نشریه علمی علوم و فنون نقشه‌برداری، دوره ۱۳، شماره ۱، صفحه ۸۳ تا ۹۷، شهریور ماه ۱۴۰۲ – مقاله پژوهشی

بهبود شبکه یادگیری عمیق YOLOv5 برای شناسایی خودرو و استخرهای روباز با استفاده از تصاویر پهپادی

آیدین ابراهیمی*[۱]، امیررضا گروسی[۱]، علی حسینی نوه[۲]، علی محمدزاده[۳]

[۱] دانشجوی کارشناسی ارشد سنجش از دور، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی
a.ebrahimi3@email.kntu.ac.ir
a.garousi@email.kntu.ac.ir

[۲] دانشیار گروه سنجش از دور و فتوگرامتری، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی
hosseininaveh@kntu.ac.ir

[۳] دانشیار گروه سنجش از دور و فتوگرامتری، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی
a_mohammadzadeh@kntu.ac.ir

(دریافت: خرداد ۱۴۰۲، تصویب: شهریور ۱۴۰۲)

چکیده

تشخیص اجسام کوچک مانند خودرو و استخرها در تصاویر پهپادی با توان تفکیک مکانی بالا، به دلیل ویژگی‌های هندسی و رنگ مشابه آن‌ها، با چالش‌هایی روبرو است. افزایش تعداد خودروها نه تنها از منظر ترافیک شهری یک چالش مهم محسوب می‌گردد بلکه منجر به مشکلات زیست‌محیطی نظیر آلودگی و گرم‌شدن هوا نیز می‌گردد؛ از این‌رو، پایش این اهداف می‌تواند نقشی مهم در مدیریت این مشکلات داشته باشد. از طرفی، ساخت و نگهداری استخرهای آبی نیز به مقدار قابل توجهی آب نیاز دارد و پایش این اهداف در محیط‌های شهری برای صرفه‌جویی در مصرف آب ضروری است. در این راستا، تصاویر سنجش‌ازدور پهپادی و شبکه‌های یادگیری عمیق که توانایی بالایی در شناسایی اشیاء از این تصاویر را دارند، ابزاری مناسب برای پایش این اهداف محسوب می‌شوند. اگرچه تاکنون پژوهش‌های ارزشمندی در این زمینه برای مقابله با هریک از چالش‌های محیط زیستی مطرح‌شده صورت گرفته‌است، اما همچنان کاستی‌هایی در آن‌ها وجود دارد. در این مطالعه، یک شبکه یادگیری عمیق جدید YOLOv5+ برای شناسایی دو هدف خود رو و استخر آبی از تصاویر پهپادی توسعه داده شده‌است، بطوری که در آن عملکرد شبکه در استخراج ویژگی‌های کارآمد به دلیل بکارگیری مکانیسم Inception در لایه‌های میانی تقویت شده‌است. همچنین، در این تحقیق، از داده‌های پهپادی مرجع DJI Mavic و DJI Mini Se که از مناطق تیانجین در کشور چین و کان در کشور فرانسه اخذ شده‌اند، برای ارزیابی عملکرد شبکه پیشنهادی و مقایسه آن با شبکه‌های یادگیری عمیق YOLOv5 و YOLOv7 استفاده گردید. در نهایت، نتایج نشان داد شبکه پیشنهادی با دقت کلی ۹۵٪، بطور میانگین عملکرد شبکه‌های قیاسی را ۲ درصد بهبود بخشیده‌است که نشان‌دهنده کارایی رویکرد پیشنهادی در این تحقیق است.

واژگان کلیدی: یادگیری عمیق، تصاویر سنجش‌ازدور ماهواره‌ای، تشخیص خودرو، استخر، قدرت تفکیک مکانی بالا، شبکه‌های عصبی پیچشی

* نویسنده رابط

۱- مقدمه

در سال های اخیر به دلیل توسعه روزافزون ارتباطات بی سیم و فناوری هوش مصنوعی، تصاویر سنجش از دور پهپادی به عنوان یکی از منابع داده ارزشمند در شناسایی اهداف زمینی موردتوجه بسیاری از محققین قرار گرفته‌است. به تبع بهبود کیفیت و توان تصویربرداری در سنجنده‌های پهپادی، تصاویر آن‌ها حاوی اطلاعات دقیق‌تر و جزئی‌تری از عوارض زمینی هستند. بطور جزئی‌تر، برخی از اهداف در تصاویر سنجش از دور پهپادی مانند خودرو، استخر رو باز و هواپیما به راحتی در این تصاویر قابل شناسایی هستند، که به همین دلیل تصاویر سنجش از دور پهپادی به موضوعی جذاب برای محققان در زمینه زمین شناسی [۱]، کشاورزی [۲]، نظامی [۳] و جنگلداری [۴] تبدیل شده است. به عنوان مثال، یائو و همکاران [۵] الگوریتم بهبودیافته مبتنی بر Mask RCNN را برای بهبود دقت شناسایی ساختارهای زمین‌شناسی پیشنهاد کردند. منگ و همکاران [۶] از الگوریتم تشخیص اشیاء برای جستجوی نشت گاز در تصاویر سنجش از دور پهپادی استفاده کردند. تحقیقات آن‌ها می‌تواند آلودگی محیط زیست را کاهش دهد و در مدیریت خسارات احتمالی مفید باشد. با این وجود، تشخیص دقیق اشیای زمینی در تصاویر سنجش از دور پهپادی همچنان یک چالش بزرگ به شمار می‌رود. اولاً، اشیاء به راحتی تحت تأثیر عوامل محیطی مانند شدت نور، تأثیرات توپوگرافی و شرایط آب و هوایی غالب در پس زمینه تصاویر سنجش از دور پهپادی قرار می‌گیرند و به راحتی نمیتوان آن‌ها را تشخیص داد. ثانیاً، توانایی استخراج ویژگی های اشیاء و ادغام روش های موجود همچنان جای پیشرفت دارد، زیرا بدیهی است استخراج ویژگی‌های غنی‌تر منجر به بهبود دقت تشخیص می‌شود. توانایی الگوریتم تشخیص اشیاء سنتی مبتنی بر روش همسان سازی الگو [۷] موثر نیست، زیرا این روش نیاز به استخراج و طبقه بندی دستی ویژگی های اشیاء زمینی را دارد، همینطور سرعت و دقت آن در هنگام مواجهه با تعداد زیادی از اشیاء کوچک قابل اطمینان نیست. الگوریتم های یادگیری ماشین [۸] اغلب نیاز به توصیف تعداد زیادی از ویژگی‌ها با استفاده از روش های آماری ریاضی پیچیده دارند و توان تشخیص و تعمیم‌پذیری آن‌ها کافی نیست. با توسعه روش‌های یادگیری عمیق در پژوهش‌های تشخیص اشیاء، تعداد زیادی روش تشخیص اشیاء مبتنی بر شبکه‌های عصبی عمیق پیشنهاد شده است. از جمله آن‌ها، الگوریتم‌های دو مرحله‌ای، الگوریتم‌های سری R-CNN [۹-۱۱] هستند. در حالی که الگوریتم‌های تک مرحله‌ای شامل الگوریتم‌های سری YOLO [۱۲-۱۸] و الگوریتم‌های سری SSD [۱۹-۲۱] هستند. ظهور این الگوریتم‌ها مبتنی بر شبکه‌های عصبی عمیق، عملکرد تشخیص اشیاء را به میزان قابل توجهی بهبود بخشیده است. تصویر سنجش از دور پهپادی به عنوان یک نوع منبع داده پرکاربرد نیز در پژوهش‌های تشخیص اشیاء مورد استفاده قرار می‌گیرد. یان و همکاران [۲۲] الگوریتم Faster-RCNN بهبودیافته را برای بهبود دقت تشخیص منابع معدنی پیشنهاد کردند. لو و همکاران [۲۳] با کاهش درگاه خروجی ادغام ویژگی شبکه گردن در الگوریتم YOLO، دقت تشخیص اشیاء هواپیما در تصاویر سنجش از دور را بهبود بخشیدند. اگرچه الگوریتم فوق در تشخیص اشیاء با یک عارضه عملکرد تشخیص عالی دارد، اما تعمیمپذیری آن ضعیف است زیرا تصاویر سنجش از دور اغلب حاوی عارضه های مختلفی از اشیاء هستند.

