مقاله پژوهشی

مجله ماشین‌بینایی و پردازش تصویر
دوره ۱۳، شماره ۱، تابستان ۱۴۰۵

تشخیص نمادهای ناهنجار در داده‌های تصویری با استفاده از مدل‌های یادگیری عمیق

مهسا سیروزیان فرد[۱]، مه لقا افراسیابی[۲]

چکیده

با گسترش روزافزون محتوای تصویری در فضای مجازی، نظارت دستی بر نمادهای ناهنجار عملاً غیرممکن شده است. در این پژوهش، یک مجموعه‌داده اختصاصی شامل ۳۶۵۹ تصویر از سه نماد «چشم جهان‌بین»، «ستاره واژگون» و «گونیا و پرگار» گردآوری و برچسب‌گذاری شد. داده‌ها به نسبت ۸۰٪ آموزش و ۲۰٪ آزمون تقسیم شدند. به‌منظور افزایش تنوع بصری، تکنیک‌های افزایش داده به مجموعه آموزش اعمال گردید. عملکرد مدل‌های Faster R-CNN، SSD300 و خانواده YOLO (نسخه‌های ۸ تا ۱۲) مورد ارزیابی قرار گرفت. نتایج نشان داد مدل Faster R-CNN در مجموعه آزمون به مقادیر Precision برابر با ۰/۹۰۱، Recall برابر با ۰/۹۰۹ و mAP50 برابر با ۰/۹۳۹ دست یافته و عملکرد بهتری نسبت به مدل‌های تک‌مرحله‌ای ارائه کرده است. یافته‌ها بیانگر آن است که مدل‌های دومرحله‌ای در مواجهه با نمادهای دارای جزئیات هندسی ظریف و شباهت ساختاری بالا، تعمیم‌پذیری بهتری دارند. این پژوهش می‌تواند مبنایی برای توسعه سامانه‌های هوشمند پایش خودکار محتوای تصویری در کاربردهای امنیت فرهنگی و پالایش رسانه‌ای باشد.

کلیدواژه‌ها

نمادهای ناهنجار، مدل تشخیص شیء، مجموعه‌داده تصویری، YOLO، SSD، Faster-RCNN

۱ مقدمه

با گسترش سریع رسانه‌های دیجیتال، شبکه‌های اجتماعی و پلتفرم‌های اشتراک‌گذاری ویدئو، حجم عظیمی از محتوای تصویری به‌صورت روزانه تولید و منتشر می‌شود. در این میان، تصاویر و ویدئوها به‌عنوان مؤثرترین بستر انتقال پیام، نقشی کلیدی در شکل‌دهی نگرش‌ها، باورها و الگوهای رفتاری ایفا می‌کنند. یکی از شیوه‌های رایج انتقال پیام‌های غیرمستقیم در رسانه‌ها، استفاده از نمادهاست؛ عناصری بصری که می‌توانند مفاهیم چندلایه و گاه پنهان را بدون تصریح مستقیم منتقل کنند. در برخی موارد استفاده گسترده و تکرارشونده از نمادهای خاص در محتوای تصویری می‌تواند آثار فرهنگی و اجتماعی قابل توجهی به همراه داشته باشد، به‌ویژه زمانی که این نمادها به‌صورت ضمنی و خارج از کانون توجه مستقیم مخاطب ارائه شوند. با وجود اهمیت این موضوع، پایش و تحلیل دستی حجم بالای محتوای تصویری عملاً امکان‌پذیر نیست و نیاز به توسعه سامانه‌های هوشمند مبتنی بر بینایی ماشین بیش از پیش احساس می‌شود. با این حال، مسئله تشخیص نمادهای خاص در تصاویر با چالش‌های فنی متعددی همراه است. نخست آنکه این نمادها اغلب در مقیاس‌های متفاوت و در ابعاد کوچک در تصویر

این مقاله در ۱۲ آبان ۱۴۰۴ دریافت و در ۲۵ تیر ۱۴۰۵ پذیرفته شد.
[۱] کارشناسی ارشد هوش مصنوعی و رباتیکز، گروه مهندسی کامپیوتر، دانشکده برق و کامپیوتر، دانشگاه صنعتی همدان
رایانامه: m.sirooziyan@stu.hut.ac.ir
[۲] استادیار، گروه مهندسی کامپیوتر، دانشکده برق و کامپیوتر، دانشگاه صنعتی همدان
رایانامه: m.afrasiabi@hut.ac.ir
نویسنده مسئول: مه لقا افراسیابی

ظاهر می‌شوند. دوم، شباهت‌های هندسی و تقارن ساختاری میان برخی نمادها می‌تواند فرآیند تمایز بین کلاس‌ها را برای مدل‌های یادگیری عمیق دشوار سازد. سوم، حضور نمادها در پس‌زمینه‌های پیچیده، شرایط نوری متنوع، چرخش‌های مختلف و حتی هم‌پوشانی چند نماد در یک تصویر، مسئله تعمیم‌پذیری مدل‌ها را با چالش مواجه می‌کند. افزون بر این، پراکندگی منابع تولید تصویر (شبکه‌های اجتماعی، بازی‌های رایانه‌ای، انیمیشن‌ها و تصاویر تولیدشده توسط مدل‌های مولد) موجب افزایش تنوع بصری و ناهمگنی داده‌ها می‌شود که آموزش پایدار مدل را دشوارتر می‌سازد.

در حوزه تشخیص شیء، مدل‌های مبتنی بر شبکه‌های عصبی کانولوشنال، به‌ویژه معماری‌های تک‌مرحله‌ای نظیر YOLO و SSD و معماری‌های دومرحله‌ای نظیر Faster R-CNN، پیشرفت‌های چشمگیری در کاربردهای مختلف از جمله نظارت تصویری، پزشکی و صنعت داشته‌اند. با این حال، عملکرد این مدل‌ها در مواجهه با الگوهای نمادین دارای جزئیات هندسی ظریف، تقارن بالا و تغییرات مقیاس شدید، نیازمند بررسی دقیق و تحلیل مقایسه‌ای است. به‌ویژه مشخص نیست که در چنین مسئله‌ای، مدل‌های سریع تک‌مرحله‌ای عملکرد بهتری ارائه می‌کنند یا مدل‌های دقیق‌تر دومرحله‌ای از نظر تعمیم‌پذیری برتری دارند.

اگرچه در سال‌های اخیر پژوهش‌های متعددی در حوزه تشخیص ناهنجاری (Anomaly Detection) در تصاویر و ویدئوها، به‌ویژه در کاربردهایی نظیر نظارت تصویری، تشخیص رفتار غیرعادی، بازرسی صنعتی و تحلیل تصاویر پزشکی انجام شده است، تمرکز این مطالعات عمدتاً بر ناهنجاری‌های آماری، رفتاری یا ساختاری در داده‌های طبیعی بوده است. در مقابل، مسئله تشخیص نمادهای خاص با بار معنایی فرهنگی یا ایدئولوژیک در بستر محتوای رسانه‌ای، به‌صورت هدفمند و مبتنی بر مجموعه‌داده ساخت‌یافته، کمتر مورد بررسی قرار گرفته است. به‌عبارت دیگر، شکاف موجود نه در اصل تشخیص ناهنجاری، بلکه در کاربرد آن برای شناسایی الگوهای نمادین خاص در داده‌های تصویری متنوع و رسانه‌ای است.

با توجه به چالش‌های مذکور، این پژوهش بر سه محور اصلی تمرکز دارد: ۱- ایجاد یک مجموعه‌داده ساخت‌یافته و برچسب‌گذاری‌شده از نمادهای منتخب با تنوع بصری بالا ۲- تحلیل مقایسه‌ای عملکرد مدل‌های تک‌مرحله‌ای و دومرحله‌ای در شرایط واقعی و متنوع تصویری و ۳- بررسی میزان تعمیم‌پذیری این مدل‌ها در مواجهه با تغییر مقیاس، چرخش و پس‌زمینه‌های پیچیده.

برای پاسخ به این چالش‌ها، ابتدا یک مجموعه‌داده تصویری شامل سه نماد «چشم جهان‌بین»، «ستاره واژگون» و «گونیا و پرگار» گردآوری و به‌صورت دستی برچسب‌گذاری شد. سپس با اعمال تکنیک‌های افزایش داده بر مجموعه آموزش، تنوع ظاهری نمونه‌ها تقویت گردید. در ادامه، چندین معماری مطرح تشخیص شیء شامل Faster R-CNN، SSD300 و نسخه‌های مختلف YOLO مورد آموزش و ارزیابی قرار گرفتند تا رفتار آن‌ها در مواجهه با نمادهای هندسی با ویژگی‌های ساختاری مشابه مورد تحلیل قرار گیرد.

نوآوری این پژوهش را می‌توان در سه جنبه خلاصه کرد:

نخست، ارائه یک مجموعه‌داده هدفمند برای مسئله تشخیص نمادهای ساختاری با تنوع منبع و شرایط تصویری؛

دوم، ارائه تحلیل مقایسه‌ای نظام‌مند میان معماری‌های تک‌مرحله‌ای و دومرحله‌ای در یک مسئله دارای شباهت هندسی بالا و سوم، بررسی تجربی تعمیم‌پذیری مدل‌ها در حضور تغییرات مقیاس، چرخش و زمینه‌های پیچیده.

نتایج حاصل از این مطالعه می‌تواند مبنایی برای توسعه سامانه‌های هوشمند پایش محتوای تصویری و همچنین بستری برای تحقیقات آینده در زمینه تشخیص الگوهای نمادین پیچیده فراهم سازد.

۲ پژوهش‌های مرتبط

تشخیص و طبقه‌بندی اشیاء از مهم‌ترین وظایف حوزه بینایی کامپیوتر است، به‌صورت کلی بدون درنظرگرفتن روش مورد استفاده برای محقق کردن این وظیفه، مراحل اصلی مرسوم برای تشخیص شیء در داده‌های تصویری شامل پنج مرحله اصلی است: پیش‌پردازش تصاویر، استخراج ویژگی، آموزش مدل مناسب، تشخیص اشیاء و پس‌پردازش پیش‌بینی‌های تولید شده توسط مدل.

