}(i, j) + (1 - \lambda) d^{(2)}(i, j) \quad (7)$$

نشان داده می شود؛ که در آن λ پارامتری است که می تواند برای تعیین تأثیر فاصله فضایی $d^{(1)}$ و فاصله بصری $d^{(2)}$ تنظیم شود.

فاصله مکانی $d^{(1)}$ با عبارت زیر نمایش داده می شود:

$$d^{(1)}(i, j) = (d_i - y_i)^T S_i^{-1} (d_i - y_i) \quad (8)$$

متوجه شود و آن‌ها را از هم متمایز کند. شکل (۱۰) یک عملکرد کلی از سیستم پیاده‌سازی شده را نمایش می‌دهد.



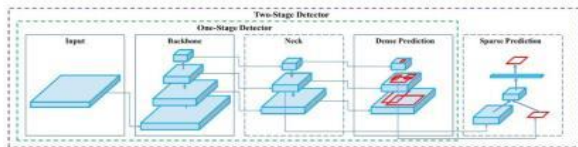
شکل ۱۰. عملکرد کلی سیستم پیشنهادی

۴-۱. روش استفاده‌شده

با توجه به مسئله اهمیت زمان در تشخیص اسلحه و شناسایی هرچه سریع‌تر افراد تروریست و سارقان مسلح و همچنین سرعت عبور آن‌ها از محدوده دید دوربین‌های نظارتی در این تحقیق از روش‌های تک‌مرحله‌ای و به طور دقیق‌تر از شبکه عصبی مصنوعی YOLO [۵] استفاده شده است. همچنین برای سرعت بخشیدن به عملکرد الگوریتم YOLO در زمان اجرا از ابزار TensorRT که در بخش‌های بعدی معرفی خواهد شد استفاده شده است.

شبکه YOLO نسخه ۵: اولین نسخه از الگوریتم YOLO [۵] در سال ۲۰۱۶ در کنفرانس CVPR^۱ ارائه شد که برگرفته شده از همان قابلیت سیستم بینایی انسان است که با یک نگاه عمل تشخیص اشیاء را انجام می‌دهد؛ بنابراین، سیستم تشخیص اشیاء YOLO باهدف ارائه روشی مشابه کارکرد سیستم بینایی انسان طراحی شده است. در شبکه YOLO تنها یک شبکه کانولوشنی وجود دارد که تصویر تغییر سایز شده ورودی را دریافت و سپس به صورت هم‌زمان چندین Box را به همراه احتمال کلاس‌ها پیش‌بینی می‌کند.

همان‌طور که در شکل (۱۱) و شکل (۱۲) نشان داده شده است، به‌طور کلی معماری YOLO دارای سه قسمت: ۱- Backbone، ۲- Neck، ۳- Head است که در ادامه هر بخش به طور مختصر توضیح داده خواهد شد.



شکل ۱۱. معماری کلی الگوریتم‌های تشخیص شی [۱۳]



شکل ۱۲. بخش‌های مختلف شبکه YOLO [۲۲]

می‌کند، همان‌طوری که در شکل (۸) مشخص هست الگوریتم یک اسلحه کمری را تشخیص داده است و سعی می‌کند آن را در مراحل بعدی رهگیری کند.



شکل ۸. رهگیری تک شی

۲. رهگیری چندگانه اشیاء: در این نوع رهگیری، هدف این است که تک‌تک اشیاء موجود موردنظر در تصویر به صورت یکتا شناسایی و همه آن‌ها را تا زمانی که از قاب خارج می‌شوند ردیابی کنیم که شکل (۹) یک مثال از الگوریتم رهگیری چندگانه را نشان می‌دهد که در آن چند اسلحه کمری متفاوت را شناسایی کرده و سعی می‌کند سلاح‌های شناسایی شده را رهگیری کند.



شکل ۹. رهگیری چندگانه اشیاء

۴. روش پیشنهادی

در این تحقیق تمرکز ما بر روی عملیات تشخیص سلاح گرم و سرد در دوربین‌های نظارتی و رهگیری آن در فریم‌های بعدی ویدئو هست که با چالش‌های مختلف روبه‌رو است و از آنجایی که عملیات تشخیص و شناسایی سلاح‌های گرم و سرد در دوربین‌های نظارتی به دلیل کم بودن کیفیت ویدئوهای دریافتی، وجود زوایای دید مختلف در دوربین‌ها و همچنین از آنجایی که ممکن است چالش‌های مختلف همچون انسداد رخ داده باشد، تشخیص و رهگیری سلاح یک مسئله چالش‌برانگیز است.

