

آشکار سازی هدف در تصاویر زمین پایه با مجموعه داده محدود مبتنی بر یادگیری عمیق

سید امین موسوی^{۱*}، مهدی نصیری^۲، محمد جواد ناطقی^۳

۱- کارشناسی ارشد، ۲- استادیار، ۳- دانشجوی دکتری، دانشگاه جامع امام حسین (ع)
(دریافت: ۱۴۰۲/۱۲/۱۱، بازنگری: ۱۴۰۳/۰۲/۰۳، پذیرش: ۱۴۰۳/۰۲/۱۲، انتشار: ۱۴۰۳/۰۳/۱۱)

چکیده

الگوریتم تشخیص شی YOLO، یکی از مشهورترین و پرکاربردترین الگوریتم‌های تشخیص اشیا در حوزه بینایی ماشین است. این الگوریتم با استفاده از تصاویر آموزشی فراوان، آموزش می‌بیند و می‌تواند در شناسایی اهداف عملکرد قابل توجهی داشته باشد. در شرایط خاص نظامی، پزشکی و محیط‌های خطرناک به دست آوردن تصاویر واقعی اهداف زمان‌بر و چالش‌برانگیز است. این چالش‌ها محققان را به دنبال راهکارهای جایگزین برای رفع مشکل کمبود داده و افزایش دقت الگوریتم می‌کند. یکی از راهکارهای جدید برای حل این مشکل، استفاده از داده‌های مصنوعی تولید شده توسط موتورهای بازی‌سازی است. هدف از این تحقیق، آشکار سازی هدف با استفاده از یادگیری عمیق در شرایط کمبود مجموعه داده آموزشی است. برای این منظور ابتدا مجموعه‌ای از داده‌های مربوط به هدف را با استفاده از موتور بازی‌سازی Unreal Engine، تولید کرده و سپس با استفاده از انواع الگوریتم‌های YoloV5، اقدام به آشکار سازی هدف مورد نظر گردید. الگوریتم‌های YoloV5x، YoloV5l، YoloV5m، YoloV5s، در این تحقیق مورد استفاده قرار گرفتند که به لحاظ معیار دقت به ترتیب نتایج ۸۵، ۸۴، ۸۲ و ۷۷ درصد به دست آمد. بر اساس نتایج، الگوریتم YoloV5l، بالاترین میزان دقت را به دست آورده که پیشنهاد می‌شود در هنگام استفاده از داده‌های مصنوعی، مورد استفاده قرار گیرد.

کلیدواژه‌ها: الگوریتم تشخیص شی، YOLO، داده‌های مصنوعی، Unreal Engine.

Target Detection Based on Deep Learning in Ground-Based Images with Limited Dataset

A. Mousavi¹, M. Nasiri², M.J. Nateghi³

Imam Hossein university, Tehran, Iran

(Received: 2024/03/01, Revised: 2024/04/22, Accepted: 2024/05/01, Published: 2024/05/31)

DOR: <https://dorl.net/dor/20.1001.1.26762935.1402.14.3.1.5>

Abstract

The YOLO object recognition algorithm is one of the most famous and widely used object recognition algorithms in the field of machine vision. This algorithm is trained by using a lot of training images and can have a significant performance in identifying targets. In military applications, medical and dangerous environments, obtaining real images of targets is time-consuming and challenging. These challenges make researchers look for alternative solutions to solve the problem of lack of data and increase the accuracy of the algorithm. One of the new ways to solve this problem is to use artificial data generated by game engines. The purpose of this research is to reveal the target using deep learning in the condition of lack of training dataset. For this reason, a set of data related to the target was first generated using the Unreal Engine, and then the generated target was revealed using various YoloV5 algorithms. Algorithms YoloV5l, YoloV5x, YoloV5m, YoloV5s were used in this research, and according to the accuracy criteria, the results were 85, 84, 82 and 77% respectively. Based on the results, the YoloV5l algorithm obtained the highest level of accuracy, which is suggested to be used when using artificial data.

Keywords: Object Recognition Algorithm, YOLO, Artificial Data, Unreal Engine.