در مرحله پیش‌پردازش تصویر ورودی، جهت آماده‌سازی تصویر برای مراحل بعدی انجام می‌شود. عملیات رایج در این مرحله شامل کاهش نویز، افزایش کنتراست، کاهش فضای حالت، هموارسازی تصاویر، نرمال‌سازی تصویر و تغییر اندازه تصویر می‌شود. در این مرحله اغلب توابع و فیلترهای مختلف حوزه پردازش تصویر استفاده می‌شوند. در این پژوهش از فیلترها و تکنیک‌های افزایش داده متنوعی استفاده شده که در بخش **Error! Reference source not found.** جزئیات آن شرح داده می‌شود. پس از آن مرحله استخراج ویژگی انجام می‌شود. هدف از این مرحله استخراج ویژگی‌های بصری از تصویر است که برای تشخیص اشیا به کار می‌روند، ویژگی‌های رایج شامل رنگ، شکل، بافت، و لبه‌ها می‌شود. نکته قابل توجه این است که الگوریتم‌های یادگیری عمیق، خود توانایی استخراج ویژگی از تصاویر را با استفاده از لایه‌های مختلف دارند، اما در صورت نیاز به‌صرفه

جویی در زمان یا استخراج ویژگی‌های خاص مدنظر از تصاویر می‌توان از الگوریتم‌های مختلف پیش از ورود داده‌ها به شبکه استفاده کرد. از الگوریتم‌های رایج در استخراج ویژگی می‌توان به الگوریتم DoG، HOG [۱] و شبکه‌های عصبی عمیق مانند CNNها [۲] اشاره کرد. می‌توان از تکنیک‌های مختلفی مانند تجزیه و تحلیل بافت، هیستوگرام [۳]، و تبدیلات هارمونیزه شده نیز برای استخراج ویژگی استفاده کرد.

پس از آن یک مدل یادگیری ماشین مناسب برای تشخیص اشیا انتخاب و آموزش داده می‌شود. مدل‌های رایج شامل شبکه‌های عصبی کانولوشنال[۱]، ماشین‌های بردار پشتیبان[۲] [۴] و جنگل‌های تصادفی[۳] [۵] می‌شود. در این مرحله مدل با استفاده از مجموعه‌داده‌های برچسب‌گذاری شده آموزش داده می‌شود. سپس مدل آموزش دیده برای تشخیص اشیا در تصاویر جدید استفاده می‌شود. مدل برای هر شیء احتمالی در تصویر، یک مرز تصمیم[۴] و یک برچسب (نوع شیء قرار گرفته در مرز تصمیم) ایجاد می‌کند. این فرایند معمولاً با استفاده از تشخیص‌دهنده‌های یک‌مرحله‌ای[۵] یا تشخیص‌دهنده‌های دومرحله‌ای[۶] انجام می‌شود. در مرحله بعد پس‌پردازش صورت می‌گیرد. این مرحله شامل پالایش نتایج تشخیص است، عملیات رایج در این مرحله شامل ادغام کادرهای همپوشانی، حذف اشیاء نادرست، و تصحیح برچسب‌های پیش‌بینی شده توسط مدل انتخابی می‌شود.

در ادامه این بخش، به بررسی و مرور مدل‌های رایج تشخیص شیء مبتنی بر شبکه‌های عصبی عمیق و همچنین مهم‌ترین بهبودهای ارائه‌شده با هدف افزایش دقت و کارایی این مدل‌ها پرداخته می‌شود. تمرکز این فصل بر معرفی معماری‌های پرکاربرد، نقاط قوت و ضعف آن‌ها و همچنین روش‌های بهبود عملکرد تشخیص اشیاء، به‌ویژه اشیای کوچک و پیچیده است.

۱-۲ مدل‌های تشخیص شیء مبتنی بر شبکه‌های عصبی

شبکه YOLO [۶][۷] یکی از شناخته‌شده‌ترین شبکه‌های عصبی کانولوشنال برای تشخیص اشیاء در تصاویر و ویدئوها به‌صورت بلادرنگ است. برخلاف بسیاری از الگوریتم‌های تشخیص شیء که فرآیند تشخیص را در چندین مرحله انجام می‌دهند، YOLO کل فرآیند تشخیص شیء را به‌صورت یک‌مرحله‌ای انجام می‌دهد. این ویژگی باعث افزایش قابل توجه سرعت پردازش شده و YOLO را به گزینه‌ای مناسب برای کاربردهای بلادرنگ نظیر رباتیک، بینایی ماشین و سیستم‌های نظارتی تبدیل کرده است.

در این الگوریتم، تصویر ورودی ابتدا به شبکه‌ای از نواحی هم‌اندازه (grid) تقسیم می‌شود. سپس، برای هر ناحیه، شبکه ویژگی‌هایی نظیر رنگ، شکل و بافت را استخراج کرده و احتمال وجود هر کلاس از پیش تعریف‌شده به همراه مختصات جعبه محاط‌کننده پیش‌بینی می‌شود. در مرحله پس‌پردازش، نواحی تکراری یا نادرست با استفاده از روش‌هایی مانند سرکوب غیرحداکثری[۸] حذف می‌گردند. YOLO با وجود ساختار نسبتاً ساده، از دقت بالایی برخوردار بوده و در مقایسه با بسیاری از روش‌های دیگر، سرعت بسیار بیشتری دارد.

خانواده مدل‌های YOLO از پرکاربردترین مدل‌ها در حوزه بینایی کامپیوتر محسوب می‌شوند. نسخه اولیه این شبکه نخستین بار در سال ۲۰۱۵ توسط Joseph Redmon ارائه شد که با زبان C و در چارچوب یادگیری عمیق Darknet پیاده‌سازی شده بود [۷]. پس از آن، نسخه YOLOv5 توسط شرکت Ultralytics توسعه یافت و زمینه‌ساز ارائه نسخه‌های متعددی نظیر Scaled-YOLOv4، YOLOR و YOLOv7 شد. از مهم‌ترین مزایای این مدل‌ها می‌توان به دقت بالا در کنار حجم نسبتاً کم مدل و قابلیت اجرا بر روی یک GPU به‌صورت مستقل اشاره کرد.

مدل YOLOv8 به‌عنوان نسخه‌ای بهبودیافته از YOLOv5، با اصلاحات ساختاری در بخش‌های مختلف، به‌ویژه در زمینه طراحی بدون لنگر[۹] معرفی شد. در این مدل، به‌جای استفاده از جعبه‌های لنگر[۱۰] از پیش تعریف‌شده، مرکز اشیاء به‌صورت مستقیم پیش‌بینی می‌شود. این رویکرد منجر به کاهش تعداد جعبه‌های پیش‌بینی‌شده، کاهش پیچیدگی محاسباتی و حذف وابستگی به تنظیم دستی لنگرها شده و در نتیجه، عملکرد تشخیص اشیاء بهبود یافته است. YOLOv8 در معیارهای سنجشی[۱۱] نظیر Microsoft COCO و Roboflow100 نتایج قابل توجهی ارائه کرده است.

مدل YOLOv9 در فوریه ۲۰۲۴ معرفی شد و با ارائه مفاهیمی نظیر PGI و GELAN به بهبود کارایی محاسباتی و کاهش از دست رفتن اطلاعات پرداخت [۸]. این معماری با بهره‌گیری از اصل «گلوگاه اطلاعاتی[۱۲]» و استفاده از روش‌های نوین آدرس‌دهی اطلاعات، توانسته است دقت و سرعت بالایی را در کاربردهای

---

۱ CNN (Convolutional Neural Networks)
۲ SVM (Support Vector Machine)
۳ Random forest
۴ Bounding Box
۵ One-Stage Detectors
۶ Two-Stage Detectors
۷ You Only Look Once
۸ Non-Maximum Suppression
۹ anchor-free
۱۰ anchor box
۱۱ Benchmark
۱۲ Information Bottleneck principle

بلادرنگ تضمین کند. به‌کارگیری GELAN[۱] موجب بهینه‌سازی استفاده از پارامترها و افزایش سازگاری معماری در سناریوهای مختلف شده است.

در ادامه، YOLOv10 در ماه مه ۲۰۲۴ معرفی شد که تمرکز اصلی آن بر حذف وابستگی به مرحله سرکوب غیرحداکثری و کاهش افزونگی محاسباتی بود [۹]. این مدل با معرفی مکانیزم تخصیص دوگانه سازگار، امکان آموزش و استنتاج بدون سرکوب غیرحداکثری را فراهم کرده و در نتیجه، تأخیر در استنتاج به‌طور قابل توجهی کاهش یافته است. همچنین، به‌منظور کاهش بار محاسباتی، از یک طبقه‌بند سبک در معماری استفاده شده است.

پس از آن، YOLOv11 با معرفی بلوک C3k2[۲] و ماژول‌های SPFF[۳] و C2PSA[۴] ارائه شد که موجب بهبود فرآیند استخراج ویژگی و افزایش دقت نهایی مدل گردید [۱۰]. جدیدترین عضو این خانواده، YOLOv12 است که با تمرکز بر ادغام مکانیزم‌های توجه کارآمد[۵] در بخش ستون فقرات مدل[۶] معرفی شده است. این مدل با بهره‌گیری از ماژول توجه ناحیه[۷]، پیچیدگی محاسباتی خودتوجه[۸] را کاهش داده و در عین حال، مزایای مدل‌سازی زمینه سراسری را حفظ می‌کند [۱۱،۱۲]. استفاده از R-ELAN، اتصالات باقیمانده مقیاس‌بندی‌شده و کاهش نسبت MLP، باعث افزایش پایداری آموزش و بهبود عملکرد، به‌ویژه در مدل‌های بزرگ، شده است.

شبکه SSD[۹][۱۳] نیز یکی دیگر از معماری‌های تک‌مرحله‌ای تشخیص شیء است که برای کاربردهای بلادرنگ توسعه یافته است. در این مدل، پس از استخراج ویژگی‌ها توسط لایه‌های کانولوشنی، مختصات جعبه‌های محاطکننده و احتمال وجود کلاس‌ها به‌طور مستقیم پیش‌بینی می‌شوند. در مرحله نهایی، با استفاده از سرکوب غیرحداکثری، تشخیص‌های تکراری حذف می‌گردند. با وجود سرعت بالا، SSD مانند YOLO در تشخیص اشیای بسیار کوچک یا دارای همپوشانی، با چالش‌هایی مواجه است.

در مقابل، خانواده شبکه‌های R-CNN شامل R-CNN، Fast R-CNN، Faster R-CNN و Mask R-CNN، جزء روش‌های دومرحله‌ای محسوب می‌شوند که بر دقت بالا تمرکز دارند. در این روش‌ها ابتدا نواحی کاندید حاوی شیء شناسایی شده و سپس ویژگی‌ها توسط شبکه کانولوشنال استخراج و برای طبقه‌بندی و پالایش جعبه‌های محاطکننده استفاده می‌شوند. Faster R-CNN با معرفی شبکه RPN فرآیند پیشنهاد ناحیه را تسریع کرد، اما همچنان با سرعت پایین‌تری نسبت به روش‌های تک‌مرحله‌ای (حدود ۷ فریم بر ثانیه) عمل می‌کند [۱۴،۱۵]. در مقابل، SSD با حذف مرحله پیشنهاد ناحیه، سرعت تشخیص را افزایش داده و با استفاده از ویژگی‌های چندمقیاسی و جعبه‌های محصورکننده[۱۰] پیش‌فرض، کاهش دقت را تا حد زیادی جبران کرده است.