یکی از عواملی که در تشخیص سلاح‌های گرم و سرد مورد توجه قرار می‌گیرد به این صورت است که سلاح‌های گرم و سرد معمولاً به طور رایج با اشکال مختلفی وجود دارد و بیشتر این اشکال به صورت اندازه کوچک می‌باشند که این اندازه‌های کوچک ممکن است عملیات تشخیص آن‌ها را سخت‌تر کند و از آنجایی که ممکن است این سلاح‌های گرم و سرد به اجسام دیگر شباهت داشته باشد الگوریتم تشخیص اشیاء باید به خوبی آموزش داده شده باشد که بتواند تفاوت مابین سلاح‌های گرم و سرد با سایر اشیاء را

¹Conference on Computer Vision and Pattern Recognition

هرچه معماری در نسخه‌های مختلف شبکه YOLO نسخه ۵ پیچیده‌تر می‌شود میزان دقت به‌دست‌آمده بر روی مجموعه داده COCO بیشتر می‌شود که با بررسی دقیق‌تر معماری‌های مختلف این شبکه درمی‌یابیم که تفاوت دقت نسخه کوچک (YOLOv5S) شبکه با نسخه خیلی بزرگ (YOLOv5X) آن حدوداً به ۱۴ درصد می‌رسد. لازم به ذکر است نسخه خیلی بزرگ (YOLOv5X) شبکه YOLO نسخه ۵ دارای ۸۹ میلیون پارامتر و نسخه کوچک آن (YOLOv5S) تنها دارای ۷/۵ میلیون پارامتر است که به دلیل افزایش پارامترهای قابل‌یادگیری در معماری باعث افزایش دقت و کاهش سرعت شبکه می‌شود که باید سعی شود یک حد میانه برای آن انتخاب کرد لازم به ذکر است که سرعت شبکه نسخه کوچک (YOLOv5S) ۲/۱ میلی‌ثانیه و در نسخه خیلی بزرگ (YOLOv5X) به سه برابر افزایش یعنی ۶/۱ میلی‌ثانیه رسیده است.

۴-۲. معرفی مجموعه داده‌ها

در علم هوش مصنوعی و به‌خصوص در زمینه آموزش شبکه‌های عصبی عمیق وجود یک دیتاست بزرگ از واجبات آموزش شبکه هست و از آنجایی که یکی از چالش‌های مهم کاربرد هوش مصنوعی در حوزه‌های نظامی، در دسترس نبودن مجموعه داده‌های موردنیاز برای آموزش شبکه‌های عصبی بر پایه یادگیری عمیق است، در این تحقیق سعی شده است که از یک مجموعه داده نزدیک ۲۸ هزار تصویر برای سلاح‌های گرم و سرد استفاده شود که این مجموعه داده‌ها شامل دو رده سلاح گرمی و چاقو است.

چالش‌های مختلفی در زمینه جمع‌آوری مجموعه داده‌ها برای این تحقیق وجود داشته است که به‌صورت زیر فهرست می‌گردد:

۱) زاویه تصویر سلاح گرم و سرد: با توجه به اینکه سلاح‌ها از زاویه‌های مختلف، به شکل‌های مختلفی دیده می‌شوند در جمع‌آوری دیتا این قضیه موردتوجه قرار گرفته است به‌طوری‌که همان‌طور که در شکل (۱۵) مشاهده می‌کنید از زاویه‌های مختلف اعم از روبه‌رو، بالا، کنار و همچنین زاویه پشت از سلاح تصویر موجود است.