۱. مقدمه

سهبعدی پیچیده است که زمان و تلاش زیادی می‌طلبد. همچنین، داده‌های ساختگی ممکن است نمایندگی کاملی از داده‌های واقعی نداشته باشند و مدل‌ها را به شرایط خاص و محدود ترغیب کنند. به‌طور کلی، استفاده از داده‌های ساختگی برای تشخیص اشیا یک راهکار قابل‌توجه است که محققان را قادر می‌سازد تا مدل‌های خود را در شرایط مختلف آموزش دهند و دقت و کارایی الگوریتم‌ها را بهبود بخشند. با پیشرفت تکنولوژی و ابزارهای مربوطه، می‌توان انتظار داشت که استفاده از داده‌های ساختگی در آینده از روش‌های رایج‌تر و پرکاربردتری برای تشخیص اشیا در بینایی رایانه باشد [۸].

۲. الگوریتم تشخیص شی YOLO

تشخیص اشیا در زمان واقعی به‌عنوان یک عنصر اساسی در کاربردهای متنوع ظاهر شده است که شامل حوزه‌های مختلفی مانند وسایل نقلیه خودران، روباتیک، نظارت تصویری و واقعیت افزوده می‌شود. یکی از چارچوب‌های برجسته در این حوزه، چارچوب YOLO است که به دلیل تعادل قابل‌توجه بین سرعت و دقت، مورد توجه قرار گرفته است و امکان شناسایی سریع و قطعی اشیا در تصاویر را فراهم می‌کند [۹]. خانواده YOLO طول زمان از طریق تکامل متعدد، ورژن‌های مختلفی را به خود اختصاص داده است.

اولین ورژن این چارچوب YOLO (You Only Look Once) نامیده می‌شود [۱۰]. این ورژن با تمرکز بر سرعت و قابلیت شناسایی به‌عنوان یک وظیفه یکپارچه، مورد استفاده قرار گرفت. YOLOv1 توسط جوزف ردمن و همکارانش [۱۱] پیشنهاد شد. این اولین آشکارساز یک مرحله‌ای در دوران یادگیری عمیق بود. ایده اصلی YOLOv1 این است که یک سلول شبکه‌ای به اندازه $S \times S$ (به طور پیش فرض 7×7) در یک تصویر اعمال شود. اگر مرکز یک شی در یک سلول شبکه باشد، آن سلول شبکه مسئول تشخیص آن شی است. بنابراین، سایر سلول‌ها حتی ظاهر شی که در سلول‌های دیگر آشکار شده است را نادیده می‌گیرند. معماری این نسخه در شکل (۱) نشان داده شده است [۱۲].

مدتی کوتاه پس از معرفی YOLOv1، ورژن YOLOv2 با نام YOLO9000 معرفی شد. این ورژن، علاوه بر بهبود سرعت و دقت، قابلیت شناسایی اشیا در تعداد بیشتری از کلاس‌ها را نیز داشت و می‌توانست از ۹۰۰۰ کلاس مختلف برای تشخیص اشیا استفاده کند. YOLOv2 دارای چندین پیشرفت نسبت به YOLO اصلی بود که آن را بهبود بخشید. در عین حال، سرعت اجرا حفظ شده و توانایی قوی‌تری در شناسایی ۹۰۰۰ دسته اشیا داشت [۱۳]. معماری این نسخه از YOLO در شکل (۲) نشان داده شده است [۱۴]. این نسخه از معماری Darknet، با استفاده از ۳۰ لایه کانولوشنی، بیشتر از لایه‌های YOLOv1 استفاده کرد.