۲-۲ بهبودهای ارایه شده در تشخیص اشیاء

برای افزایش عملکرد مدل‌های تشخیص اشیاء مبتنی بر یادگیری عمیق، روش‌های متنوعی استفاده می‌شود که عمدتاً شامل یادگیری ویژگی‌های چندمقیاسی[۱۱] [۱۶]، افزایش داده[۱۲] [۱۷]، استراتژی‌های حین آموزش، تشخیص مبتنی بر زمینه[۱۳] [۱۸] و تشخیص مبتنی بر GAN[۱۴] [۱۹] هستند.

تشخیص اشیاء در مقیاس‌های مختلف اهمیت بالایی دارد، به‌ویژه برای اشیای کوچک. برای این منظور، تکنیک‌هایی مانند هرم‌های تصویری ویژگی محور، شبکه‌های هرم ویژگی و ادغام چندمقیاسی به‌کار گرفته می‌شوند. همچنین افزایش داده با تبدیلاتی مانند برش، چرخش و مقیاس‌بندی، تعداد نمونه‌ها و عملکرد مدل را بهبود می‌بخشد. روش‌های پیشرفته‌تری مانند SNIP[۱۵] [۲۰] و SNIPER [۲۱] نیز برای آموزش مؤثر چندمقیاسی و تمرکز روی نواحی مهم ارائه شده‌اند.

اطلاعات زمینه‌ای نقش مهمی در شناسایی اشیاء دارند، به‌ویژه برای اشیای کوچک یا با ویژگی‌های محدود. روش‌هایی مانند MRCNN[۱۶] [۲۲]، MPNet[۱۷] و GBDNet [۲۳] با استخراج ویژگی از نواحی زمینه‌ای محلی و چندمقیاسی، عملکرد تشخیص را بهبود می‌دهند. همچنین استفاده از زمینه سراسری و ماژول‌های سراسری [۲۴] به مدل کمک می‌کند تا با بهره‌گیری از کل تصویر، دقت تشخیص افزایش یابد.

---

۱ Generalized Efficient Layer Aggregation Network
۲ Cross Stage Partial with kernel size 2
۳ Spatioal Pytamid Pooling-Fast
۴ Convolutional block with Parallel Spatial Attention
۵ Efficient Attention
۶ Backbone
۷ Area Attention
۸ Self-attention
۹ Single Shot Multibox Detector
۱۰ bounding box
۱۱ Multi-scale Feature Learning
۱۲ Data Augmentation
۱۳ Context-Based Detection
۱۴ Generative Adversarial Networks
۱۵ Scale Normalization for Image Pyramids
۱۶ multi-region CNN
۱۷ multipath network

تکنیک‌های پیش‌پردازش مانند افزایش وضوح تصویر، عملیات Tiling و یادگیری خودکار لنگرها نیز برای بهبود تشخیص اشیای کوچک و مقیاس‌های متفاوت پیشنهاد شده‌اند.

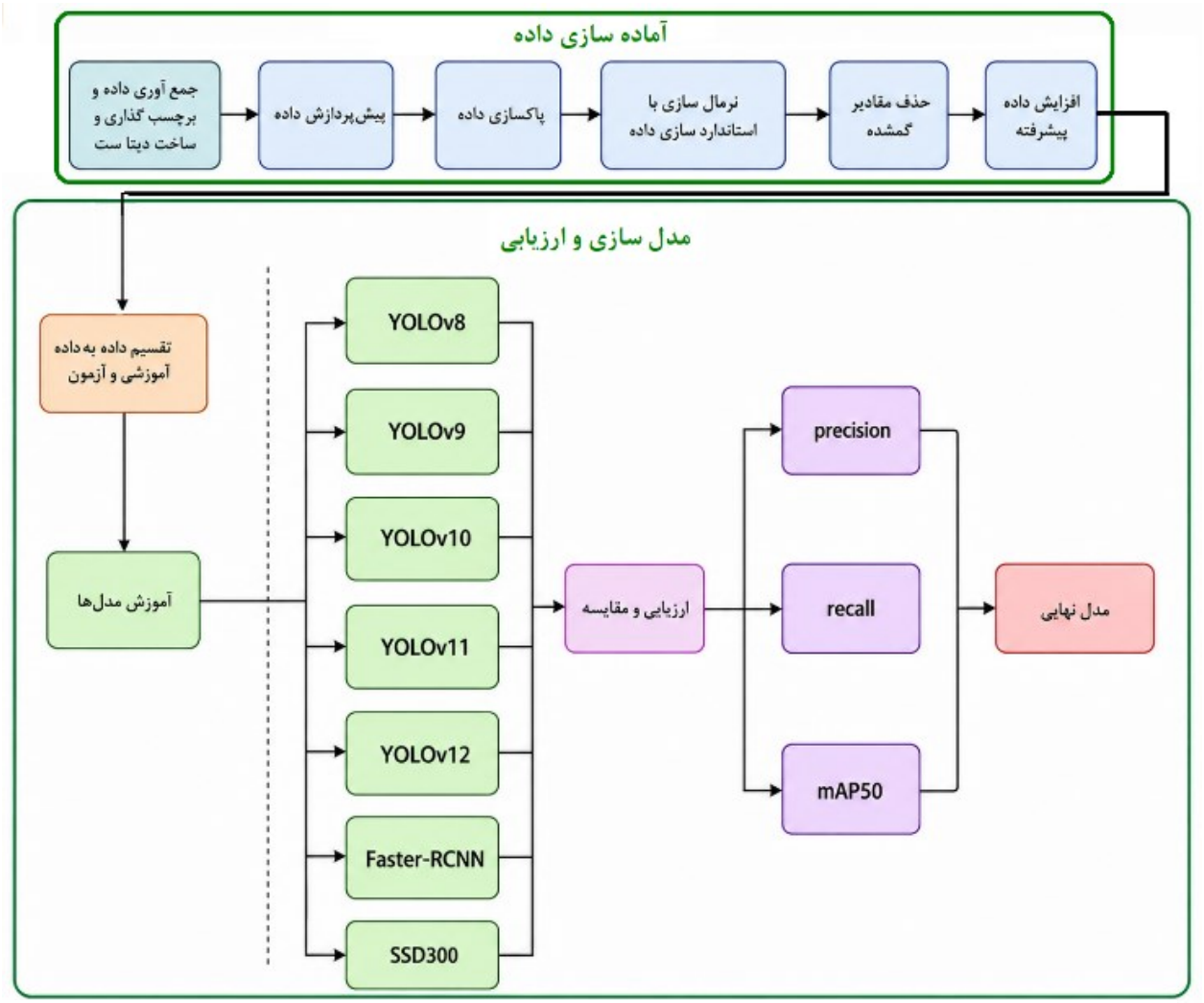
بر اساس مرور انجام‌شده، مدل‌های تک‌مرحله‌ای نظیر YOLO و SSD از مزیت سرعت بالا و ساختار سبک برخوردار بوده و برای کاربردهای بلادرنگ مناسب هستند. با این حال، در سناریوهایی که اشیاء دارای ابعاد کوچک، جزئیات هندسی ظریف یا شباهت ساختاری بالا باشند، این معماری‌ها ممکن است با افت دقت و کاهش توان تفکیک کلاس‌ها مواجه شوند. در مقابل، مدل‌های دومرحله‌ای نظیر Faster R-CNN به دلیل بهره‌گیری از سازوکار پیشنهاد ناحیه (Region Proposal Network) و استخراج ویژگی عمیق‌تر در مرحله دوم، معمولاً در تشخیص اشیاء کوچک و پیچیده عملکرد دقیق‌تری ارائه می‌دهند، هرچند با هزینه محاسباتی بالاتر همراه هستند.

با توجه به ماهیت نمادهای مورد بررسی در این پژوهش که شامل الگوهای هندسی با تقارن بالا، تغییر مقیاس، چرخش و حضور در پس‌زمینه‌های پیچیده هستند، انتظار می‌رود معماری‌های دارای دقت مکانی بالاتر و توان تفکیک ساختاری قوی‌تر عملکرد مناسب‌تری داشته باشند. ازاین‌رو، در این پژوهش مدل Faster R-CNN به‌عنوان روش پایه جهت تحلیل عمیق‌تر انتخاب شده و عملکرد آن در کنار معماری‌های تک‌مرحله‌ای به‌صورت تجربی مورد ارزیابی قرار گرفته است. این رویکرد امکان مقایسه نظام‌مند میان سرعت و دقت را فراهم ساخته و مبنای انتخاب نهایی مدل پیشنهادی را شکل می‌دهد.

در این پژوهش، مجموعه‌داده‌ای شامل سه کلاس نماد ناهنجار ایجاد شد و مدل‌های SSD [۱۳]، Faster-RCNN [۲۵]، YOLOv8 [۲۶]، YOLOv9 [۲۷]، YOLOV10 [۲۸]، YOLOv11 [۱۰] و YOLOv12 [۱۲] بر روی آن ارزیابی شدند. نتایج نشان داد بهترین عملکرد متعلق به Faster-RCNN است که با هایپرپارامترهای بهینه برای تشخیص نمادهای ناهنجار انتخاب شد.

۳ روش پیشنهادی

در این مقاله، ابتدا سه کلاس نماد تصویری به‌منظور ایجاد مجموعه‌داده انتخاب شد. سپس با انجام فرایندهای پیش‌پردازش



شکل (۱): روند انجام پژوهش

تصاویر و به‌کارگیری تکنیک‌های افزایش داده، کیفیت و تنوع داده‌ها افزایش یافت. در ادامه، مدل‌های مختلف تشخیص شیء بر روی این مجموعه‌داده آموزش داده شده و عملکرد آن‌ها در شناسایی نمادهای انتخاب‌شده مورد ارزیابی و مقایسه قرار گرفت. جزئیات مربوط به هر یک از این مراحل در این بخش تشریح می‌شود.

بطور کلی روش پیشنهادی در نمودار شکل (۱) آمده است.