در این تحقیق، برای شناسایی دو عارضه زمینی مختلف شامل خودرو و استخر آبی روباز یک رویکرد نوآورانه با ادغام ماژول Inception در ساختار YOLOv5 ارائه می‌شود که امکان استخراج ویژگی های متنوع و غنی از داده های ورودی را فراهم می‌کند. ترکیب کرنل های ۳x۳، ۵x۵ و ۷x۷ در ماژول Inception، باعث بهبود قابل توجهی در دقت تشخیص میشود. این رویکرد از نقاط قوت YOLOv5 و ماژول Inception استفاده می‌کند و این امکان را فراهم می‌کند تا مدل عملکرد برتری در تشخیص اشیا داشته باشد. این نواوری گام رو به جلوی قابل توجهی در حوزه تشخیص اشیا محسوب می‌شود. روش پیشنهادی نه تنها دقت را افزایش می‌دهد، بلکه کاربرد مدل را در پس‌زمینه‌های پیچیده و حوزه های مختلف، از جمله سیستم های خودران، نظارت و تصویربرداری پزشکی گسترش می‌دهد. با افزایش تقاضا برای سیستم های تشخیص اشیای قوی، مدل +YOLOv5 گواهی بر قدرت تکنیک های استخراج ویژگی نوآورانه در پیشبرد قابلیت های مدل های یادگیری عمیق است.

۱-۲- کارهای مرتبط

۱-۲-۱- الگوریتم‌های تشخیص اشیاء متداول

مطالعات زیادی برای غلبه بر چالش‌های مذکور انجام شده است. برای مثال، قبل از سال ۲۰۱۰، اگرچه روش‌های تشخیص اشیاء سنتی مانند تطبیق الگو [۷] و یادگیری ماشین [۸] برای پیاده‌سازی راحت بودند، ولی عملکرد آن‌ها ضعیف بود. در سال ۲۰۱۲، کرزژوسکی و همکارانش [۲۴] یک شبکه عصبی پیچشی عمیق (D-CNN) به نام AlexNet را پیشنهاد کردند تا تشخیص اشیاء را به عصر یادگیری عمیق وارد کنند. گیرشیک [۱۰] یک الگوریتم تشخیص اشیاء دو مرحله‌ای به نام شبکه عصبی پیچشی ناحیه‌ای (R-CNN) را پیشنهاد کرد که از سایر الگوریتم‌ها در دقت تشخیص پیشی می‌گیرد، اما سرعت تشخیص آن کندتر است. بر اساس این، سان و همکارانش [۱۱] و رن و همکارانش [۹] به ترتیب Fast R-CNN وFaster R-CNN را پیشنهاد کردند که سرعت و دقت الگوریتم R-CNN را بهبود می‌بخشند. اگرچه الگوریتم دو مرحله‌ای بهبود یافته کارایی تشخیص را بهبود بخشیده است، اما همچنان برآوردن نیازهای تشخیص عملی دشوار است. بنابراین، برای افزایش کاربردی بودن الگوریتم، آنگولوف و همکارانش [۲۰] یک الگوریتم تشخیص تک مرحله‌ای (SSD) را پیشنهاد کردند. الگوریتم‌های سری SSD دقت و سرعت تشخیص بالایی دارند، اما یک عیب آن‌ها این است که به پارامترهای دستی زیادی نیاز دارند که باعث ایجاد مشکل در کاربردهای عملی می‌شود. از سال ۲۰۱۶، ردمون و همکارانش [۱۷] ماژول YOLO[1] و YOLOv2 را معرفی کردند. این ماژول علاوه بر افزایش دقت تشخیص، سرعت آن را نیز افزایش میدهد. در عین حال، به دلیل چارچوب سلول شبکه منحصر به فرد آن، عملکرد تشخیص اشیاء کوچک نسبتاً ضعیف است. در سال ۲۰۱۸، ردمون و همکارانش [۱۵] YOLOv3 را با افزودن افزایش داده موزاییک برای تقویت توانایی شبکه برای پیشی گرفتن از مدل دو مرحله‌ای در بحث دقت پیشنهاد کردند و YOLOv4 [۱۴] در سال ۲۰۲۰ ایجاد شد. YOLOv5 شناخته شده‌ترین دنباله الگوریتم‌های سری YOLO است. اگرچه دقت تشخیص متوسط و کارایی تشخیص نسبت به کارهای قبلی بهبود یافته است، اما اثر تشخیص اشیاء با پس‌زمینه‌های پیچیده همچنان نیاز به بهبود دارد.

۱-۲-۲- تشخیص اشیاء در تصاویر سنجش از دور پهپادی

در سال‌های اخیر، با توسعه کاربردهای شبکه عصبی پیچشی عمیق در تصاویر، بسیاری از محققان در زمینه تشخیص اشیاء در تصاویر سنجش از دور پهپادی تحقیقات ارزشمند زیادی انجام داده‌اند. دا و همکارانش [۲۵] ماژول انتقال[2] و شبکه باقیمانده[3] را در شبکه اصلی YOLOv3 تعبیه کردند که می‌توانست دقت تشخیص کشتی‌های کوچک در تصاویر سنجش از دور را بهبود بخشد. لو و همکارانش [۲۳] با کاهش درگاه‌های خروجی ادغام ویژگی شبکه گردن در الگوریتم YOLO، دقت تشخیص هواپیما را در تصاویر سنجش از دور بهبود بخشیدند. کائو و همکارانش [۲۶] چارچوب CSP[4] را در شبکه اصلی YOLO بهینه سازی کردند. همچنین آن‌ها توانایی طبقه‌بندی غیرخطی شبکه را با تابع فعال‌سازی Mish افزایش دادند. روش‌های فوق توانایی ادغام ویژگی را تقویت کرده و دقت تشخیص اشیاء را افزایش دادند. لی و همکارانش [۲۷] یک روش مؤثر تشخیص اشیاء در محیط‌های ارتفاع پایین برای تصاویر سنجش از دور را پیشنهاد کردند. این روش اگرچه روی تصاویر پهپادی دقت خوبی دارد ولی بروی تصاویر هوایی نتیجه مطلوبی ندارد. وانگ و همکارانش [۲۸] در سال ۲۰۲۱ ماژول توجه CBAM[۲۹] را در شبکه رایج YOLOv5 تعبیه کردند تا توانایی الگوریتم را برای تشخیص اشیاء کوچک تقویت کنند. اگرچه این الگوریتم دقت تشخیص را بهبود می‌بخشد، اما نیاز به محاسبات زیادی دارد و کارایی عملیات را کاهش می‌دهد. بر این اساس، یانگ و همکارانش [۳۰] ماژول ECA [۳۱] و چارچوب SAHI [۳۲] را در YOLOv5 معرفی کردند و آن‌ها را روی سه مجموعه داده تصویر سنجش از دور پهپادی آزمایش کردند و به نتایج تشخیص عالی دست یافتند.

---

۱ You Only Look Once

۲ Transition Module

۳ Residual Network

۴ Cross Stage Partial

۱-۲-۳- اصول چارچوب تشخیص YOLOv5 رایج

چارچوب تشخیص اشیاء YOLOv5 که در سال ۲۰۲۰ توسط Ultralytics LLC پیشنهاد شد، یک چارچوب بهبود یافته بر اساس سری YOLO می‌باشد. از نظر ساختاری، دارای چارچوب تشخیص تک مرحله‌ای است که از چهار واحد تشکیل شده است: ورودی، شبکه ستون فقرات[۱]، شبکه گردن و خروجی. YOLOv5 با بهره‌گیری از مزایای نسخه‌های قبلی سری YOLO و سایر الگوریتم‌های تشخیص، لایه Focus را برای افزایش داده[۲] در ورودی تعبیه می‌کند. در همین حال، از DarkNet53 در ستون فقرات برای استخراج ویژگی‌های اصلی از تصویر استفاده میکند. یک چارچوب ادغام ویژگی که شامل ساختار هرم ویژگی [۳۳] (FPN) و شبکه تجمع مسیر از پایین به بالا [۳۴] در شبکه گردن تعبیه شده‌است تا پیوند اتصال کوتاه و ادغام بین لایه ای را در ویژگی های چند مقیاسی تقویت کند. چارچوب کامل YOLOv5 در شکل ۱ نشان داده شده است. چهار واحد تشکیل دهنده در شبکه YOLOv5 به شرح زیر نشان داده شده است:

ورودی: YOLOv5 مانند YOLOv4 از ماژول موزائیک[۳] برای افزایش داده استفاده می‌کند. YOLOv5 به دلیل استفاده از چهار عکس به طور قابل توجهی مقدار ویژگی را افزایش می‌دهد. ادغام، توانایی تعمیم تشخیص اطلاعات پس زمینه و اشیاء را افزایش می‌دهد و بار محاسباتی را کاهش می‌دهد. چارچوب بهبودیافته، اندازه تصاویر ورودی را از طریق ماژول مقیاس گذاری تطبیقی تصویر[۴] به ابعاد یکپارچه ۶۴۰ x 640 پیکسل تنظیم می‌کند تا پیچیدگی شناسایی اشیا را کمتر کند.