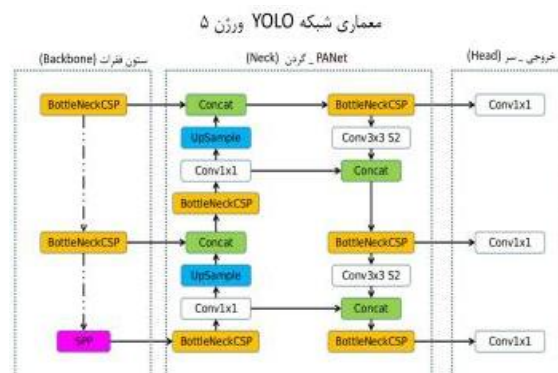


شکل ۱۵. الف) تصاویر سلاح‌ها از زاویه دید روبه‌رو، ب) تصاویر سلاح‌ها از زاویه دید بالا، ج) تصاویر سلاح‌ها از زاویه دید بغل، د) تصاویر سلاح‌ها از زاویه دید پشت

۱- **Backbone یا ستون فقرات:** از این قسمت به‌عنوان بخش استخراج‌کننده ویژگی شبکه استفاده می‌شود. در این قسمت معمولاً از شبکه‌هایی مثل VGG [۱۴]، DenseNet [۱۵]، ResNet [۱۶] و دیگر شبکه‌های مشابه استفاده می‌شود. این شبکه‌ها از قبل بر روی مجموعه داده‌های معروف مانند Imagenet یا COCO [۱۷] آموزش داده شده‌اند و با توجه به معماری عمیقی که دارند، ویژگی‌های خوبی را در سطوح مختلف استخراج می‌کنند که می‌توان آن‌ها را به‌عنوان ورودی مرحله بعد در نظر گرفت.

۲- **Neck یا گردن:** میان دو قسمت ستون فقرات (Backbone) و قسمت سر (Head) شبکه، لایه‌های اضافه به نام گردن (Neck) قرار گرفته شده است. وظیفه این لایه‌ها استخراج نقشه‌های ویژگی^۱ مختلف از مراحل مختلف ستون فقرات است. معماری‌هایی که معمولاً در این قسمت استفاده می‌شوند FPN [۱۸]، PANet [۱۹] یا Bi-FPN [۲۰] و... است.

۳- **Head یا سر:** این بخش از شبکه، خود شبکه‌ای مجزا است. این شبکه لنگرها اشیاء و یا به‌اصطلاح لنگرها^۲ را شناسایی و دنبال می‌کند و در نهایت آن‌ها را طبقه‌بندی می‌کند. خروجی این قسمت مختصات جعبه پیش‌بینی یعنی x, y, w, h است، که در کنار آن‌ها احتمال‌های وجود شیء نیز هست. شکل (۱۳) نیز به‌طور دقیق‌تر معماری شبکه YOLO نسخه ۵ را نشان می‌دهد.



شکل ۱۳. معماری شبکه YOLO نسخه ۵ و اجزای آن [۱۳]

همانند دیگر نسخه‌های شبکه YOLO، نسخه ۵ این شبکه نیز بر روی مجموعه داده COCO آموزش‌دیده است و شامل چهار معماری مختلف است که در شکل (۱۴) نمایش داده شده‌اند:

کوچک	متوسط	بزرگ	خیلی بزرگ
YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
14 MB _{FP16}	41 MB _{FP16}	90 MB _{FP16}	168 MB _{FP16}
2.0 ms _{Y100}	2.7 ms _{Y100}	3.8 ms _{Y100}	6.1 ms _{Y100}
37.2 mAP _{COCO}	44.5 mAP _{COCO}	48.2 mAP _{COCO}	50.4 mAP _{COCO}

شکل ۱۴. معماری‌های مختلف شبکه YOLO نسخه ۵

^۱ Feature Map

^۲ Anchor



شکل ۱۸. تصاویر سلاح در حالت‌های شب و روز

در نهایت مجموعه داده مورد استفاده در این تحقیق شامل ۶۰۰۰ تصویر از سلاح‌های کمری و چاقو همانند شکل (۱۹) است که بعد از اعمال عملیات افزایش داده‌ها تعداد مجموعه داده‌ها برای دو رده به ۲۸ هزار تصویر رسیده است. این مجموعه داده شامل تصاویر سلاح‌های مختلف کمری و چاقو از فاصله‌های دور و نزدیک و شرایط مختلف بحث شده، است.



شکل ۱۹. نمونه‌های تصاویر مجموعه داده

۳-۴. افزایش داده‌ها

هدف از تقویت داده‌ها افزایش تنوع تصاویر ورودی است، به طوری که مدل تشخیص شیء طراحی شده استحکام بالاتری نسبت به تصاویر به دست آمده از محیط‌های مختلف داشته باشد. به عنوان مثال، اعوجاج‌های فتومتریکی و اعوجاج هندسی دو روش متداول افزایش داده‌ها هستند و قطعاً برای کار تشخیص اشیاء مفید هستند. در برخورد با اعوجاج فتومتریکی، روشنایی، کنتراست، رنگ، اشباع و نویز یک تصویر را تنظیم می‌کنیم. برای اعوجاج هندسی، مقیاس بندی تصادفی، برش، برعکس کردن و چرخش را اضافه می‌کنیم.