۳-۱ نمادهای‌ناهنجار انتخاب شده

در بخش پیشین، روش‌های تشخیص شیء در داده‌های تصویری و مدل‌های پرکاربرد این حوزه مرور شد. با توجه به هدف این پژوهش، مجموعه‌داده‌ای شامل تصاویری حاوی الگوهای نمادین خاص گردآوری شد. پس از بررسی منابع موجود و با در نظر گرفتن میزان تکرار، تنوع بصری و چالش‌های تشخیص، سه نماد تصویری

به‌عنوان کلاس‌های اصلی مجموعه‌داده انتخاب شدند: ستاره واژگون[۱] [۲۹]، گونیا و پرگار[۲] [۳۰]، چشم جهان‌بین[۳] [۳۱]. این نمادها از نظر ساختار هندسی، تقارن، پیچیدگی بصری و نحوه قرارگیری در تصاویر، ویژگی‌های متفاوتی دارند که آن‌ها را برای ارزیابی مدل‌های تشخیص شیء مناسب می‌سازد. در ادامه، هر یک از این نمادها از منظر بصری معرفی می‌شود.

۳-۱-۱ نماد ستاره واژگون

ستاره واژگون یا پنتاگرام در برخی منابع به‌عنوان یکی از نمادهای مرتبط با شیطان‌گرایی معرفی شده است. پنتاگرام شیطانی معمولاً به‌صورت یک ستاره پنج‌پر وارونه با یک رأس رو به پایین و گاه همراه با تصویر سر بز نمایش داده می‌شود. این نوع کاربرد، که به نیمه دوم قرن بیستم بازمی‌گردد، از شناخته‌شده‌ترین و در عین حال بحث‌برانگیزترین تفاسیر پنتاگرام محسوب می‌شود [۲۹].

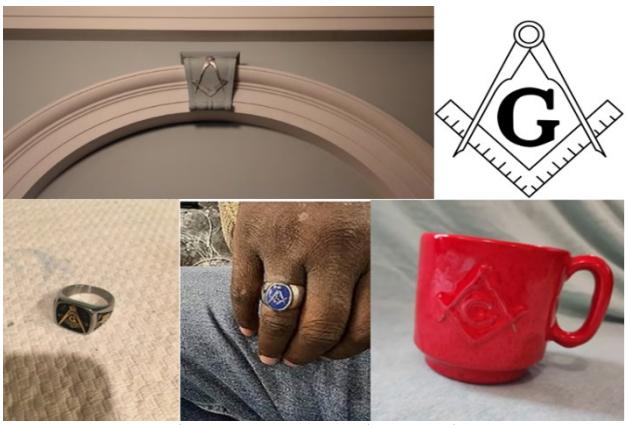
گسترش این نماد با شکل‌گیری شیطان‌پرستی مدرن و تأسیس کلیسای شیطان توسط «آنتوان لاوی» همراه بوده است. در این چارچوب، پنتاگرام شیطانی اغلب به‌عنوان نمادی از جادوی سیاه و برتری گرایش‌های مادی بر باورهای دینی تفسیر می‌شود [۳۲]. نماد «سیگیل بافومت» به‌عنوان نشان رسمی کلیسای شیطان شناخته می‌شود که نخستین بار در سال ۱۹۶۸ معرفی و در سال‌های بعد به‌طور گسترده در محصولات فرهنگی مختلف مورد استفاده قرار گرفته است [۳۳]. شکل (۲) نمونه‌هایی از این نماد را نشان می‌دهد.

شکل (۲) : نماد ستاره واژگون و نمونه‌ای از قرارگیری نماد در انیمیشن‌ها

۳-۱-۲ نماد گونیا و پرگار

نماد گونیا و پرگار اصلی‌ترین نشان فراماسونری است و از دو ابزار معماری که به‌صورت نمادین در آیین‌های ماسونی استفاده می‌شوند، تشکیل شده است. این نماد در مواردی با تغییرات جزئی یا در قالب‌های متفاوت، توسط نهادها و سازمان‌های غیرماسونی در حوزه‌های فرهنگی، آموزشی و تجاری مورد استفاده قرار گرفته است [۳۰].

با این حال، هدفمند بودن یا تصادفی بودن این کاربردها همواره محل بحث بوده است. حضور این نماد در تصاویر و اشیایی که انتظار مشاهده آن‌ها وجود ندارد، توجه پژوهشگران را به خود جلب کرده است. نمونه‌هایی از این موارد در شکل (۳) ارائه شده‌اند.



شکل (۳) : نماد گونیا و پرگار و نمونه‌ای از قرارگیری نماد در اشیای مختلف

۳-۱-۳ نماد چشم جهان‌بین

چشم جهان‌بین (چشم ایزدی، Eye of providence یا All seeing eye) نمادی است که معمولاً به‌صورت یک چشم درون یک مثلث نمایش داده می‌شود و امروزه غالباً با فراماسونری مرتبط دانسته

---

۱ Pentagram

۲ Square and Compass

۳ All-Seeing Eye

می‌شود. نخستین استفاده مستند از این نماد در پیکرنگاری فراماسونی به سال ۱۷۹۷ بازمی‌گردد.
ریشه‌های این نماد را می‌توان در نظام‌های نمادین تمدن‌های مختلف، از جمله اسطوره‌شناسی مصر باستان (چشم حوروس)، آیین بودایی و پیکرنگاری مسیحی در قرون وسطی و رنسانس مشاهده کرد. در تصویرسازی‌های قرن هفدهم، عناصر بصری مانند پرتوهای نور پیرامون این نماد دیده می‌شوند [۳۱]. در برخی نمونه‌ها، چشم جهان‌بین در ترکیب با نماد گونیا و پرگار نیز به کار رفته است. نمونه‌هایی از این کاربردها در شکل (۴) نمایش داده شده‌اند.

شکل (۴) : نماد چشم جهان‌بین و نمونه‌ای از قرارگیری نماد در بازی‌های رایانه‌ای

۳-۲ مشخصات مجموعه‌داده

با توجه به سه نماد انتخاب‌شده (ستاره واژگون، گونیا و پرگار، و چشم جهان‌بین)، مجموعه تصاویر مربوط به هر نماد از منابع در دسترس گردآوری و برچسب‌گذاری آن‌ها به‌صورت دستی انجام شد. در مجموع، این مجموعه‌داده شامل ۳۶۵۹ تصویر است که از این تعداد، ۱۰۲۴ تصویر به کلاس «چشم جهان‌بین»، ۱۴۹۳ تصویر به نماد «گونیا و پرگار» و ۱۱۴۲ تصویر به نماد «ستاره واژگون» اختصاص دارد.
در برخی تصاویر، بیش از یک نماد به‌صورت هم‌زمان حضور دارد؛ ازاین‌رو، تعداد نمونه‌های برچسب‌گذاری‌شده در هر کلاس حدود ۲۰۰۰ نمونه است. تصاویر این مجموعه‌داده از منابع متنوعی شامل شبکه‌های اجتماعی، بازی‌های رایانه‌ای، انیمیشن‌ها و فیلم‌ها، وب‌سایت‌های فروشگاهی، تصاویر تولیدشده توسط مدل‌های مولد تصویر مبتنی بر هوش مصنوعی، منابع تاریخی و وب‌سایت‌های اینترنتی مختلف گردآوری شده‌اند تا تنوع مناسبی از حالات ظاهری نمادها در شرایط مختلف تصویری به‌طور متعادل در هر کلاس پوشش داده شود.
در فرآیند تقسیم داده‌ها، ۸۰٪ از تصاویر به‌صورت تصادفی برای آموزش و ۲۰٪ برای آزمون در نظر گرفته شد. همچنین، از مجموعه آموزش، ۲۰٪ از داده‌ها به‌صورت تصادفی جهت اعتبارسنجی[۱] انتخاب شدند. لازم به ذکر است که تقسیم‌بندی داده‌ها به مجموعه آموزش، اعتبارسنجی و آزمون پیش از اعمال هرگونه افزایش داده انجام شد و مجموعه آزمون به‌صورت کامل از فرآیند افزایش داده مستثنی گردید.

۳-۳ پیش‌پردازش و آماده‌سازی داده‌ها

پیش از مرحله افزایش داده و آموزش مدل‌ها، مجموعه تصاویر گردآوری‌شده تحت فرایند پیش‌پردازش و پاک‌سازی قرار گرفت. در این مرحله، تصاویر تکراری، تصاویر دارای کیفیت بسیار پایین و نمونه‌های دارای برچسب‌گذاری نادرست حذف شدند، تا کیفیت و سازگاری داده‌ها افزایش یابد. همچنین تمامی تصاویر از نظر قالب ذخیره‌سازی و ساختار داده‌ها یکسان‌سازی شدند.
در ادامه، تصاویر متناسب با ابعاد ورودی موردنیاز مدل‌ها تغییر اندازه داده شدند و مقادیر پیکسل‌ها در بازه استاندارد نرمال‌سازی گردیدند تا اثر تفاوت‌های ناشی از شرایط تصویربرداری کاهش یابد. لازم به ذکر است که به دلیل ماهیت داده‌های تصویری مورد استفاده، داده گمشده‌ای در مجموعه تصاویر وجود نداشت؛ بنابراین مرحله حذف داده‌های گمشده صرفاً به بررسی و حذف تصاویر ناقص یا غیرقابل استفاده محدود شد.
این مراحل با هدف کاهش نویز، افزایش یکنواختی داده‌های ورودی و فراهم‌سازی شرایط مناسب برای آموزش پایدار مدل‌های تشخیص شیء انجام شد.