شبکه ستون فقرات: YOLOv5 از CSPDarknet53 [۱۵] به عنوان شبکه ستون فقرات استفاده می‌کند. شبکه ستون فقرات از یک لایه Focus، چارچوب CSPNet و ماژول ادغام هرم فضایی[۵] [۳۵] تشکیل شده است. هنگام برخورد با ویژگی های بیشتر، YOLOv5 می‌تواند ابتدا یک نقشه ویژگی را تقسیم و به هم متصل کند. سپس تصاویر را از طریق لایه های اتصال دهنده روی هم قرار دهد تا بتواند بازنمایی های ویژگی در سطوح مختلف را از طریق لایه های کانولوشن استخراج کند. چارچوب CSPNet، شبکه ستون فقرات را تشکیل می‌دهد و توانایی ادغام نقشه های ویژگی با ابعاد مختلف را از طریق اتصالات باقیمانده افزایش می‌دهد که اساس بازگشت به عقب در شبکه است.

شبکه گردن: این شبکه اطلاعات بافت و اطلاعات موقعیت را در نقشه ویژگی ادغام می‌کند تا توانایی ادغام اطلاعات در اشیاء را تقویت کند، شبکه گردن YOLOv5 از ساختار ادغام PAFPN [۳۶] استفاده می‌کند. ساختار FPN ویژگی‌های ابعاد مختلف را در شبکه از طریق نمونه افزایی[۶]، توانایی ادغام گراف و توانایی تشخیص اشیاء با اندازه های مختلف بیشتر را تقویت می‌کند. چارچوب PAN می‌تواند اطلاعات لایه کم عمق را از طریق پیوندهای اتصال کوتاه به لایه پایینی بیاورد که نتایج تشخیص اشیاء مزاحم را بهبود می‌بخشد.

خروجی: خروجی از ماژول NMS[۷] یا حذف غیر حداکثرها و تابع هزینه تشکیل شده است. YOLOv5 پایه از تابع هزینه CIoU در خروجی استفاده می‌کند. این عمل مشکل عدم تطابق IoU در موارد خاص را برطرف می‌کند و زمانی که عوارض پیش بینی شده همپوشانی دارند، اثر تشخیص را بهبود می‌بخشد. از NMS وزن دار برای ثبت عملکرد تشخیص در محیط چند هدفه استفاده می‌شود، بنابراین چارچوب تشخیص بهینه به دست می‌آید.

اگرچه YOLOv5 عملکرد عالی در زمینه های مختلف تشخیص اشیاء نشان داده است، اما به دلیل بهره‌گیری از مکانیسم استخراج ویژگی ضعیف همچنان با مشکلاتی در زمینه تشخیص دقیق اشیاء از تصاویر پهپادی مواجه است که در این مطالعه برای حل این مسئله ماژول Inception در معماری آن گنجانده می‌شود.

۲- روش تحقیق

۲-۱- مناطق موردمطالعه

اولین محدوده موردمطالعه شهر تیانجین در استان هوبئی چین قرارگرفته‌است که تصاویر مربوط به خودرو از

۱ Back Bone
۲ Data Augmentation
۳ Mosaic
۴ Adaptive Image Scaling Module
۵ Spatial Pyramid Pooling
۶ Upsampling
۷ Non-Maximum Suppression

این منطقه تهیه‌شده است. دومین محدوده موردمطالعه ما شهر کان در استان نرماندی فرانسه انتخاب‌شده است که تصاویر مربوط استخر از این منطقه اخذ شده است.

۲-۲- دیتا ست‌های مورداستفاده

تصاویر موردنظر از سایت www.Kaggle.com جمع‌آوری گردیده‌اند که شامل تصاویر پهپادی از پارکینگ‌ها و استخرها می‌باشد. تصاویر پهپادی استفاده‌شده برای این تحقیق شامل تصاویر پهپادی اخذشده از DJI Mini Se برای کلاس خودرو است که از محدوده‌های پارکینگ در شهر تیانجین چین اخذشده‌است و دارای قدرت تفکیک مکانی ۵ سانتی‌متر است. برای کلاس استخر نیز از تصاویر پهپادی DJI Mavic استفاده‌شده است که از مناطق ویلایی شهر کان فرانسه باقدرت تفکیک مکانی ۲۰ سانتی‌متر جمع آوری شده است.

۲-۳- مروری بر شبکه YOLO

YOLO مخفف عبارت You Only Look Once، به معنای **"شما فقط یک‌بار به تصویر نگاه می‌کنید"** هست. درواقع، این عبارت به همان قابلیت سیستم بینایی انسان اشاره دارد که با یک نگاه عمل تشخیص اشیا را انجام می‌دهد. بنابراین، سیستم تشخیص YOLO باهدف ارائه روشی مشابه کارکرد سیستم بینایی انسان طراحی‌شده است.

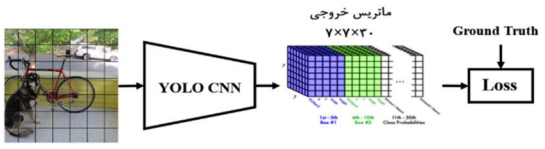
سامانه‌های تشخیص اشیای قبلی YOLO، از طبقه‌بندی کننده‌ها در کار تشخیص اشیا استفاده می‌کردند. این سامانه‌ها برای تشخیص یک شی، یک طبقه‌بندی کننده را در موقعیت‌ها و مقیاس‌های مختلف به تصویر ورودی اعمال می‌کردند. به‌عنوان‌مثال، سامانه‌هایی مانند Deformable Part Models یا DPM از پنجره‌های لغزان[۱] بهره می‌برند که طبقه‌بندی کننده را به موقعیت‌های مختلف در سراسر تصویر اعمال می‌کنند. این اعمال طبقه‌بندی کننده به موقعیت‌های مختلف تصویر، کار زمان‌بری است که البته شباهت چندانی هم به سیستم بینایی انسان در تشخیص اشیا ندارد.

YOLO معماری سامانه‌های تشخیص اشیا را دست‌خوش تغییراتی کرده‌است و به مسئله تشخیص اشیا به‌صورت یک مسئله رگرسیون می‌نگرد که مستقیم از پیکسل‌های تصویر به مختصات کادر و احتمال کلاس‌ها می‌رسد. با استفاده از سیستم YOLO، برای تشخیص اشیای موجود در تصویر، به هر تصویر شما فقط یک‌بار می‌نگرید. این رویکرد قابل مقایسه با رویکردهای DPM و R-CNN می‌باشد.

YOLO تنها یک شبکه پیچشی دارد که تصویر تغییریافته در ابعاد ورودی را دریافت و سپس به‌صورت هم‌زمان چندین کادر را به همراه احتمال کلاس‌ها پیش‌بینی می‌کند YOLO روی تصاویر آموزش داده می‌شود و مستقیماً کارآیی تشخیص را بهبود می‌دهد.

ساختار کلی الگوریتم YOLO در شکل ۱ نشان داده‌شده است. تصویر ورودی با ابعاد ۴۴۸x۴۴۸x۳ به یک Grid یا شبکه SxS تقسیم‌بندی می‌شود. این تصویر به شبکه YOLO داده می‌شود. خروجی شبکه پیچشی، ماتریسی به ابعاد SxSx۳۰ خواهد بود. هریک از درایه‌های ماتریس SxS خروجی معادل با یک سلول در شبکه SxS ورودی است .

خروجی SxSx۳۰ شامل مختصات کادرها و احتمال‌هاست. اگر در فرآیند آموزش[۲] باشیم، خروجی SxSx۳۰ به همراه کادرهای واقعی یا هدف[۳] به تابع اتلاف داده می‌شود. مقدار S در YOLO نسخه ۱، برابر با ۷ در نظر گرفته‌شده است. اگر در فرآیند آزمایش[۴] باشیم، خروجی SxSx۳۰ به الگوریتم حذف غیر حداکثرها داده می‌شود تا کادرهای ضعیف از بین بروند و تنها کادرهای درست در خروجی نمایش داده شوند.