در این تحقیق برای افزایش داده از روش‌های مثل برعکس کردن عمودی، افقی، چرخش، تار کردن متوسط، تغییرات رنگی استفاده شده است که تصاویر حدود ۶ هزار (۳ هزار برای سلاح کمری، ۳ هزار برای چاقو) را به تعداد ۲۸ هزار تصویر تبدیل کرده‌ایم که در شکل (۲۰) چند نمونه از این تصاویر نمایش داده شده است.

۲) **میزان فاصله از دوربین:** موضوع فاصله شیء از دوربین‌های نظارتی یکی از چالش‌های مهم در مجموعه داده‌ها است که در این کار سعی شده است این موضوع رعایت شود شکل (۱۶) نمونه‌ای از این تصاویر است.



شکل ۱۶. نمونه‌ای تصاویری که سلاح کمری در آن‌ها به دلیل فاصله زیاد، کوچک است.

۳) **انسداد:** وجود یک شیء در جلوی شیء دیگر و یا نمایش بخشی از شیء از حالت‌های مختلف انسداد است که به طور کلی یکی از چالش‌های حوزه تشخیص اشیاء است. انسداد زمانی به وجود می‌آید که بخشی از شیء مورد نظر به دلایل مختلفی از دید پنهان می‌شود. در مورد سلاح‌های گرم و سرد نیز در بیشتر اوقات ممکن است انسداد پیش آید مثلاً انسداد می‌تواند به وسیله دستان فرد مسلح ایجاد یا اینکه به دلیل پنهان کردن سلاح از دید، بخشی از آن بیرون و بخش دیگر اسلحه به وسیله اشیاء دیگر پنهان شده باشد. به منظور اینکه سیستم تشخیص سلاح گرم و سرد بتواند در صورت انسداد هم عملیات شناسایی شیء را با موفقیت انجام دهد، سعی شده است که از حالت‌های مختلف انسداد در مجموعه داده‌ها استفاده شود که شکل (۱۷) حالت‌های مختلف را نشان می‌دهد که به دلایل مختلف دچار انسداد شده‌اند.



شکل ۱۷. تصاویر سلاح در حالت انسداد

۴) **شرایط نوری:** از آنجایی که در دنیای واقعی شرایط نوری متفاوتی وجود دارد در مجموعه داده‌ها سعی بر آن شده که داده‌ها در حالت‌های کم‌نور و پر نور همانند آنچه در شکل (۱۸) نشان داده شده است، در مجموعه داده موجود باشند.

۵. نتایج و بحث

همان‌طور که گفته شد در این تحقیق از تمام معماری YOLO نسخه ۵ استفاده شده است. اطلاعات این معماری‌ها در جدول (۱) آورده شده است:

جدول ۱. اطلاعات معماری‌های استفاده‌شده از شبکه YOLO

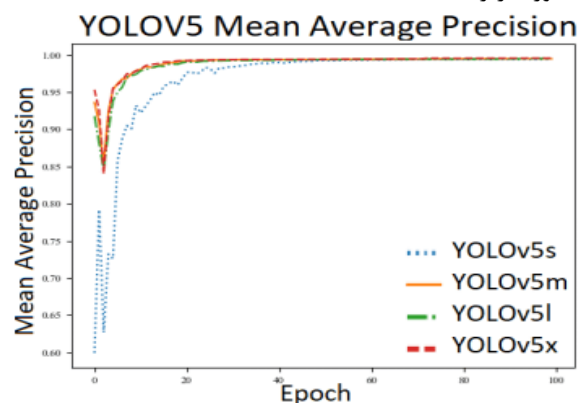
نوع معماری	تعداد لایه‌ها	تعداد پارامترها	زمان تشخیص	mAP (COCO)
YOLOv5s	۲۷۰	۷/۵ میلیون	۲/۱ ms	۳۶/۸
YOLOv5m	۳۶۹	۲۱/۸ میلیون	۳/۰ ms	۴۴/۵
YOLOv5l	۴۶۸	۴۷/۸ میلیون	۳/۹ ms	۴۸/۱
YOLOv5x	۵۶۷	۸۹ میلیون	۶/۱ ms	۵۰/۱

معماری‌های نامبرده را توسط مجموعه داده با هایپرپارامترهای که در جدول (۲) به طور خلاصه آورده شده‌اند آموزش داده‌ایم. لازم به ذکر است که آموزش در بستر سیستم با واحد پردازش گرافیکی A5000 صورت گرفته است.