۳-۴ افزایش داده

برای آموزش مؤثر مدل‌های عمیق، وجود تعداد کافی نمونه برای هر کلاس ضروری است. منظور برآورده کردن به این نیاز، از تصاویر اولیه مجموعه‌داده، نمونه‌های جدیدی با استفاده از تکنیک‌های افزایش داده تولید شد. این تبدیلات شامل تغییر کنتراست و روشنایی، افزودن نویز رنگی و فلفل‌نمکی، برش، چرخش و برگرداندن تصویر، اعمال فیلترهای گوسی و تارکننده، کشیدگی تصویر و افزایش داده ویژه برای اشیای مقیاس کوچک بود. پس از جداسازی اولیه داده‌ها، تکنیک‌های افزایش داده تنها بر روی مجموعه آموزش اعمال شد. در نتیجه، تعداد نمونه‌های آموزشی به ۴۷۳,۸۳۷ تصویر افزایش یافت، در حالی که مجموعه آزمون بدون هیچ‌گونه افزایش داده و با همان تصاویر خام اولیه برای ارزیابی نهایی مورد استفاده قرار گرفت. با توجه به هزینه محاسباتی بالای آموزش مدل‌های تشخیص عمیق و وجود یک مجموعه آزمون کاملاً مستقل، از اعتبارسنجی متقاطع استفاده نشد؛ در عوض، جداسازی سخت‌گیرانه میان داده‌های آموزش، اعتبارسنجی و آزمون به‌منظور کاهش بیش‌برازش اعمال گردید.
در خصوص کفایت تعداد نمونه‌های اولیه، باید توجه داشت که در بسیاری از مسائل تشخیص شیء با دامنه کاربردی خاص، مجموعه‌داده‌های سفارشی در مقیاس چند هزار تصویر نقطه شروع

[۱] Validation

متداولی محسوب می‌شوند، مشروط بر آنکه تنوع بصری مناسب و برچسب‌گذاری دقیق وجود داشته باشد. در این پژوهش، ۳۶۵۹ تصویر خام از منابع متنوع گردآوری شد که شامل تغییرات مقیاس، زاویه دید، شرایط نوری و پس‌زمینه‌های مختلف بودند. افزایش تعداد داده‌ها در این مطالعه صرفاً مبتنی بر تبدیلات روتین و ساده نظیر چرخش یا تغییر روشنایی نبوده است، بلکه از تکنیک‌های پیشرفته‌تری استفاده شده است که تنوع ساختاری نمونه‌ها را به‌طور معنادار افزایش می‌دهند. از جمله این روش‌ها می‌توان به تکرار تصادفی نواحی دارای برچسب (نمادها) در موقعیت‌های مختلف تصویر، ایجاد هم‌پوشانی مصنوعی بین چند نماد، برش‌های موضعی شدید[1] همراه با تغییرات مقیاس، افزودن نویزهای ساختاری و اغتشاشات رنگی شدید، و همچنین استفاده از روش افزایش داده موزائیک[2] اشاره کرد. در روش موزائیک، چند تصویر به‌صورت ترکیبی در یک قاب جدید ادغام می‌شوند که این امر موجب ایجاد تنوع مکانی، مقیاسی و زمینه‌ای قابل توجه در داده‌های آموزشی می‌شود.

به دلیل ترکیب چندین تکنیک افزایش داده و اعمال آن‌ها به‌صورت تصادفی و چندمرحله‌ای، تعداد نمونه‌های آموزشی به‌صورت تصاعدی افزایش یافته و به بیش از ۴۷۰ هزار تصویر رسیده است. لازم به ذکر است که این فرآیند تنها بر مجموعه آموزش اعمال شده و مجموعه آزمون بدون هیچ‌گونه افزایش داده برای ارزیابی نهایی مورد استفاده قرار گرفته است. هدف از این رویکرد، افزایش توان تعمیم مدل در مواجهه با تغییرات شدید تصویری و جلوگیری از بیش‌برازش بر روی مجموعه اولیه بوده است.

۳-۵ معیارهای ارزیابی

در این پژوهش، برای ارزیابی عملکرد مدل‌های تشخیص شیء از معیارهای Precision، Recall و mAP@0.5 استفاده شده است. معیار Precision بیانگر نسبت پیش‌بینی‌های صحیح به کل پیش‌بینی‌های انجام‌شده توسط مدل بوده و مقادیر بالاتر آن نشان‌دهنده کاهش خطای مثبت کاذب و افزایش اطمینان مدل در تشخیص اشیاء است. معیار Recall نیز نشان می‌دهد چه نسبتی از اشیای واقعی موجود در تصاویر توسط مدل به‌درستی شناسایی شده‌اند.

معیار Average Precision (AP) بیانگر مساحت زیر منحنی Precision–Recall برای هر کلاس است. برای محاسبه AP، ابتدا با تغییر آستانه اطمینان (Confidence Threshold)، مقادیر مختلفی از Precision و Recall محاسبه شده و منحنی Precision–Recall ترسیم می‌شود. سپس مقدار AP به‌صورت انتگرال عددی یا تقریب گسسته از مساحت زیر این منحنی به‌دست می‌آید. این معیار به‌طور هم‌زمان توانایی مدل در شناسایی صحیح نمونه‌های مثبت و کنترل خطاهای پیش‌بینی را ارزیابی می‌کند. هرچه مقدار AP به ۱ نزدیک‌تر باشد، عملکرد مدل در آن کلاس بهتر است.

معیار mean Average Precision (mAP) نیز میانگین مقادیر AP در تمامی کلاس‌های مسئله است و یکی از مهم‌ترین شاخص‌های ارزیابی در مسائل تشخیص شیء محسوب می‌شود. در این پژوهش از mAP@0.5 استفاده شده است؛ به این معنا که تنها پیش‌بینی‌هایی صحیح در نظر گرفته می‌شوند که میزان همپوشانی (IoU)بین جعبه پیش‌بینی‌شده و جعبه واقعی برابر یا بزرگ‌تر از ۰/۵ باشد.

در روابط زیر، TP بیانگر تعداد تشخیص‌های صحیح، FP تعداد تشخیص‌های نادرست، FN تعداد نمونه‌های واقعی تشخیص‌داده‌نشده و IoU میزان همپوشانی میان جعبه واقعی و جعبه پیش‌بینی‌شده است:

$$Precision = \frac{TP}{TP + FP} \quad (۱)$$

$$Recall = \frac{TP}{TP + FN} \quad (۲)$$

$$AP = \int_0^1 Precision(Recall)\, dRecall \quad (۳)$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (۴)$$

۳-۶ مدل‌های آموزش دیده

برای آموزش روی مجموعه‌داده ایجاد شده و مقایسه عملکرد از مدل‌های پرکاربرد حوزه تشخیص شیء YOLOv8، YOLOv9، YOLOv10، YOLOv11، YOLOv12، Faster-RCNN و SSD300 استفاده کرده‌ایم. برای حفظ انصاف مقایسه، تنظیمات پایه آموزش برای همه مدل‌ها یکسان در نظر گرفته شد و تنها تفاوت‌ها ناشی از معماری ذاتی مدل‌هاست.

همان‌طور که در بخش مقدمه اشاره شد، هر کدام از این معماری‌ها دارای ویژگی‌ها و مزایای خاص خود هستند که هریک می‌تواند در تشخیص شیء بسته به خاصیت داده‌ها، عملکرد قابل‌قبولی را نتیجه دهد. از آنجایی که مدل‌های تشخیص دومرحله‌ای دارای دقت بالاتر اما محاسبات سنگین‌تر و کندتر هستند و آموزش طولانی‌تری دارند، این دسته از مدل‌ها را با تعداد دوره آموزشی کمتری نسبت به مدل‌های یک‌مرحله‌ای آموزش دادیم. تمامی فرایند آموزش و تست مدل‌ها با کارت گرافیکی

[1] Random cropping
[2] Mosaic Augmentation

RTX 4090 صورت پذیرفته‌اند. ابتدا ۸۰ درصد داده‌ها جهت آموزش به‌صورت تصادفی انتخاب و ۲۰ درصد برای تست مدل‌ها جداسازی شد و از مجموعه آموزش نیز ۲۰ درصد جهت ارزیابی به‌صورت تصادفی انتخاب شده است. به‌منظور بررسی حساسیت مدل‌ها نسبت به تنظیمات آموزشی، تحلیل محدودی بر روی برخی هایپرپارامترهای کلیدی انجام شد. در این مطالعه، نرخ یادگیری اولیه برابر با ۰/۰۰۱ و اندازه دسته آموزشی برابر با ۱۶ در نظر گرفته شد که بر اساس آزمایش‌های اولیه منجر به همگرایی پایدار گردید. افزایش نرخ یادگیری موجب نوسان در منحنی آموزش و کاهش پایداری مدل شد، در حالی که کاهش بیش از حد آن باعث کندی همگرایی گردید. همچنین تعداد دوره آموزشی برای مدل‌های تک‌مرحله‌ای بیشتر از مدل دومرحله‌ای تنظیم شد، زیرا Faster R-CNN با وجود تعداد دوره آموزشی کمتر، همگرایی سریع‌تری نشان داد. آزمایش‌های تکمیلی نشان داد افزایش بیش از حد دوره آموزشی در مدل‌های YOLO منجر به کاهش فاصله train-test و نشانه‌هایی از بیش‌برازش می‌شود، در حالی که Faster R-CNN پایداری بیشتری در تعمیم‌پذیری نشان داد. این نتایج نشان می‌دهد عملکرد مدل‌ها نه‌تنها به معماری، بلکه به تنظیم دقیق پارامترهای آموزشی نیز وابسته است و انتخاب بهینه هایپرپارامترها نقش مهمی در دستیابی به تعادل میان دقت و تعمیم‌پذیری دارد.

نتایج آموزشی این مدل‌ها روی مجموعه‌داده ساخته شده در جدول (۱) آمده است. لازم به توضیح است که مقادیر Precision، mAP50 و Recall گزارش‌شده در فاز آموزش، صرفاً به‌منظور تحلیل رفتار یادگیری و بررسی همگرایی مدل‌ها ارائه شده‌اند و مبنای اصلی ارزیابی و مقایسه عملکرد، نتایج حاصل از مجموعه آزمون مستقل بوده است.

جدول (۱) : مقایسه معیارهای عملکرد مدل‌ها در فاز آموزش (جهت تحلیل یادگیری) و آزمون (جهت ارزیابی نهایی)

مدل	فاز	Precision	Recall	mAP50
YOLOv8n	آموزش	۰,۹۶۸	۰,۹۲۵	۰,۹۶۱
	تست	۰,۸۸۴	۰,۷۷۲	۰,۸۳۴
YOLOv9n	آموزش	۰,۹۸۷	۰,۹۶۷	۰,۹۸۵
	تست	۰,۸۸۶	۰,۷۷۵	۰,۸۳۶
YOLOv10n	آموزش	۰,۹۶۸	۰,۹۲۵	۰,۹۶۵
	تست	۰,۸۹۵	۰,۷۷۵	۰,۸۴
YOLOv11n	آموزش	۰,۹۶۱	۰,۹۱۲	۰,۹۵۳
	تست	۰,۸۹	۰,۷۸	۰,۸۳۹
YOLOv12n	آموزش	۰,۹۷۲	۰,۹۲۳	۰,۹۶۳
	تست	۰,۸۸۶	۰,۷۸۹	۰,۸۳۷
Faster-RCNN	آموزش	**۰,۹۹۱**	**۰,۹۹۹**	**۰,۹۹۵**
	تست	**۰,۹۰۱**	**۰,۹۰۹**	**۰,۹۳۹**
SSD300	آموزش	۰,۹۶۸	۰,۹۹	۰,۹۹۶
	تست	۰,۷۹۲	۰,۸۹۹	۰,۸۸۶

بر اساس نتایج ارائه‌شده در جدول (۱)، تفاوت معناداری میان عملکرد مدل‌های تک‌مرحله‌ای و دومرحله‌ای در مجموعه آزمون مشاهده می‌شود. اگرچه تمامی مدل‌ها در فاز آموزش به مقادیر بالایی از mAP50 (بیش از ۰,۹۵) دست یافته‌اند، اما عملکرد آن‌ها در فاز آزمون متفاوت بوده و شکاف تعمیم‌پذیری در برخی معماری‌ها مشهود است.