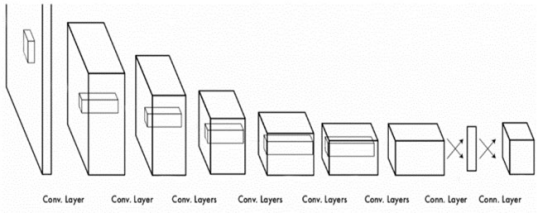
شکل ۱- ساختار CNN شبکه YOLO

YOLO شامل یک شبکه عصبی پیچشی با ۲۴ لایه پیچشی برای استخراج ویژگی و همچنین ۲ لایه کاملاً متصل[۵] برای پیش‌بینی احتمال و مختصات اشیا است که

۱ Sliding Window

۲ Train
۳ Ground Truth
۴ Test
۵ Fully Connected

معماری آن را در شکل ۲ مشاهده می‌کنید. همچنین، یک نسخه به روز شده از YOLO برای تشخیص سریع اشیا طراحی‌شده‌است.



شکل ۲- ساختار کلی شبکه YOLO

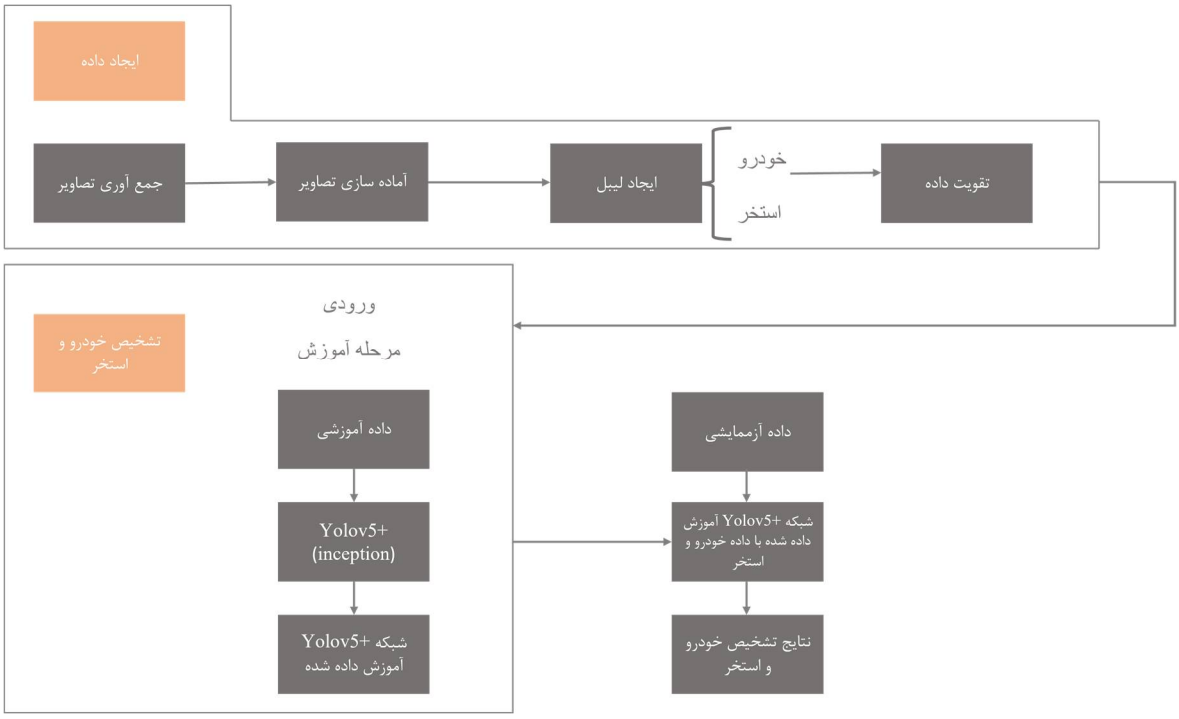
YOLO سریع، یک شبکه عصبی با تعداد لایه‌های پیچشی کمتر است که در آن از ۹ لایه پیچشی بجای ۲۴ لایه پیچشی اصلی استفاده‌شده و البته تعداد فیلترهای هر لایه در YOLO به روز رسانی شده نسبت به YOLO اصلی کمتر است. اندازه ورودی هر دو شبکه ۴۴۸x۴۴۸x۳ و خروجی شبکه نیز یک تنسور ۷x۷x۳۰ از پیش‌بینی‌ها است. در تمامی لایه‌ها از Leaky ReLU استفاده‌شده است.

YOLOv5 زمانی که با اندازه دسته بزرگ‌تری آزمایش شود، سرعت استنباط بالاتری نسبت به بسیاری از مدل‌های تشخیص‌دهنده دارد، ازاین‌رو در ادامه شبکه YOLOv5 را به‌عنوان شبکه هدف در نظر گرفتیم. YOLOv5 مدلی از خانواده مدل‌های بینایی کامپیوتری YOLO است. YOLOv5 معمولاً برای تشخیص اشیا استفاده می‌شود. YOLOv5 در چهار نسخه اصلی ارائه می‌شود: کوچک (s)، متوسط (m)، بزرگ (l) و فوق‌العاده بزرگ (x) که هرکدام نرخ دقت بالاتری را ارائه می‌دهند. هر نوع نیز زمان متفاوتی را برای آموزش نیاز دارد که در ادامه بیشتر به توضیح آن‌ها پرداخته خواهد شد.

YOLOv5 بر اساس چارچوب PyTorch پیاده سازی می‌شود. تفاوتی که بین YOLOv5 و نسخه‌های YOLO وجود دارد این است که YOLOv4 از cfg برای پیکربندی استفاده می‌کند درحالی‌که YOLOv5 از فایل yaml برای پیکربندی استفاده می‌کند.

۲-۴- روند نمای تحقیق

همان‌طور که در شکل ۳ مشاهده می‌شود در ابتدا به جمع‌آوری داده‌های موردنیاز پرداخته‌شده‌است و پس از آن تصاویر صحنه بزرگ به‌دست‌آمده را برای آموزش بهتر شبکه و یکسان‌سازی ابعاد تصاویر به تصاویر کوچک‌تر برش داده شده است.



شکل ۳- روندنمای تحقیق

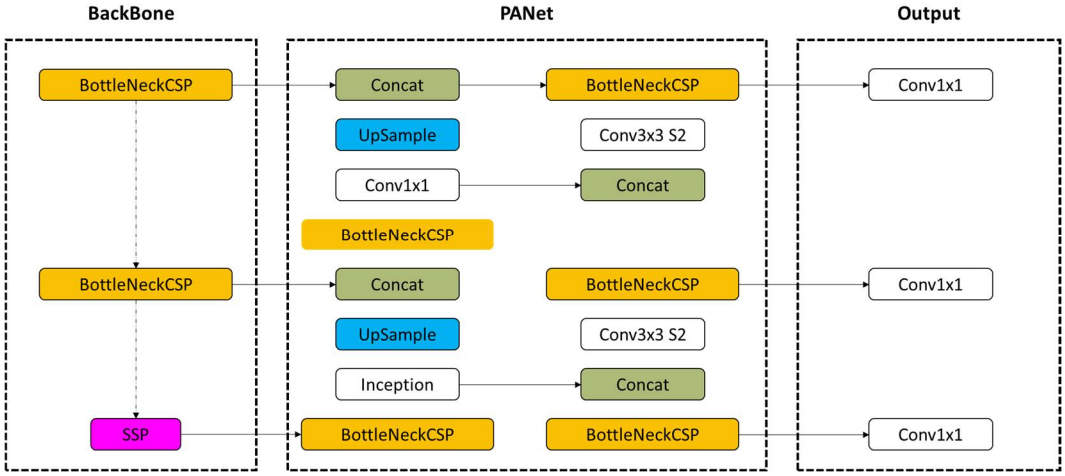
از روی این تصاویر ایجادشده برچسب‌های موردنظر را ایجاد شده است. برای افزایش بیشتر تصاویر آن‌ها با انجام عملیات‌هایی نظیرFlip، Brightness، Rotating و ... افزایش یافته است. سپس شبکه YOLOv5+ را با استفاده

از داده‌های آموزشی، تحت آموزش قرار گرفته و درنهایت برای آزمایش شبکه، از داده‌های آزمایشی استفاده شده است. داده‌های آموزشی و آزمایشی و اعتبار سنجی به ترتیب ۷۰، ۲۰ و ۱۰ درصد از داده‌ها اخذشده را تشکیل می‌دهند. درنهایت به بررسی نتایج شناسایی استخر و خودرو پرداخته‌شده است.

۲-۵- روند نمای کلی شبکه YOLOv5+

در تلاش برای بهبود تشخیص اشیاء در تصاویر پهپادی، به‌صورت نوآورانه‌ای شبکه YOLOv5 با اضافه کردن ماژول‌های Inception به لایه‌های آن ارتقاء داده‌شده است. یکپارچگی ماژول‌های Inception که به دلیل توانایی در استخراج ویژگی‌های چندمرحله‌ای شناخته می‌شوند، به‌عنوان یک ارتقاء مهم در ساختار YOLOv5 عمل کرده است. این افزایش، درک دقیق‌تری از داده‌های بصری پیچیده را فراهم می‌کند و به شبکه اجازه می‌دهد که الگوها و اشیاء پیچیده در داخل تصاویر پهپادی را با دقت بی‌سابقه تشخیص دهد.