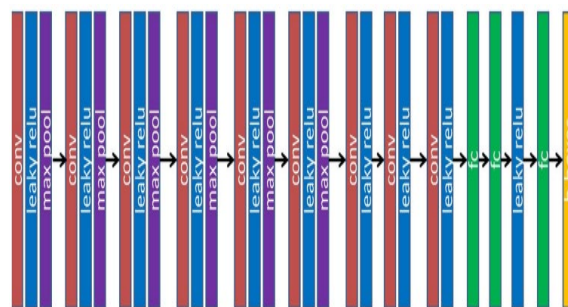
تشخیص اشیا یک برنامه کاربردی یادگیری عمیق است که معمولاً در بینایی کامپیوتر استفاده می‌شود. این برنامه به هوش مبتنی بر نرم‌افزار اجازه می‌دهد تا اشیا را در یک تصویر طبقه‌بندی، محلی‌سازی و شناسایی کند. روش سنتی این کار شامل آموزش یک مدل با داده‌های واقعی است که به‌صورت دستی گرفته شده و برچسب‌گذاری شده است [۱]. اما این روش زمان‌بر و چالش‌برانگیز است و محققان را به دنبال جایگزین‌های دیگری می‌کشاند [۲]. یکی از جایگزین‌های مورد استفاده برای جمع‌آوری و برچسب‌گذاری داده‌ها، استفاده از داده‌های مصنوعی است. این داده‌ها توسط کاربران تولید می‌شوند و امکان تولید تعداد زیادی داده در مدت زمان کوتاه را فراهم می‌کنند. همچنین، داده‌های مصنوعی به محققان امکان می‌دهند تا در مواردی که تکرار با داده‌های واقعی چالش‌برانگیز است، بتوانند در محیط‌های مجازی سفارشی شده، آموزش داده و الگوریتم‌ها را بهبود بخشند [۳].

الگوریتم تشخیص شی^۱ YOLO به دلیل داشتن توأم دقت و سرعت یکی از معروف‌ترین و پرکاربردترین الگوریتم‌های تشخیص شی در حوزه بینایی ماشین است. این الگوریتم با آموزش روی تصاویر آموزشی زیادی که به آن ارائه می‌شود، عملکرد قابل‌توجهی در شناسایی اهداف دارد [۴]. با این حال، در برخی از شرایط خاص مانند فضای نظامی، پزشکی و محیط‌های پرخطر، بدست آوردن تصاویر واقعی از اهداف مشکل است و محدودیت‌هایی وجود دارد که محققان را مجاب به جستجوی راهکارهای جدیدی برای رفع مشکل کمبود داده و افزایش دقت الگوریتم می‌کند.

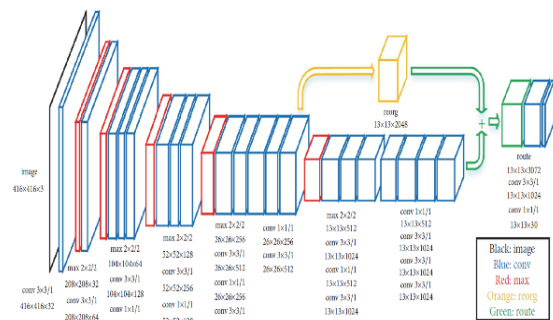
استفاده از داده‌های مصنوعی تولید شده توسط موتورهای بازی‌سازی یکی از راهکارهای جدید برای حل مشکل کمبود داده است. این روش نوآورانه با کاهش وابستگی به داده‌های واقعی، عملکرد و دقت الگوریتم را بهبود می‌بخشد [۵]. استفاده از داده‌های مصنوعی می‌تواند شامل داده‌های ساختگی شامل اشیا، صحنه‌ها و شرایط مختلف باشد. موتورهای بازی‌سازی معمولاً دارای قابلیت ساخت محیط‌های سهبعدی با اشیا و شرایط متنوع هستند. با استفاده از این موتورها و ابزارهای گرافیکی، می‌توان داده‌های ساختگی را با تنوع بالا و با دقت بالا تولید کرد [۶].

با استفاده از داده‌های ساختگی، می‌توان مدل‌های تشخیص اشیا را در شرایط مختلف و متنوع آموزش داد. به‌عنوان مثال، می‌توان شرایط نوری متفاوت، زوایای دید مختلف، اشیا در حال حرکت و تغییر شکل را شبیه‌سازی کرد [۷]. این کار به محققان امکان می‌دهد تا مدل‌های خود را برای مواجهه با شرایط واقعی مورد آزمایش و بهینه‌سازی قرار دهند. با این حال، استفاده از داده‌های ساختگی همچنان با برخی چالش‌هایی همراه است. یکی از چالش‌های اصلی، نیاز به طراحی و ساخت محیط‌های

^۱You Only Look Once



شکل ۳. معماری شبکه YOLOv3 که در آن از تابع Leaky ReLU به عنوان تابع فعال‌ساز استفاده شده است [۱۷].