در میان مدل‌های بررسی‌شده، Faster R-CNN بالاترین مقدار mAP50 در مجموعه آزمون (۰,۹۳۹) را ثبت کرده است که نشان‌دهنده توان بالاتر این مدل در تعمیم به داده‌های دیده‌نشده است. مقدار Recall برابر با ۰,۹۰۹ نیز بیانگر آن است که این مدل توانایی شناسایی بخش عمده‌ای از نمادهای موجود در تصاویر را دارد، که برای مسئله حاضر از اهمیت ویژه‌ای برخوردار است. در مقابل، مدل‌های خانواده YOLO اگرچه در فاز آموزش عملکرد بسیار مطلوبی نشان داده‌اند (mAP50 در بازه ۰,۹۵۳ تا ۰,۹۸۵)، اما در فاز آزمون با افت قابل توجهی مواجه شده‌اند (mAP50 حدود ۰,۸۳ تا ۰,۸۴). این اختلاف می‌تواند نشانه‌ای از حساسیت بیشتر این معماری‌ها به تنوع بصری داده‌ها و پیچیدگی ساختاری نمادها باشد. کاهش Recall در این مدل‌ها نیز حاکی از آن است که بخشی از نمادهای کوچک یا دارای شباهت هندسی بالا به‌درستی شناسایی نشده‌اند. مدل SSD300 رفتار متفاوتی از خود نشان داده است؛ به‌طوری‌که در فاز آزمون Recall نسبتاً بالایی (۰,۸۹۹) دارد اما مقدار Precision آن پایین‌تر (۰,۷۹۲) است. این موضوع بیانگر افزایش نرخ تشخیص‌های نادرست (False Positive) در این مدل است که در کاربردهای عملی می‌تواند منجر به افزایش خطای نوع اول و نیاز به بازبینی انسانی شود. مقایسه شکاف میان عملکرد آموزش و آزمون نشان می‌دهد که مدل‌های تک‌مرحله‌ای بیش از Faster R-CNN مستعد بیش‌برازش بر داده‌های آموزشی بوده‌اند. در حالی که Faster R-CNN با وجود تعداد epoch کمتر، توانسته تعادل مناسب‌تری میان دقت و تعمیم برقرار کند. این مسئله احتمالاً به ساختار دومرحله‌ای و استخراج ویژگی عمیق‌تر در مرحله پیشنهاد ناحیه بازمی‌گردد که امکان تمرکز دقیق‌تر بر نواحی دارای نماد را فراهم می‌کند. در مجموع، تحلیل جدول (۱) نشان می‌دهد که اگرچه مدل‌های تک‌مرحله‌ای از نظر سرعت و سادگی معماری مزیت دارند، اما در مسئله‌ای با نمادهای دارای جزئیات هندسی ظریف و تنوع مقیاسی بالا، معماری دومرحله‌ای Faster R-CNN عملکرد پایدارتر و قابل اتکاتری ارائه می‌دهد.

۳-۷ بررسی و تحلیل نتایج

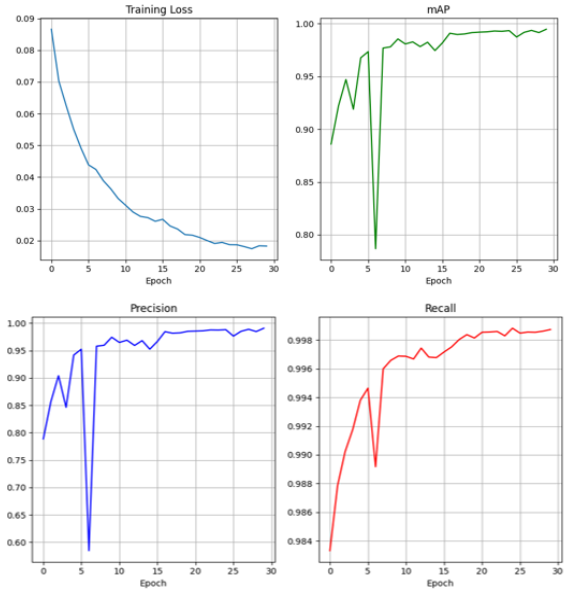
بر اساس مقادیر ارائه‌شده در جدول (۱)، تمرکز اصلی در ارزیابی عملکرد مدل‌ها بر نتایج مجموعه تست قرار داده شده است؛ چرا که این نتایج نشان‌دهنده توان تعمیم مدل‌ها به داده‌های جدید و دیده‌نشده هستند. در این میان، مدل Faster R-CNN با اختلاف معنادار نسبت به سایر مدل‌ها، بهترین عملکرد را از خود نشان داده است. این مدل در فاز تست به مقدار mAP50 برابر با ۰/۹۳۹ دست یافته که به‌طور قابل توجهی بالاتر از تمام مدل‌های خانواده YOLO (حدود ۰/۸۳ تا ۰/۸۴) و مدل SSD300 (۰/۸۸۶) است.

از نظر معیار Recall نیز Faster R-CNN با مقدار ۰/۹۰۹ بالاترین نرخ تشخیص صحیح اشیاء را دارد که نشان می‌دهد این مدل در شناسایی اکثر نمادهای موجود در تصاویر، حتی در شرایط پیچیده، بسیار موفق عمل کرده است. این ویژگی برای مسئله تشخیص نمادهای ناهنجار اهمیت ویژه‌ای دارد، زیرا عدم تشخیص حتی یک نماد می‌تواند منجر به باقی ماندن محتوای نامطلوب در فضای مجازی شود. علاوه بر این، مقدار Precision برابر با ۰/۹۰۱ بیانگر آن است که پیش‌بینی‌های انجام‌شده توسط مدل از دقت مناسبی برخوردار بوده و نرخ تشخیص‌های نادرست در سطح پایینی قرار دارد.

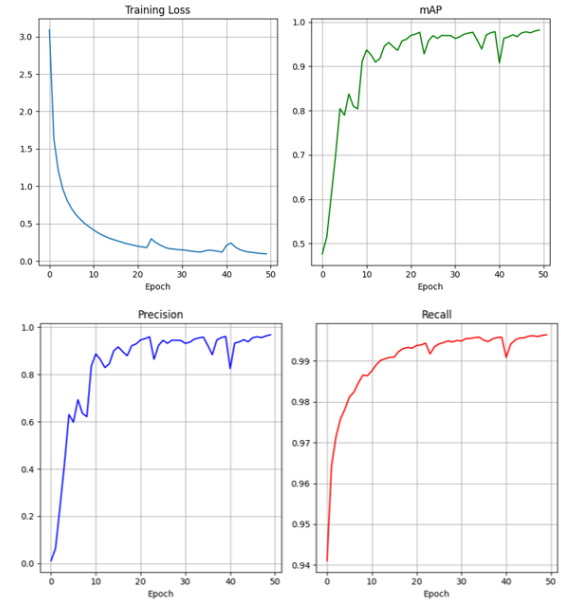
نکته قابل توجه دیگر، بهره‌وری بالای آموزشی Faster R-CNN است؛ به‌طوری‌که این مدل با ۳۰ دوره آموزشی توانسته به عملکردی بهتر از مدل‌های YOLO و SSD دست یابد که با تعداد دوره‌های آموزشی بیشتری آموزش داده شده‌اند. این موضوع نشان‌دهنده توان بالاتر Faster R-CNN در استخراج ویژگی‌های معنادار و یادگیری مؤثر الگوهای مرتبط با نمادها است.

در مقابل، مدل‌های YOLOv8n تا YOLOv12n اگرچه در فاز آموزش عملکرد بسیار خوبی (mAP50 بالاتر از ۰/۹۵) داشته‌اند، اما افت قابل‌توجهی در فاز تست مشاهده می‌شود؛ به‌گونه‌ای که Recall آن‌ها در بازه ۰/۷۷۲ تا ۰/۷۸۹ قرار گرفته است. این فاصله میان عملکرد آموزش و تست می‌تواند نشان‌دهنده چالش تعمیم‌پذیری این مدل‌ها در مواجهه با تنوع ظاهری نمادها، تغییر مقیاس، چرخش و پس‌زمینه‌های پیچیده باشد. از آنجا که نمادهای ناهنجار معمولاً دارای شباهت‌های هندسی بالا و جزئیات ظریف هستند، مدل‌های تک‌مرحله‌ای ممکن است در تشخیص دقیق این تفاوت‌ها با محدودیت مواجه شوند.

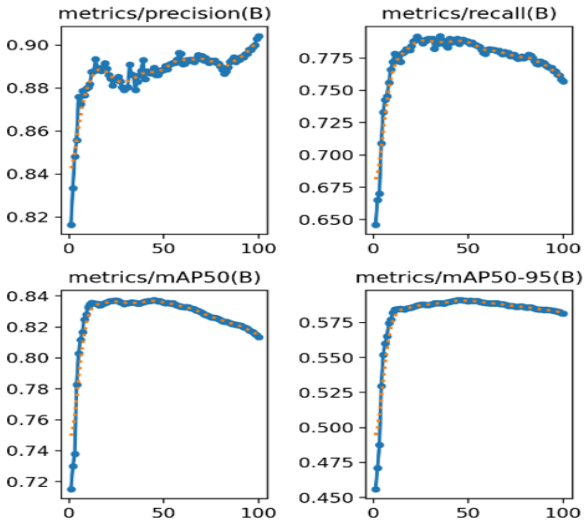
مدل SSD300 نیز با وجود دستیابی به Recall نسبتاً بالا در فاز تست (۰/۸۹۹)، از مقدار Precision پایین‌تری (۰/۷۹۲) برخوردار است که نشان‌دهنده افزایش تشخیص‌های نادرست است. این مسئله می‌تواند در کاربردهای واقعی منجر به حذف اشتباه محتوای سالم یا افزایش بار انسانی در بازبینی نتایج شود.