جدول ۲. نمونه‌هایی از هایپرپارامترهای آموزش

Epoch	Threshold IOU	آموزش دوباره (لایه)	Lr (Learning rate)	تصاویر ارزیابی	تصاویر آموزش
۱۰۰	۰/۲	متغیر	۰/۰۱	۴۰۰۰	۲۴۰۰۰

همان‌طور که در جدول (۲) مشخص است، از میان تصاویر مجموعه داده‌ها، تعداد ۲۴۰۰۰ تصویر را برای آموزش و تعداد ۴۰۰۰ را نیز برای ارزیابی شبکه در نظر گرفته شده است. حد آستانه برای محاسبه TP برحسب IOU نیز برابر با ۰/۲ قرار داده شده است و در نهایت شبکه به تعداد ۱۰۰ مرحله آموزش داده شده است. معیارهای مورد استفاده شده برای ارزیابی قبلاً در بخش ۳ معرفی شدند. با توجه به شکل (۲۲) معیار mAP برای حد آستانه برابر با ۰/۵ در ۱۰۰ مرحله آموزش محاسبه شده است که نتایج آن به صورت زیر است.



شکل ۲۲. مقایسه معیار mAP چهار معماری YOLOv5 بر روی مجموعه داده



شکل ۲۰. نمونه تصاویر افزایش داده‌ها (چرخش عمودی، افقی، تغییر شدت رنگ، تار کردن و...)

۴-۴. معرفی TensorRT

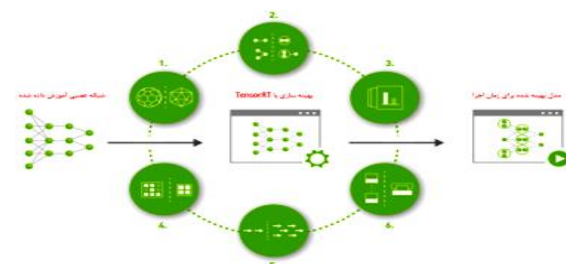
TensorRT که به اختصار TRT خطاب می‌شود، کتابخانه‌ای است که توسط شرکت NVIDIA برای استنتاج سریع‌تر با استفاده از سخت‌افزارهای پردازش گرافیکی NVIDIA ارائه شده است [21]. TensorRT بهینه‌سازی‌هایی را برای انواع داده ۸ INT و ۱۶ FP را باهدف استقرار کاربردهای استنتاج یادگیری عمیق از جمله پخش ویدیو، تشخیص گفتار، سیستم‌های پیشنهاددهی، تشخیص کلاهبرداری و پردازش زبان طبیعی فراهم می‌کند. به‌طور کلی سه عملیات در مرحله بهینه‌سازی فرایند ذکر شده در بالا انجام می‌شود:

۱- تقسیم‌بندی گراف TensorRT: گراف‌های محاسباتی را برای زیر گراف‌هایی اسکن می‌کند که می‌توانند بر اساس عملیات پشتیبانی شده بهینه‌سازی شوند.

۲- لایه تبدیل: این لایه، لایه‌های شبکه پشتیبانی شده در هر زیر گراف را به لایه‌های TensorRT تبدیل می‌کند.

۳- بهینه‌سازی موتور: در نهایت، زیر گراف‌ها به موتورهای TensorRT تبدیل می‌شوند و در گراف والد شبکه جایگزین می‌شوند.