تفاوتی که شبکه YOLOv5+ با شبکه متداول YOLOv5 دارد این است که در مرحله بعدBottleNeckCSP بجای Conv 1x1 از Inception استفاده‌شده است که بجای یک Conv 1x1 از پیچش‌های۳x۳، ۵x۵، ۷x۷ استفاده‌شده و با استفاده از یک ۳x۳ Max pooling ابعاد آن‌ها یکسان شده و در ادامه عملیات ادغام صورت گرفته‌شده است. این روند نما در شکل ۴ قابل‌مشاهده است. در ادامه پردازش موازی به شبکه این امکان را می‌دهد که جزئیات پیچیده‌تر را با کرنل کوچک‌تر استخراج کرده و ویژگی‌های پیچیده و غنی از تصویر را با کرنل‌های بزرگ‌تر استخراج کند. یکی از مزایای اصلی این رویکرد در قابلیت شناسایی اشیا در صحنه‌های پیچیده تصاویر پهپادی است. تصاویر پهپادی اغلب اشیایی با اندازه و اشکال مختلف، در برابر زمینه‌های پیچیده دارند. شبکه‌های رایج مانند YOLOv5 ممکن است در تشخیص دقیق این اشیاء به دلیل پیچیدگی آن‌ها قادر به شناسایی اشیا نباشند. بااین‌حال، شبکهYOLOv5+ که با ماژول‌های Inception ارتقا داده‌شده است، در مقابله با این پیچیدگی به‌مراتب با دقت بالاتری نسبت به شبکه YOLOv5 عمل می‌کند. علاوه بر دقت بالا، شبکه YOLOv5+ در کارایی نیز نتایج بهتری بجا گذاشته است. علیرغم پیچیدگی ایجادشده توسط ماژول‌های Inception، ساختار شبکه بهینه‌سازی شده‌است تا توانایی شبکه در شناسایی اشیا را افزایش دهد. این تعادل بین دقت و کارایی برای کاربردهای عملی همچون شناسایی و تحلیل تصاویر پهپادی به‌صورت آنی امری اساسی است. درنهایت، این رویکرد نوین، یعنی ادغام ماژول‌های Inception با اندازه‌های مختلف کرنل در شبکه YOLOv5یک پیشرفت قابل‌توجه در حوزه تشخیص اشیاء در تصاویر پهپادی را نمایان می‌کند. به دلیل مقابله مؤثر با چالش‌های شناسایی اشیاء متنوع و پیچیدگی‌های زمینه‌ای در تصویر، شبکه YOLOv5+ به‌عنوان یک راه‌حل قوی برای تشخیص اشیاء با دقت بالا در تصاویر پهپادی هست. از طریق آزمایش‌ها و تنظیمات دقیق، نه‌تنها دقت و کارایی شبکه قابل ارتقا است، بلکه مسیر را برای استفاده‌های بیشتر در زمینه‌هایی نظیر نظارت محیطی و برنامه‌ریزی شهری فراهم کرده‌است.



شکل ۴-روندنمای شبکه YOLOv5+

۲-۶- آماده‌سازی تصاویر

در این مرحله سعی بر این بود با برش تصاویر در ابعاد مساوی دیتاست را تولید کرده و تعداد تصاویر لازم برای ورودی شبکه فراهم شود. نمونه‌ای از مجموعه داده‌های آموزشی آماده‌شده در شکل ۵ قابل مشاهده می‌باشد.

شکل ۵- نمونه‌تصاویر آماده‌شده برای آموزش شبکه

۲-۷- استخراج برچسب‌ها

مرحله بعدی ایجاد برچسب‌ها از روی تصاویر به‌صورت فایلtext است که با استفاده از کتابخانه Labelme در محیط برنامه نویسی پایتون برچسب‌ها استخراج شده اند. در ادامه با استفاده از کتابخانه Albumentation افزایش داده را بر روی تصاویر اعمال شده است. با استفاده از این کتابخانه میتوان عملیات‌هایی نظیر Flip، Brightness، Rotating و ... را روی تصاویر اعمال کرد و باعث افزایش داده شد.

۲-۸- آموزش شبکه

برای سادگی این آموزش، مدل اندازه پارامترهای کوچک YOLOv5+ آموزش داده شده است، اگرچه می‌توان از مدل‌های بزرگ‌تر برای نتایج بهتر استفاده کرد. رویکردهای آموزشی متفاوتی ممکن است برای موقعیت‌های مختلف در نظر گرفته شود، و در اینجا رایج‌ترین روش های مورداستفاده پیاده‌سازی شده است. وقتی یک مجموعه داده بزرگ آموزش داده می‌شود از ابتدا بیشترین سود را خواهد داشت. وزن‌ها به‌طور تصادفی با ارسال یک string خالی (‘’) به آرگومان وزن‌ها مقداردهی اولیه می‌شوند. در ادامه با قرار دادن مقادیر ذیل برای ورودی‌های شبکه عملیات آموزش شبکه صورت گرفته است.

ازآنجایی‌که مجموعه داده مورد استفاده بسیار حجیم نیست و اشیاء زیادی در هر تصویر وجود ندارد، پس با کوچک‌ترین مدل یعنیYOLOv5s شروع به آموزش دادن داده‌ها شده‌است و از برازش بیش‌ازحد جلوگیری شده است. تعداد تصاویر وردی در هر اپوک ۳۲، اندازه تصویر ۳۲۰ و تعداد اپوک ۱۰۰ را برای آموزش شبکه در نظر گرفته شده است. شبکه شروع به آموزش دیدن می‌کند و در هر اپوک دوباره آموزش می‌بیند تا تعداد اپوکی که مشخص شده‌است به اتمام برسد. در شکل ۶ تصاویر آموزش‌دیده قابل‌مشاهده می‌باشند که با استفاده از Bounding box دور اشیا مشخص‌شده‌اند.

۲-۹- مقایسه صحت کلی نسخه‌های شبکه با مقایسه پارامترها

مدل YOLOv5 شامل ۱۰ معماری جداگانه به نام‌های YOLOv5n، YOLOv5s، YOLOv5m، YOLOv5l، YOLOv5x، YOLOv5n6، YOLOv5s6، YOLOv5m6، YOLOv5l6 و YOLOv5x6 است. بااین‌حال، معمولاً فقط چهار مورد اول برای تحقیق در نظر گرفته می‌شوند: کوچک، متوسط، بزرگ و خیلی بزرگ. تفاوت اصلی بین آن‌ها در تعداد ماژول‌های استخراج ویژگی و هسته‌های پیچش نهفته‌است و متعاقباً از منظر عملی در کار با مدل YOLO در تعداد پارامترهای شبکه عصبی مهم است.

در جدول ۱ نتایج آموزش دیتاست جمع‌آوری‌شده بر روی شبکه YOLOv5 و YOLOv5+ نمایش داده‌شده است. صحت کلی آموزش شبکه‌های YOLOv5 و YOLOv5+ بر روی دیتاست در تمامی نسخه‌ها بیشتر بوده و در جدول ۱ مشخص‌شده است. صحت کلی شبکه پیشنهادی (شبکه YOLOv5+) در نسخه Xlarge بیشتر از بقیه زیر مجموعه‌ها می‌باشد.

۲-۱۰- انتخاب بهترین زیر مجموعه شبکه

پس از آزمون‌وخطا کردن و بررسی پارامترهای مختلف به این نتیجه رسیدیم هرچقدر ظرفیت شبکه ما بیشتر باشد و همچنین با تعداد اپوک مطلوب شبکه را آموزش بدهیم، خروجی ما بهتر خواهد شد و صحت کلی

بالایی را به دست خواهیم آورد. همچنین همان‌طور که در جدول ۱ قابل‌مشاهده است می‌توان به این نتیجه رسید با استفاده از عمل Inception و افزودن کرنل‌های بیشتر می‌توان صحت کلی شبکه را افزایش داد و باعث بهبود شناسایی شبکه شد. همانطور که در جدول ۱ مشاهده می‌شود شبکه بهبود یافته نتایج بهتری نسبت به شبکه‌های مرسوم YOLOv5 و YOLOv7 در هر چهار نسخه کوچک، متوسط، بزرگ و خیلی بزرگ به دست آورده است. همچنین، بیشترین دقت شناسایی در زیرمجموعه خیلی بزرگ به دست آمده‌است.