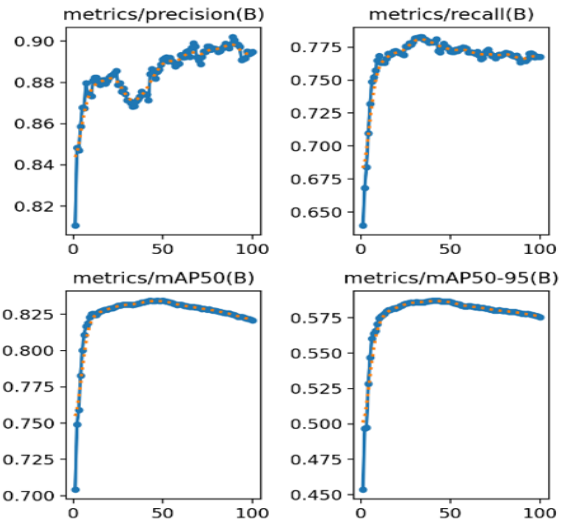
شکل (۵) : معیارهای ارزیابی در آموزش مدل Faster-RCNN



شکل (۶) : معیارهای ارزیابی در آموزش مدل SSD300



شکل (۷) : معیارهای ارزیابی در آموزش مدل YOLOv8



شکل (۸) : معیارهای ارزیابی در آموزش مدل YOLOv12

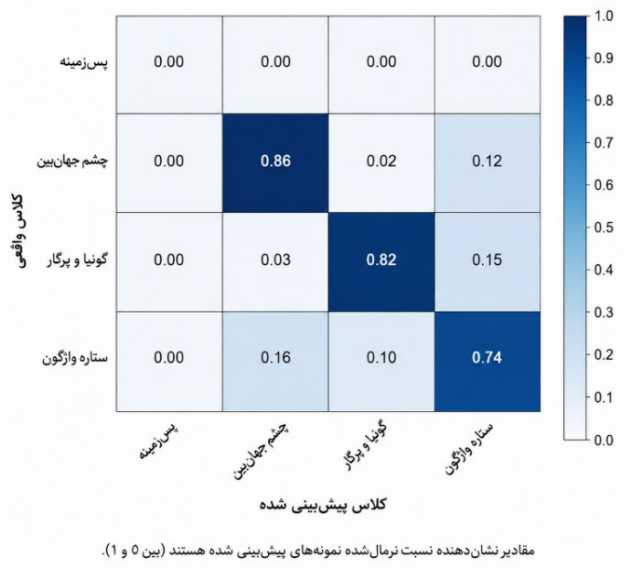
تحلیل روند آموزش مدل‌ها در شکل‌های (۵) تا (۸) نشان می‌دهد که تمامی معماری‌های مورد بررسی در طول فرایند آموزش به همگرایی قابل قبولی دست یافته‌اند، اما رفتار یادگیری و پایداری آن‌ها متفاوت است. مدل‌های خانواده YOLO در دوره‌های ابتدایی آموزش با سرعت بیشتری به مقادیر بالای mAP و Precision دست یافته‌اند که نشان‌دهنده توان مناسب این معماری‌ها در یادگیری سریع ویژگی‌های داده‌ها است. با این حال، مقایسه نتایج نهایی نشان می‌دهد که افزایش عملکرد در داده‌های آموزشی لزوماً با همان میزان بهبود در مجموعه آزمون همراه نبوده است.

در مقابل، مدل Faster R-CNN روند یادگیری یکنواخت‌تر و پایدارتری را در طول آموزش نشان می‌دهد. تغییرات معیارهای ارزیابی در این مدل با نوسان کمتری همراه بوده و مقادیر نهایی Precision، Recall و mAP در مجموعه آزمون نیز نسبت به سایر مدل‌ها بالاتر است. این نتایج بیانگر عملکرد متوازن‌تر Faster R-CNN در مسئله مورد بررسی است.

مدل SSD300 نیز رفتار نسبتاً پایداری در آموزش داشته است، اما اختلاف میان Recall و Precision در نتایج آزمون نشان می‌دهد که این مدل، با وجود شناسایی بخش قابل توجهی از نمادها، در برخی موارد تشخیص‌های نادرست بیشتری نسبت به Faster R-CNN داشته است. این موضوع می‌تواند ناشی از شباهت ساختاری برخی نمادها و پیچیدگی پس‌زمینه تصاویر باشد.

برآیند نتایج کمی و کیفی نشان می‌دهد که مدل Faster R-CNN در مقایسه با سایر معماری‌های مورد بررسی، تعادل مناسب‌تری میان دقت تشخیص و حساسیت در مجموعه آزمون برقرار کرده است. مقادیر بالاتر Recall و mAP50 در مجموعه آزمون بیانگر آن است که این مدل توانسته بخش بیشتری از نمادهای موجود در تصاویر را با دقت قابل قبول شناسایی کند، در حالی که میزان خطاهای تشخیص نیز در سطح مناسبی حفظ شده است. همچنین پایداری بیشتر روند آموزش و اختلاف کمتر میان نتایج آموزش و آزمون، بیانگر عملکرد پایدارتر Faster R-CNN در مواجهه با داده‌های آزمون است. در مجموع، نتایج این پژوهش نشان می‌دهد که اگرچه معماری‌های تک‌مرحله‌ای نظیر YOLO و SSD از نظر سرعت و سادگی ساختار مزایای قابل توجهی دارند، اما در مسئله حاضر که شامل نمادهای دارای شباهت هندسی، تغییر مقیاس و پس‌زمینه‌های پیچیده است، مدل Faster R-CNN به دلیل ساختار دومرحله‌ای خود، عملکرد پایدارتر و قابل اتکاتری ارائه داده است.

شکل (۹) نمونه‌ای از ماتریس درهم‌ریختگی حاصل از ارزیابی مدل‌های مورد بررسی را نشان می‌دهد. این ماتریس حاکی از تفکیک مناسب میان کلاس‌های مختلف نمادها بوده و نتایج آن بیانگر عملکرد مناسب مدل در تشخیص کلاس‌ها است. همچنین شکل (۱۰) بیانگر تطابق دقیق جعبه‌های مرزی پیش‌بینی‌شده با نواحی واقعی نمادها می‌باشد.



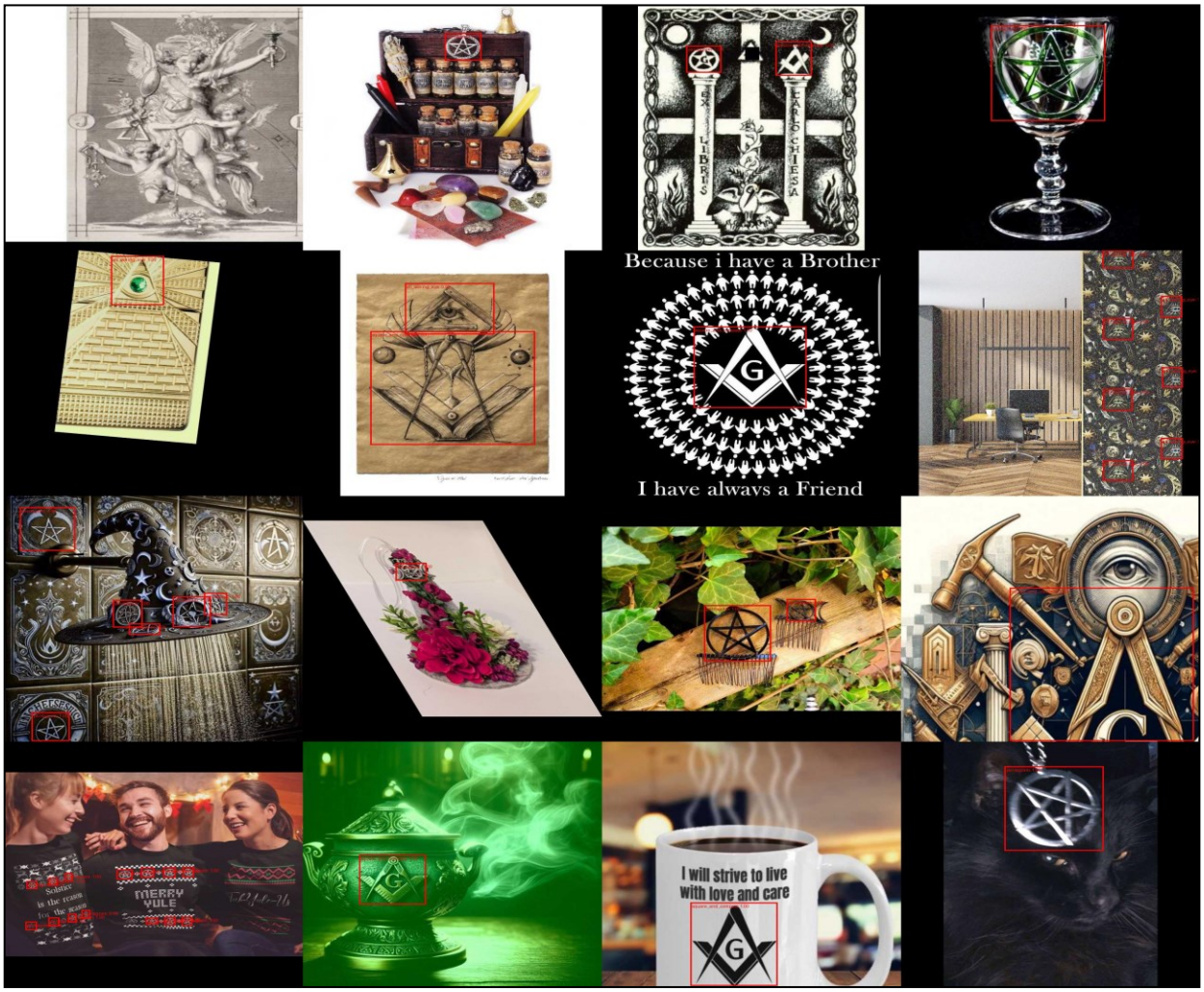
شکل (۹) : ماتریس درهم‌ریختگی در تست مدل Faster-RCNN

علاوه بر مقایسه مقادیر کمی، رفتار کیفی مدل‌ها نیز مورد بررسی قرار گرفت. نتایج نشان داد که خطاهای مدل‌های تک‌مرحله‌ای عمدتاً در شرایطی رخ می‌دهد که نمادها در مقیاس کوچک، با چرخش شدید یا در پس‌زمینه‌های شلوغ قرار دارند. در این موارد، برخی پیش‌بینی‌ها یا حذف شده‌اند (False Negative) یا با اشیای دارای شباهت هندسی اشتباه گرفته شده‌اند. در مقابل، Faster R-CNN در تشخیص نواحی دارای جزئیات ساختاری پیچیده عملکرد پایدارتری از خود نشان داده است. تحلیل کیفی نمونه‌های اشتباه نشان می‌دهد که این مدل توانسته تفکیک بهتری میان الگوهای متقارن و اشکال مشابه ایجاد کند، هرچند در برخی موارد افزایش پیچیدگی پس‌زمینه موجب کاهش اطمینان پیش‌بینی شده است. همچنین بررسی شکاف عملکرد میان آموزش و آزمون نشان می‌دهد مدل‌های تک‌مرحله‌ای حساسیت بیشتری به تغییرات توزیع داده داشته‌اند، در حالی که Faster R-CNN پایداری بیشتری در تعمیم‌پذیری نشان داده است. این یافته‌ها حاکی از آن است که در مسئله تشخیص نمادهای هندسی با شباهت ساختاری بالا، انتخاب معماری تأثیر قابل توجهی بر توازن میان دقت، حساسیت و تعمیم دارد. در مجموع، نتایج کمی و کیفی نشان می‌دهد که اگرچه مدل‌های سریع تک‌مرحله‌ای برای کاربردهای بلادرنگ مناسب هستند، اما در کاربردهایی که دقت تشخیص اولویت اصلی است، معماری‌های دومرحله‌ای گزینه مطمئن‌تری محسوب می‌شوند.