شکل (۲۱) معماری کلی انجام شده توسط tensorRT را نمایش می‌دهد



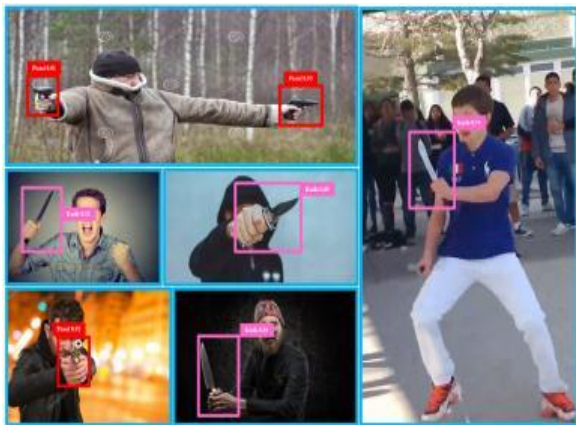
شکل ۲۱. معماری TensorRT [۲۳]

در نهایت نتایج معماری‌های موردنظر به‌وسیله یادگیری انتقالی و بعد از آموزش به‌صورت خلاصه در جدول (۳) آورده شده است.

جدول ۳. نتایج مربوط به آموزش نسخه‌های مختلف شبکه YOLO نسخه ۵ بر روی مجموعه داده

شبکه	precision	recall	mAP [۰,۵]
YOLOv5s	۹۸/۶۰	۹۹/۰۶	۹۹/۴۱
YOLOv5m	۹۹/۸۰	۹۹/۷۰	۹۹/۴۹
YOLOv5l	۹۹/۹۲	۹۹/۸۴	۹۹/۵۳
YOLOv5x	۹۹/۹۷	۹۹/۹۲	۹۹/۵۶

همچنین در شکل (۲۴) نمونه‌هایی از آزمایش سیستم پیشنهادی در تشخیص اسلحه کمری و چاقو را مشاهده می‌کنید، همان‌طور که در تصاویر نیز مشخص است، اندازه‌های اشیاء موردنظر (اسلحه کمری و چاقو) نسبت به اندازه تصویر کوچک است؛ ولی باین حال سیستم پیشنهادی ما با موفقیت توانسته است در زمینه تشخیص آن‌ها به‌دقت قابل قبولی دست پیدا کند.



شکل ۲۴. نتایج سیستم تشخیص اسلحه کمری و چاقو در دنیای واقعی

تغییرات اعمال‌شده با بهینه‌ساز tensorRT بر روی مدل آموزش دیده شده YOLO نسخه ۵ نتایج به‌صورت جدول (۴) است.

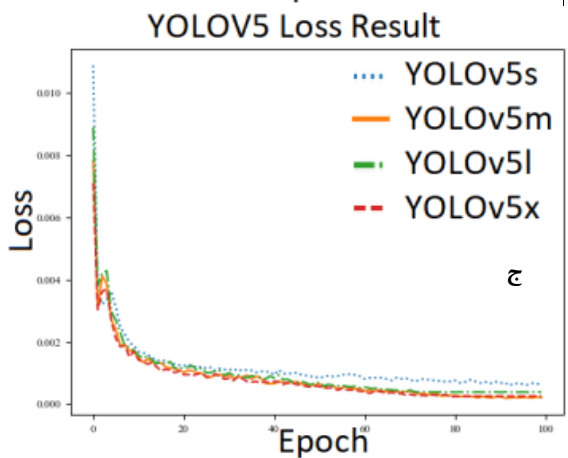
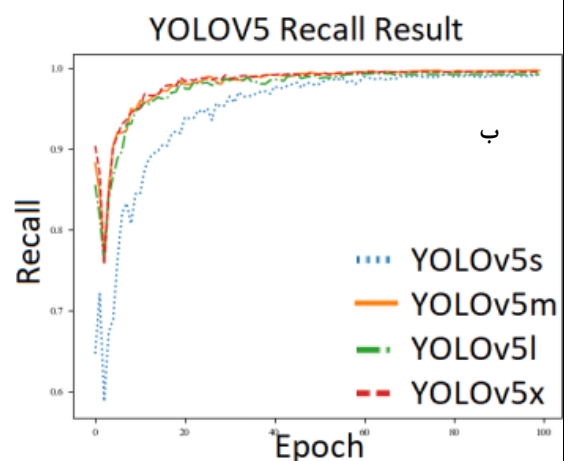
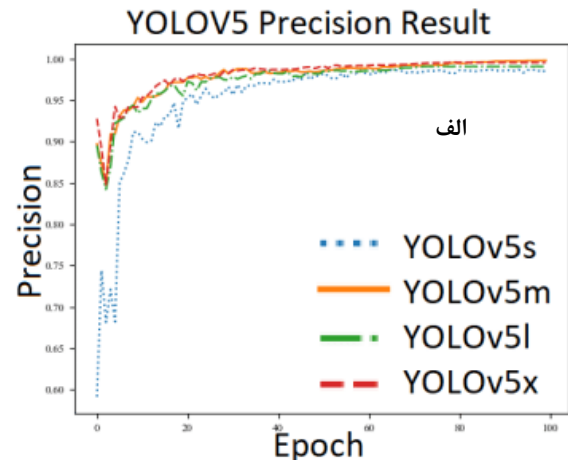
جدول ۴. بررسی سرعت اجرا سیستم با استفاده از TensorRT و بدون استفاده از آن