شکل ۴. معماری YOLOv4 که در آن از شبکه CSPDarknet53 برای استخراج ویژگی‌های هدف استفاده می‌شود [۱۹].

YOLOv3 که توسط جوزف ردمون و علی فرهادی [۱۵] منتشر شد، این نسخه تغییرات قابل توجهی را به همراه داشت و یک معماری بزرگتر را برای دستیابی به عملکرد در زمان واقعی ارائه کرد. یکی از تغییرات اصلی آن، استفاده از یک استخراج کننده ویژگی پیشرفته بود که ترکیبی از YOLOv2 و شبکه Darknet-53 بود. شبکه Darknet-53 شامل یک ساختار با ۵۳ لایه کانولوشنی برای استخراج ویژگی‌های آموزشی استفاده می‌شد. سپس، به این استخراج کننده ویژگی، ۵۳ لایه دیگری اضافه شد که به طور خاص برای تشخیص اشیاء و آموزش آشکارسازی استفاده می‌شد [۱۶]. با اضافه کردن این لایه‌ها، معماری YOLOv3 در کل دارای ۱۰۶ لایه کانولوشنی شد که یک معماری کاملاً جدید مطابق شکل (۳)، ایجاد کرد [۱۷].

الگوریتم YOLO توسط جوزف ردمن، نویسنده چارچوب Darknet، توسعه داده شده است. پس از چندین سال تحقیق و توسعه، ردمن توسعه الگوریتم YOLO را متوقف کرده و از حوزه بینایی کامپیوتر خارج شده است. او نگران استفاده نادرست از تحقیقاتش در برنامه‌های نظامی بوده است. با این حال، تحقیقات بر اساس ایده‌های اولیه YOLO همچنان ادامه دارد. پس از مدتی، الکسی بوچکوفسکی به همراه همکارانش [۱۸]، معماری YOLOv4 را مطابق شکل (۴) بر اساس چارچوب Darknet و ایده‌های ردمن منتشر کردند. YOLOv4 سرعت بالا و دقت بالایی را در تشخیص اشیاء ارائه می‌دهد. با دقت متوسط ۴۳/۵ و سرعت ۶۵ فریم در ثانیه، این مدل یکی از سریع‌ترین و دقیق‌ترین مدل‌های تشخیص اشیاء است [۱۹].

۳. استفاده از داده‌های مصنوعی در آموزش شبکه

ظهور شبکه‌های عصبی کانولوشنی و شبکه‌های یادگیری عمیق در حوزه بینایی رایانه، نیاز به داده‌های تصویری بیشتر برای آموزش و آزمایش این شبکه‌ها را افزایش داده است. با این حال، روش‌های سنتی برای جمع‌آوری این داده‌ها ممکن است با چالش‌هایی مواجه شوند، به‌ویژه در شرایط خاصی مانند محیط‌های نظامی و پزشکی. برای جمع‌آوری داده‌های تصویری، روش‌های سنتی غالباً شامل گرفتن تصاویر از دنیای واقعی و سپس برچسب‌گذاری و حاشیه‌نویسی دستی بر روی آنها است. اما در شرایط خاصی، جمع‌آوری داده‌های تصویری از دنیای واقعی با مشکلاتی همراه است. این مشکلات می‌توانند شامل عدم دسترسی به تصاویر، خطرات جمع‌آوری تصاویر (مانند محیط‌های نظامی)، نیاز به حفظ حریم خصوصی، زمان‌بری و هزینه‌بر بودن جمع‌آوری تصاویر باشند [۲].

برای حل این مشکلات، استفاده از داده‌های مصنوعی می‌تواند راهکاری مناسب باشد. با استفاده از تکنیک‌های تولید داده مصنوعی، می‌توان داده‌های تصویری جدید را ایجاد کرده و به شبکه‌ها برای آموزش و آزمایش ارائه کرد [۵]. این روش به صورت گسترده در حوزه‌های دیگری مانند پردازش زبان طبیعی استفاده می‌شود و با نتایج قابل قبولی همراه بوده است. برای تولید داده‌های مصنوعی در حوزه بینایی رایانه، روش‌هایی مانند تولید تصاویر تصادفی، استفاده از موتورهای بازی‌سازی برای تولید تصاویر، استفاده از شبکه‌های مولد، تزویج تصاویر، تغییر ظاهر تصاویر و ترکیب داده‌های موجود می‌تواند مورد استفاده قرار گیرد. با استفاده از این تکنیک‌ها، می‌توان داده‌های تصویری متنوعی را ایجاد کرده و برای آموزش شبکه‌ها استفاده کرد. همچنین، می‌توان از تکنیک‌های تقویت داده؛ مانند تغییر اندازه، چرخش، تغییر روشنایی و اعمال فیلترهای تصویری استفاده کرد تا داده‌های موجود را تنوع بخشیم و افزایش دهیم [۲۲].