با وجود عملکرد مطلوب مدل‌ها، این پژوهش با محدودیت‌هایی همراه است. عملکرد سامانه در تصاویر دارای نویز شدید، محوشدگی قابل‌توجه یا کیفیت بصری بسیار پایین به‌صورت مستقل ارزیابی نشده است و ممکن است در چنین شرایطی دقت تشخیص کاهش یابد. همچنین، ارزیابی مدل‌ها بر روی مجموعه‌داده گردآوری‌شده در این پژوهش انجام شده و آزمون بر روی داده‌های خارجی نیز صورت گرفته است. نتایج نشان داد که مدل در شرایط واقعی نیز عملکرد قابل قبولی در تشخیص این دسته نمادها دارد. همچنین بررسی تعمیم‌پذیری مدل در مواجهه با داده‌های برون‌دامنه با شرایط تصویری نامطلوب، می‌تواند به‌عنوان یکی از محورهای پژوهش‌های آتی مدنظر قرار گیرد. انتخاب مدل‌های مورد مقایسه در این پژوهش به‌صورت هدفمند و با تمرکز بر معماری‌های متداول و پایدار تشخیص شیء انجام شده است.

۴ نتیجه و جمع‌بندی

نمادها به‌عنوان ابزارهایی قدرتمند برای انتقال مفاهیم پنهان،



شکل (۱۰) : نمونه‌هایی از پیش‌بینی و طبقه‌بندی مدل Faster-RCNN

همواره نقش مهمی در شکل‌دهی پیام‌های فرهنگی و رسانه‌ای ایفا کرده‌اند و در بسیاری از موارد، بدون جلب توجه مستقیم مخاطب، بر ادراک و نگرش او تأثیر می‌گذارند. در این مقاله، با هدف شناسایی خودکار نمادهای ناهنجار در محتوای تصویری، یک مجموعه‌داده تصویری جدید شامل سه نماد «چشم جهان‌بین»، «ستاره واژگون» و «گونیا و پرگار» گردآوری و ایجاد شد و عملکرد چندین مدل مطرح پردازش تصویر برای تشخیص این نمادها مورد بررسی قرار گرفت.

با توجه به عدم وجود مجموعه‌داده‌ای اختصاصی در این حوزه، دیتاست پیشنهادی می‌تواند به‌عنوان یک مرجع پایه برای توسعه مجموعه‌داده‌های گسترده‌تر و همچنین آموزش و ارزیابی مدل‌های پیشرفته‌تر تشخیص شیء مورد استفاده قرار گیرد. در این راستا، مدل‌های Faster R-CNN، SSD300 و معماری‌های مختلف YOLO (نسخه‌های ۸ تا ۱۲) روی مجموعه‌داده پیشنهادی آموزش داده شده و نتایج آن‌ها با یکدیگر مقایسه گردید. یافته‌ها نشان داد که مدل Faster R-CNN با وجود تعداد دوره‌های آموزشی کمتر، به مقادیر بالاتری از معیارهای ارزیابی دست یافته و نسبت به مدل‌های تک‌مرحله‌ای عملکرد پایدارتر و دقیق‌تری از خود نشان داده است.

این نتایج بیانگر توان بالای Faster R-CNN در استخراج ویژگی‌های معنادار و تشخیص دقیق نمادهای دارای شباهت‌های هندسی و جزئیات ظریف است. در مجموع، نتایج این پژوهش نشان می‌دهد که استفاده از مدل‌های تشخیص شیء مبتنی بر یادگیری عمیق، به‌ویژه مدل Faster R-CNN، می‌تواند راهکاری مؤثر برای شناسایی هوشمند نمادهای ناهنجار در داده‌های تصویری باشد و به کاهش خطای انسانی در پالایش محتوای رسانه‌ای کمک کند.

در ادامه مسیر پژوهش، می‌توان با گسترش مجموعه‌داده از نظر تعداد نمونه‌ها، تنوع بصری نمادها و لحاظ‌کردن شرایط پیچیده‌تری نظیر چرخش، تغییر نور و هم‌پوشانی نمادها، عملکرد و قابلیت تعمیم مدل‌ها را بهبود داد. همچنین بررسی مدل‌های پیشرفته‌تر مبتنی بر یادگیری نیمه‌نظارتی یا بدون‌ناظر و توسعه سامانه‌های بلادرنگ برای تحلیل ویدئوها و شبکه‌های اجتماعی، می‌تواند به‌عنوان مسیرهای آینده پژوهش در این حوزه مورد توجه قرار گیرد.

مراجع

[1] S. Routray, A. K. Ray, and C. Mishra, "Analysis of various image feature extraction methods against noisy image: SIFT, SURF and HOG," in 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2017: IEEE, pp. ۱–۵.

[2] K. O'shea and R. Nash, "An introduction to convolutional neural networks," arXiv preprint arXiv:1511.08458, 2015.

[3] W. Zhou, S. Gao, L. Zhang, and X. Lou, "Histogram of oriented gradients feature extraction from raw bayer pattern images," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 67, no. 5, pp. 946–950, 2020.

[4] M. A. Chandra and S. Bedi, "Survey on SVM and their application in image classification," International Journal of Information Technology, vol. 13, no. 5, pp. 1–11, 2021.

[5] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in 2007 IEEE 11th international conference on computer vision, 2007: Ieee, pp. 1–8 .

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788 .

[7] J. Solawetz. "What is YOLOv8? A Complete Guide." https://blog.roboflow.com/what-is-yolov8/ (accessed Oct 23, 2024.

[8] P .Potrimba. "What is New in YOLOv9? An Architecture Deep Dive." Roboflow Blog. https://blog.roboflow.com/yolov9-deep-dive/ (accessed May 20, 2024.

[9] P. Potrimba. "What is YOLOv10? An Architecture Deep Dive." https://blog.roboflow.com/what-is-yolov10/ (accessed Jun 14, 2024.

[10] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," arXiv preprint arXiv:2410.17725, 2024.

[11] "YOLOv12." https://roboflow.com/model/yolov12 (accessed Feb 18, 2025.

[12] Y. Tian, Q. Ye, and D. Doermann, "Yolov12 : Attention-centric real-time object detectors," arXiv preprint arXiv:2502.12524, 2025.

[13] W. Liu et al., "Ssd: Single shot multibox detector," in European conference on computer vision, 2016: Springer, pp. 21–37 .

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, vol. 28, 2015.

[15] J. Hui. "SSD object detection: Single Shot MultiBox Detector for real-time processing." Medium. https://jonathan-hui.medium.com/ssd-object-

detection-single-shot-multibox-detector-for-real-time-processing-9bd8deac0e06 (accessed Mar 14, 2018.

[16] C. Guo, B. Fan, Q. Zhang, S. Xiang, and C. Pan, "Augfpn: Improving multi-scale feature learning for object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 12595–12604 .

[17] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in 2018 international interdisciplinary PhD workshop (IIPhDW), 2018: IEEE, pp. 117–122 .

[18] Torralba, "Context-based vision system for place and object recognition," in Proceedings Ninth IEEE International Conference on Computer Vision, 2003: IEEE, pp. 273–280 vol. 1 .

[19] J. Li, X. Liang, Y. Wei ,T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1222–1230 .

[20] B. Singh and L. S. Davis, "An analysis of scale invariance in object detection snip," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 3578–3587 .

[21] B. Singh, M. Najibi, and L. S. Davis, "Sniper: Efficient multi-scale training," Advances in neural information processing systems, vol. 31, 2018.

[22] S. Gidaris and N. Komodakis, "Object detection via a multi-region and semantic segmentation-aware cnn model," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1134–1142 .

[23] X. Zeng, W. Ouyang ,B. Yang, J. Yan, and X. Wang, "Gated bi-directional cnn for object detection," in European conference on computer vision, 2016: Springer, pp. 354–369 .

[24] Z. Chen, Q. Xu, R. Cong, and Q. Huang, "Global context-aware progressive aggregation network for salient object detection," in Proceedings of the AAAI conference on artificial intelligence, 2020, vol. 34, no. 07, pp. 10599–10606 .

[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 6, pp. 1137–1149, 2016.

[26] M. Sohan, T. Sai Ram, and C. V. Rami Reddy, "A review on yolov8 and its advancements," in International Conference on Data Intelligence and Cognitive Informatics, 2024: Springer, pp. 529–545 .

[27] C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, "Yolov9: Learning what you want to learn using programmable gradient information," in European conference on computer vision, 2024: Springer, pp. 1–21 .

[28] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, and J. Han, "Yolov10: Real-time end-to-end object detection," Advances in Neural Information Processing Systems, vol. 37, pp. 107984–108011, 2024.

[29] "Pentagram." https://en.wikipedia.org/wiki/Pentagram (accessed.

[30] "Square and Compasses ". https://en.wikipedia.org/wiki/Square_and_Compasses accessed.

[31] "Eye of Providence." https://en.wikipedia.org/wiki/Eye_of_Providence accessed.

[32] "Satanism." https://en.wikipedia.org/wiki/Satanism accessed.

[33] "Sigil of Baphomet." https://en.wikipedia.org/wiki/Sigil_of_Baphomet accessed.

مهسا سیروزیان فرد مدرک کارشناسی و کارشناسی ارشد خود را از دانشگاه صنعتی همدان دریافت کرده است. به تحقیق در زمینه بینایی کامپیوتر علاقمند بوده و در حال حاضر انجام دهنده پژوهش در موضوعات حوزه‌های پردازش تصویر و یادگیری عمیق است.

مهلقا افراسیابی مدرک کارشناسی، کارشناسی ارشد و دکتری خود را در رشته مهندسی کامپیوتر گرایش هوش مصنوعی از دانشگاه بوعلی سینا دریافت کرده است. در حال حاضر استادیار دانشگاه صنعتی همدان است. زمینه‌های پژوهشی ایشان پردازش تصویر، پردازش سیگنال، یادگیری ماشین و یادگیری عمیق می‌باشد.