شبکه	بدون TensorRT	با TensorRT
YOLOv5	51ms / 16FPS	12ms / 80FPS
YOLOv5+DeepSort	25ms / 33FPS	17ms / 52FPS

۶. نتیجه‌گیری

از آنجایی که حفظ امنیت اماکن عمومی یکی از چالش‌های مهم هر جامعه امروزی است پس به‌منظور حفظ و افزایش آن و همچنین کاهش خطر، سعی شده است که کارهای زیادی صورت بگیرد. سیستم تشخیص خودکار اسلحه کمری و چاقو می‌تواند به‌صورت بلادرنگ از طریق دوربین‌های نظارتی اماکن عمومی حالت‌های

در ادامه در شکل (۲۳) اطلاعات دیگری را از چهار معماری شبکه Yolo ورژن ۵ که آموزش داده شده‌اند را مشاهده می‌کنید که شکل (۲۳-الف) نشان‌دهنده Precision چهار معماری در ۱۰۰ مرحله آموزش، شکل (۲۳-ب) دهنده Recall چهار معماری در ۱۰۰ مرحله آموزش، شکل (۲۳-ج) نشان‌دهنده Loss چهار معماری در ۱۰۰ مرحله آموزش است.



شکل ۲۳. الف) نمودار precision برای چهار معماری (ب) نمودار recall برای چهار معماری (ج) نمودار مقایسه loss برای چهار معماری

- [5] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. "You Only Look Once: Unified, Real-time Object Detection"; Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2016.
- [6] Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation"; Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2014, 580-587.
- [7] Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. "Imagenet: A Large-Scale Hierarchical Image Database"; IEEE Conf. Computer Vision and Pattern Recognition, 2009.
- [8] Egiazarov, A.; Mavroeidis, V.; Zennaro, F. M.; Vishi, K. "Firearm Detection and Segmentation Using an Ensemble of Semantic Neural Networks"; European Intelligence and Security Informatics Conference, 2019.
- [9] Wu, J. "Introduction to Convolutional Neural Networks"; National Key Lab for Novel Software Technology. Nanjing University. China, 2017.
- [10] Agarwal, N.; Sondhi, A.; Chopra, K.; Singh, G. "Transfer Learning: Survey and Classification"; Smart Innovations in Communication and Computational Sciences, Springer, 2021.
- [11] Yilmaz, A.; Javed, O.; Shah, M. "Object Tracking: A Survey"; ACM Computing Surveys 2006, 38, 13.
- [12] Veeramani, B.; Raymond, J. W.; Chanda, P. "DeepSort: Deep Convolutional Networks for Sorting Haploid Maize Seeds"; BMC bioinformatics, 2018.
- [13] Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y. Mark Liao "Yolov4: Optimal Speed and Accuracy of Object Detection"; arXiv 2020.
- [14] Russakovsky, O. "Imagenet Large Scale Visual Recognition Challenge"; Int. J. Computer Vision 2004, 10934.
- [15] Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K. Q. "Densely Connected Convolutional Networks"; Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2017.
- [16] He, K.; Zhang, X.; Ren, S.; Sun, J. "Deep Residual Learning for Image Recognition"; Proc. IEEE conf. Computer Vision and Pattern Recognition, 2016.
- [17] Lin, T.-Y. "Microsoft Coco: Common Objects in Context"; European Conf. Computer Vision, 2014.
- [18] Li, X. "Weighted Feature Pyramid Networks for Object Detection"; IEEE Int. Conf. Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking, 2019.
- [19] Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. "Path Aggregation Network for Instance Segmentation"; Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2018.
- [20] Tan, M.; Pang, R.; Le, Q. V. "Efficientdet: Scalable and Efficient Object Detection"; Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2020.
- [21] Vanholder, H. "Efficient Inference with Tensorrt"; GPU Technology Conference, 2016.
- [22] Tohidifar, A.; Dadashtabar Ahmadi, K.; Kiyae Khosh Rudbari A. K. "Detection of Lumbar Guns in Surveillance Cameras in Difficult Situations, Including Occlusion, Different Angles and Long Distances Using Deep Learning"; Information and Communication Technology in Policing 1400, 2, 9-21 (In Persian)
- [23] <https://developer.nvidia.com/tensorrt>