جدول ۱- مقایسه صحت‌های میانگین بدست‌آمده برای دو هدف با استفاده از سه شبکه YOLOv5، YOLOv7 و Yolov5+

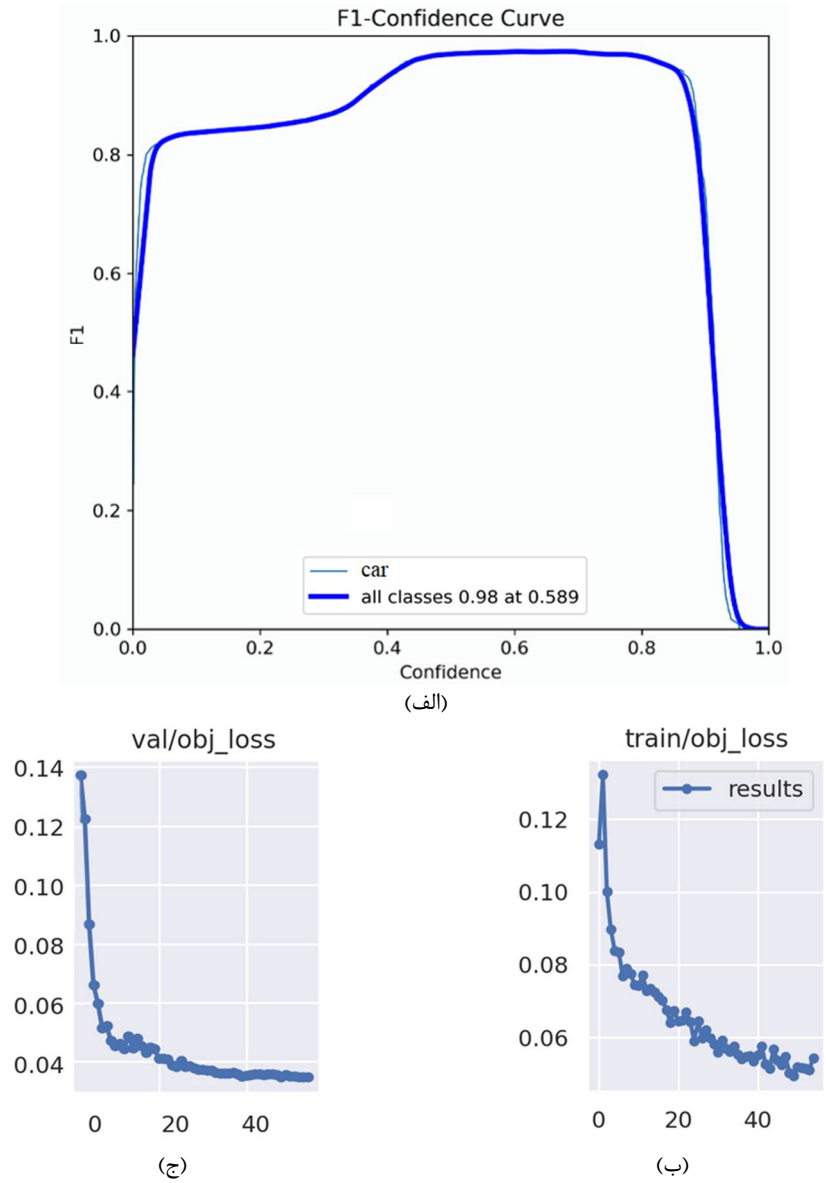
YOLOv5					YOLOv7					YOLOv5+				
OA	epochs	batch	img	Sub-version	OA	epochs	batch	img	Sub-version	OA	epochs	batch	Img	Sub-version
۸۸%	۵۵	۳۲	۳۲۰	Small	۸۶%	۵۵	۳۲	۳۲۰	Small	۸۹%	۵۵	۳۲	۳۲۰	Small
۸۹%	۵۵	۳۲	۳۲۰	Medium	۹۰%	۵۵	۳۲	۳۲۰	Medium	۹۱%	۵۵	۳۲	۳۲۰	Medium
۹۱%	۵۵	۳۲	۳۲۰	Large	۹۱%	۵۵	۳۲	۳۲۰	Large	۹۳%	۵۵	۳۲	۳۲۰	Large
۹۳%	۵۵	۳۲	۳۲۰	Xlarge	۹۳%	۵۵	۳۲	۳۲۰	Xlarge	۹۵%	۵۵	۳۲	۳۲۰	Xlarge

شکل۶- خروجی بصری نتایج مدل آموزش دیده توسط YOLOv5+

۳- ارزیابی و تحلیل نتایج

ارزیابی یک شبکه پیچش نیازمند بررسی دقت طبقه‌بندی آن بر روی کلاس‌های موردنظر است. به‌منظور ارزیابی دقت روش ارائه‌شده، قسمتی از مجموعه داده به داده‌های ارزیابی اختصاص‌یافته است. این داده‌ها پس از آموزش، مورد ارزیابی قرار می‌گیرد تا میزان اختلاف بین معماری‌های قبلی و معماری ارائه‌شده مشخص شود. این عملیات شامل چندین معیار برای ارزیابی است که از میانگین دقت کلی (Average Precision) با IoU (Intersection over Union) در بازه ۰/۵ تا ۰/۹۵ به‌عنوان معیار اصلی یاد می‌شود و طبق مقادیر AP بر روی‌داده‌های ارزیابی محاسبه می‌شود.

جدول شماره یک میانگین دقت به‌دست‌آمده از اعمال معماری ارائه‌شده و معماری YOLOv5+ با تعداد لایه‌ها و نرخ رشد متفاوت درIoU بین ۰/۵ تا ۰/۹۵ را نمایش می‌دهد. با بررسی الگوی رفتاری شبکه‌ها، عمل آموزش روان‌تر دنبال می‌شود. برای پایش وضعیت آموزش شبکه، به‌طور منظم و هر ده دقیقه یک با گزارشی از دقت متوسط هر یک از شبکه‌ها ذخیره می‌شود.



شکل ۷- الف) . نمودار F1-Curve، ب) نمودار تابع هزینه آموزش و ج) نمودار تابع هزینه اعتبارسنجی مربوط به کلاس خودرو

۳-۱- آموزش شبکه

برای تحلیل نمودار امتیاز[۱] F-1 در مرحله اول باید درک درستی از محور های این نمودار داشته باشیم. محور افقی نشان دهنده ی آستانه ی اطمینان شبکه ی مورد نظر می‌باشد. بسیاری از مدل های تشخیص اشیا با استفاده از معیار امتیاز اطمینان، اشیا شناسایی می‌کنند. محور افقی نمودار امتیاز F-1 مقادیر متفاوتی از ۰ تا ۱ را نشان می‌دهد. در ادامه محور عمودی میانگین هارمونیک صحت[۲] و پوشش[۳] می‌باشد و تعادل بین این دو معیار را نشان می‌کند.

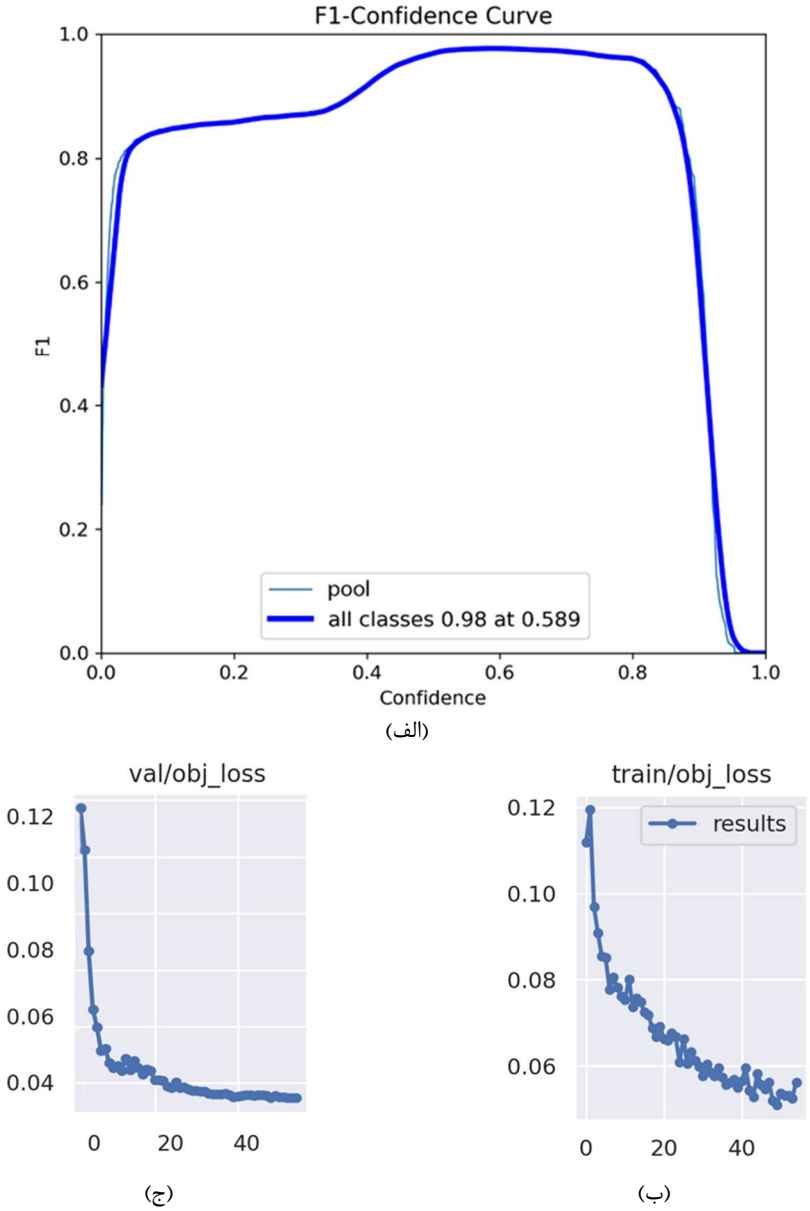
قله‌ها در این نمودار مناطقی را نشان می‌دهند که مدل به تعادل خوبی بین مقادیر صحت و پوشش دست یافته است. همانطور که در شکل ۷ الف و ۸ الف دیده میشود نقطه ی تلاقی هر دو کلاس خودرو و استخر های رو باز در محدوده ی ۰.۹ الی یک می‌باشند که نشانگر مورد اطمینان بودن آموزش شبکه می‌باشد و همینطور خروجی های مطلوبی به نمایش گذاشته شده است. همچنین ناحیه سمت چپ نمودار عموما پوشش بالا اما صحت کمتری دارند به همین دلیل با افزایش اپوک و مدت زمان آموزش شبکه همزمان میتوان به تعادل

۱ F-1 score
۲ Precision
۳ Recall

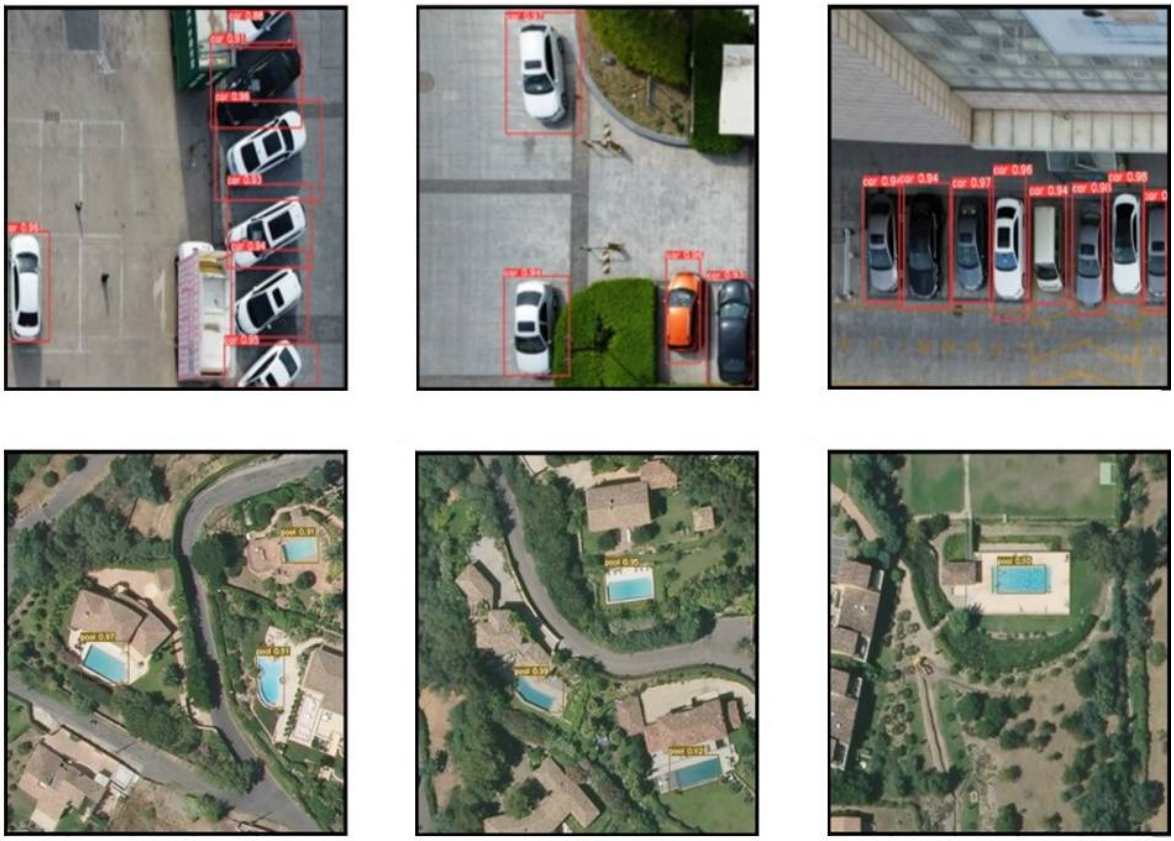
مطلوبی بین پوشش و صحت بالایی دست پیدا کرد و در نتیجه شبکه ی آموزش دیده میتواند اشیا را با صحت بالا شناسایی کند. ازآنجایی‌که داده‌های ما دو کلاس دارند، شناسایی اشتباه کلاسی بسیار کم است و خطای طبقه‌بندی به صفر نزدیک است.

تفسیر نمودار F1-Curve : امتیاز F1 (همچنین به‌عنوان F-measure یا F-score شناخته می‌شود) یک معیار خطا است (شکل ۷ و ۸، الف) که عملکرد مدل را با محاسبه میانگین هارمونیک دقت و یادآوری کلاس مثبت را اندازه‌گیری می‌کند. مقدار عددی F1 را می‌توان به‌عنوان معیاری برای عملکرد کلی مدل از ۰ تا ۱ تفسیر کرد که در آن ۱ بهترین عملکرد را نشان می‌دهد. به‌طور دقیق‌تر، مقدار F1 را می‌توان به‌عنوان توانایی متعادل مدل برای ثبت پوشش و صحت در مواردی که محاسبه می‌کند تفسیر کرد. مقدار عددی F1 از رابطه ۱ قابل‌محاسبه است:

$$F1 = (2precision * recall) / (precision + recall) \quad (۱)$$



شکل ۸- الف) نمودار F1-Curve، ب) نمودار تابع هزینه آموزش و ج) نمودار تابع هزینه اعتبارسنجی مربوط به کلاس استخر



شکل ۹- مجموعه تصاویر اشیا شناسایی‌شده توسط شبکه YOLOv5+

۳-۲- آزمایش شبکه

پس از این که مدل عملکرد قابل اطمینانی را در مرحله آموزش و صحت‌سنجی از خود نشان داد، عملکرد مدل بر روی داده‌های آزمایشی نیز مورد ارزیابی قرار گرفت. پارامترهای داده‌های آزمایشی مورداستفاده برای آزمایش شبکه به شرح زیر است:

source - مسیر ورودی تصاویر

weights - مسیر بهترین وزن آموزش‌دیده شده

img - اندازه تصویر برای آزمایش شبکه، برحسب پیکسل

conf - آستانه اطمینان

در شکل ۹ نتایج بصری مدل در استخراج اشیای هدف شامل خودرو و استخر نشان داده شده‌است. براساس شکل می‌توان گفت، شبکه پیشنهادی در استخراج تمامی اشیای هدف با موفقیت عمل کرده‌است که دلیل آن تقویت بخش استخراج ویژگی شبکه با ادغام ماژول کارآمد Inception در ساختار آن است؛ بنابراین، بطور کلی می‌توان گفت نتایج بصری مؤید صحت نتایج کمی حاصله در این تحقیق و همچنین پتانسیل بالای رویکرد پیشنهادی در استخراج عوارض هدف است.