مختلف موجود در صحنه را مورد بررسی قرار دهد و در زمان رخ دادن حالت و حرکت‌های مشکوک همانند چاقوکشی و یا تهدید با اسلحه خیلی سریع به این موضوع واکنش نشان دهد و سریعاً پیام‌های خاص موردنیاز را به نیروهای حفاظتی و پلیس اطلاع دهد. در مجموعه داده مورد استفاده به منظور طراحی یک سیستم قوی برای تشخیص اسلحه کم‌ری و چاقو با چالش‌های مختلف موجود در تشخیص شیء روبه‌رو بوده‌ایم که برای حل این چالش سعی شده است که به اکثر چالش‌های موجود در زمینه تشخیص شیء اعم از فاصله شیء از دوربین، انسداد، زوایای مختلف و... پرداخته شود و برای اینکه سیستم عملکرد مناسبی در زمان روبه‌رو شدن با این چالش‌ها از خود نشان دهد، مجموعه داده‌ها را متناسب با آن چالش‌ها پیاده‌سازی کرده‌ایم. از آنجایی که تشخیص و شناسایی سریع اسلحه کم‌ری و چاقو برای حفظ امنیت مورد انتظار است برای حل این چالش تصمیم گرفته شد که از شبکه تک‌مرحله‌ای YOLO نسخه ۵ استفاده شد که خود شامل ۴ نسخه متفاوت است. پس از آموزش شبکه YOLO نسخه ۵ با استفاده از مجموعه داده‌ای به‌اندازه ۲۸ هزار تصویر برای دو کلاس اسلحه کم‌ری و چاقو در نسخه بزرگ‌معماری YOLO نسخه ۵ به‌دقت ۹۹/۵۶٪ (mAP) و دقت نسخه کوچک آن ۹۹/۴۱٪ (mAP) دست‌یافته‌ایم که در هر دو حالت دقت قابل قبولی است؛ ولی به‌منظور استفاده کاربردی از این سیستم پیشنهاد می‌شود که از نسخه کوچک آن یعنی از معماری YOLOv5s برای سیستم‌های عملیاتی استفاده شود. همچنین به‌منظور ره‌گیری اسلحه کم‌ری و چاقو تشخیص داده‌شده توسط الگوریتم Yolo از الگوریتم DeepSort استفاده شده است که می‌تواند با دقت مناسب اسلحه کم‌ری و چاقو را در فریم‌های ویدئو ره‌گیری کند. به‌منظور افزایش سرعت سیستم در زمان اجرا از فناوری TensorRT استفاده شد که باعث افزایش سرعت پردازش نسبت به سرعت قبلی سیستم بدون استفاده از TensorRT شده است و استفاده از این فناوری باعث اجرایی‌تر شدن سیستم پیشنهادی شده است.

۷. مراجع

- [1] Tiwari, R. K.; Verma, G. K. "A Computer Vision based Framework For Visual Gun Detection Using Harris Interest Point Detector"; Procedia Computer Science, 2015.
- [2] Lim, J.; Al Jobayer, M. I.; Baskaran, V. M.; Lim, J. M.; See, J.; Wong, K. "Deep Multi-level Feature Pyramids: Application For Non-canonical Firearm Detection in Video Surveillance"; Engineering Applications of Artificial Intelligence, 2021.
- [3] Mehta, P.; Kumar, A.; Bhattacharjee, S. "Fire and Gun Violence Based Anomaly Detection System Using Deep Neural Networks"; International Conference on Electronics and Sustainable Communication Systems, 2020, 199-204.
- [4] Warsi, A.; Abdullah, M.; Husen, M. N.; Yahya, M.; Khan, S.; Jawaid, N. "Gun Detection System Using YOLOv3"; IEEE Int. Conf. Smart Instrumentation, Measurement and Application, 2019.