۴- نتیجه‌گیری

استفاده از تصاویر سنجش‌ازدوری و الگوریتم‌های یادگیری عمیق در شناسایی خودروها و استخرهای آبی می‌تواند به‌عنوان یک‌راه حل مؤثر و اقتصادی برای مدیریت و پایش منابع و عوامل زیست‌محیطی مرتبط در نظر گرفته شود. در این تحقیق، با توجه به اهمیت شناسایی و طبقه‌بندی اشیا در تصاویر سنجش‌ازدوری، فرآیندی مبتنی بر شبکه‌های پیچشی ارائه گردیده است که در عین کاهش تعداد پارامترهای آموزشی، با استخراج بهتر ویژگی‌های مناسب طبقه‌بندی، دقت و سرعت روش‌های شناسایی اهداف را افزایش می‌دهد. افزایش تعداد لایه‌های پیچشی یکی از مؤلفه‌های اصلی در افزایش دقت طبقه‌بندی شبکه‌های پیچشی عمیق است؛ اما از طرفی افزایش بیش‌ازحد این لایه‌ها موجب محو شدن گرادیان و اطلاعات واردشده در لایه‌های انتهایی می‌گردد. این

موضوع در بسیاری از تحقیقات صورت گرفته درزمینه‌ی شبکه‌های پیچش مطرح‌شده است.

فرآیند آموزش معماری پیچش پیشنهادی توسط ۶۴۰ قطعه تصویر انجام‌شده‌است و پس‌ازآن، این معماری بر روی دو منطقه مطالعاتی شهر تیانجین و شهر کان مورد ارزیابی قرار گرفت که به ترتیب مقادیر ۹۵ و ۹۵ درصد برای معیار F1-Measure به‌دست‌آمده است. در راستای بهبود معماری ارائه‌شده در این مقاله می‌توان از اپراتورهای پیچش منبسط‌شده (Dilated Convolution) باهدف استخراج ویژگی‌های بارزتر استفاده کرد. از سوی دیگر باهدف توسعه و عمومی‌سازی هرچه بهتر روش ارائه‌شده، می‌توان به طبقه‌بندی انواع خودرو (اعم از خودروهای سواری، شاسی‌بلند و خودروهای سنگین) و همچنین طبقه‌بندی انواع استخر (اعم از استخرهای با مساحت بزرگ و کوچک) پرداخت.

مراجع

[۱] D. Zhao, D. Xie, F. Yin, L. Liu, J. Feng, and T. Ashraf, "Estimation of Pb Content Using Reflectance Spectroscopy in Farmland Soil near Metal Mines, Central China," *Remote Sensing,* vol. 14, no. 10, p. 2420, 2022.]Online[. Available: https://www.mdpi.com/2072-4292/14/10/2420.

[۲] Z. Chen *et al.*, "Automatic Estimation of Apple Orchard Blooming Levels Using the Improved YOLOv5," *Agronomy,* vol. 12, no. 10, p. 2483, 2022.]Online[. Available: https://www.mdpi.com/2073-4395/12/10/2483.

[۳] A. D. W. Sumari, A. S. Pranata, I. A. Mashudi, I. N. Syamsiana, and C. O. Sereati, "Automatic Target Recognition and Identification for Military Ground-to-Air Observation Tasks using Support Vector Machine and Information Fusion," in *2022 International Conference on ICT for Smart Society (ICISS)*, 10-11 Aug. 2022 2022, pp. 01-08, doi: 10.1109/ICISS55894.2022.9915256.

[۴] Z. Wang, X. Li, Y. Mao, L. Li, X. Wang, and Q. Lin, "Dynamic simulation of land use change and assessment of carbon storage based on climate change scenarios at the city level: A case study of Bortala, China," *Ecological Indicators,* vol. 134, p. 108499, 2022/01/01/ 2022, doi: https://doi.org/10.1016/j.ecolind.2021.108499.

[۵] Y. Liu *et al.*, "Study of the Automatic Recognition of Landslides by Using InSAR Images and the Improved Mask R-CNN Model in the Eastern Tibet Plateau," *Remote Sensing,* vol. 14, no. 14, p. 3362, 2022.]Online[. Available: https://www.mdpi.com/2072-4292/14/14/3362.

[۶] J. Meng, J. Yan, and J. Zhao, "Bubble Plume Target Detection Method of Multibeam Water Column Images Based on Bags of Visual Word Features," *Remote Sensing,* vol. 14, no. 14, p. 3296, 2022.]Online[. Available: https://www.mdpi.com/2072-4292/14/14/3296.

[۷] S. Jin, X. Li, X. Yang, J. A. Zhang, and D. Shen, "Identification of Tropical Cyclone Centers in SAR Imagery Based on Template Matching and Particle Swarm Optimization Algorithms," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 57, no. 1, pp. 598-608, 2019, doi: 10.1109/TGRS.2018.2863259.

[۸] S. Jian, J. Jiang, K. Lu, and Y. Zhang, "SEU-tolerant Restricted Boltzmann Machine learning on DSP-based fault detection," in *2014 12th International Conference on Signal Processing (ICSP)*, 19-23 Oct. 2014 2014, pp. 1503-1506, doi: 10.1109/ICOSP.2014.7015250.

[۹] 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.

[۱۰] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.

[۱۱] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 7-13 Dec. 2015 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[۱۲] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430,* 2021.

[۱۳] M. Kasper-Eulaers, N. Hahn, S. Berger, T. Sebulonsen, Ø. Myrland, and P. E. Kummervold, "Short Communication: Detecting Heavy Goods Vehicles in Rest Areas in Winter Conditions Using YOLOv5," *Algorithms,* vol. 14, no. 4, p. 114, 2021.]Online[. Available: https://www.mdpi.com/1999-4893/14/4/114.

[۱۴] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934,* 2020.

[۱۵] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767,* 2018.

[۱۶] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263-7271.

[۱۷] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.]Online[. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.91.

[۱۸] Y. Zhang, Z. Guo, J. Wu, Y. Tian, H. Tang, and X. Guo, "Real-time vehicle detection based on improved yolo v5," *Sustainability,* vol. 14, no. 19, p. 12274, 2022.

[۱۹] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "Dssd: Deconvolutional single shot detector," *arXiv preprint arXiv:1701.06659,* 2017.

[۲۰] W. Liu *et al.*, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, 2016: Springer, pp. 21-37.

[۲۱] Z. Chen, L. Cao, and Q. Wang, "Yolov5-based vehicle detection method for high-resolution UAV images," *Mobile Information Systems,* vol. 2022, 2022.

[۲۲] D. Yan *et al.*, "An improved faster R-CNN method to detect tailings ponds from high-resolution remote sensing images," *Remote Sensing,* vol. 13, no. 11, p. 2052, 2021.

[۲۳] S. Luo, J. Yu, Y. Xi, and X. Liao, "Aircraft target detection in remote sensing images based on improved YOLOv5," *Ieee Access,* vol. 10, pp. 5184-5192, 2022.

[۲۴] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems,* vol. 25, 2012.

[۲۵] Y. Da, X. Gao, and M. Li, "Remote Sensing Image Ship Detection Based on Improved YOLOv3," in *2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP)*, 15-17 April 2022 2022, pp. 1776-1781, doi: 10.1109/ICSP54964.2022.9778531.

[۲۶] C. Cao *et al.*, "Research on Airplane and Ship Detection of Aerial Remote Sensing Images Based on Convolutional Neural Network," *Sensors,* vol. 20, no. 17, p. 4696, 2020.]Online[. Available: https://www.mdpi.com/1424-8220/20/17/4696.

[۲۷] Z. Li, A. Namiki, S. Suzuki, Q. Wang, T. Zhang, and W. Wang, "Application of Low-Altitude UAV Remote Sensing Image Object Detection Based on Improved YOLOv5," *Applied Sciences,* vol. 12, no. 16, p. 8314, 2022.]Online[. Available: https://www.mdpi.com/2076-3417/12/16/8314.

[۲۸] Z. Wang, H. Lu, J. Jin, and K. Hu, "Human Action Recognition Based on Improved Two-Stream Convolution Network," *Applied Sciences,* vol. 12, no. 12, p. 5784, 2022.

[۲۹] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3-19.

[۳۰] L. Yang, G. Yuan, H. Zhou, H. Liu, J. Chen, and H. Wu, "RS-YOLOX: A High-Precision Detector for Object Detection in Satellite Remote Sensing Images," *Applied Sciences,* vol. 12, no. 17, p. 8707, 2022.]Online[. Available: https://www.mdpi.com/2076-3417/12/17/8707.

[۳۱] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11534-11542.

[۳۲] F. C. Akyon, S. O. Altinuc, and A. Temizel, "Slicing aided hyper inference and fine-tuning for small object detection," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022: IEEE, pp. 966-970.

[۳۳] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117-2125.

[۳۴] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A Deep Network Architecture for Pan-Sharpening," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 22-29 Oct. 2017 2017, pp. 1753-1761, doi: 10.1109/ICCV.2017.193.

[۳۵] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 37, no. 9, pp. 1904-1916, 2015.

[۳۶] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8759-8768.