هرچند داده‌های مصنوعی می‌توانند به شبکه‌ها کمک کنند تا در شرایط نظامی بهتر عمل کنند، اما باید توجه داشت که دقت و کارایی آن‌ها به طور مستقیم به کیفیت داده‌های مصنوعی و تطبیق آن‌ها با شرایط واقعی وابسته است. همچنین باید در نظر داشت که داده‌های مصنوعی تنها بخشی از زمینه عمل شبکه‌ها هستند و عوامل دیگری مانند ساختار شبکه و الگوریتم‌های آموزش نیز بر کارایی آن‌ها تأثیر دارند.

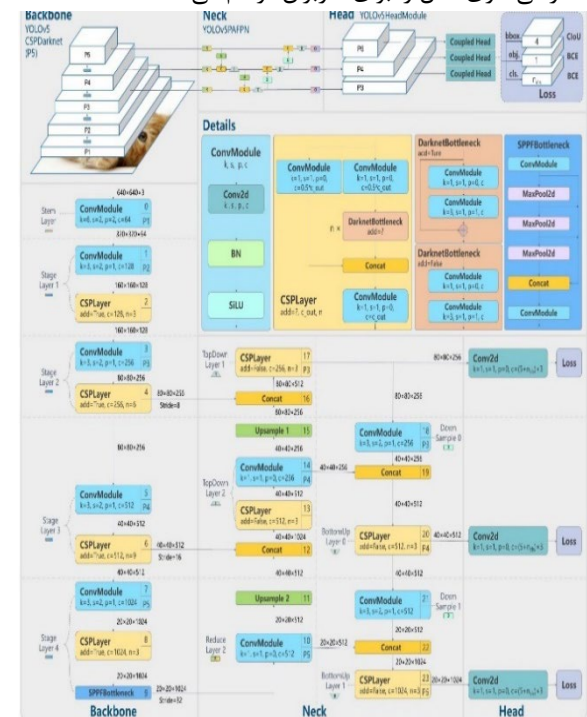
۴. مزایا و کاربردهای استفاده از داده مصنوعی

استفاده گسترده از موتورهای بازی‌سازی مانند Unity و Unreal Engine در سال‌های اخیر، برای تولید و جمع‌آوری داده‌های مصنوعی به منظور آموزش شبکه‌های عصبی، به‌ویژه شبکه‌های عصبی کانولوشنی که برای آموزش نیاز به تعداد زیادی تصاویر آموزشی دارند، افزایش یافته است. این موتورها همراه با

گردن (Neck): وظیفه این لایه‌ها استخراج نقشه‌های ویژگی مختلف از مراحل مختلف ستون فقرات است. در این معماری برای گردن از $SPPF^1$ و CSP-PAN اصلاح شده استفاده می‌کند.

۳. سر (Head): وظیفه تشخیص و تمایز اشیاء در تصاویر را بر عهده دارد. این بخش با استفاده از ویژگی‌های استخراج شده توسط بخش ستون فقرات، با استفاده از لایه‌های پیچشی، لایه‌های تمام متصل و توابع فعال‌سازی، بازیابی مربوط به موقعیت و دسته‌بندی اشیاء را انجام می‌دهد. به‌طور کلی سر شامل جعبه‌های مرزبندی با اطلاعات مربوط به موقعیت و دسته‌بندی اشیاء می‌شود.

این شبکه همانند دیگر نسخه‌های شبکه YOLO بر روی مجموعه داده COCO آموزش دیده است که شامل چهار معماری کوچک، متوسط، بزرگ و خیلی بزرگ است که معماری نسخه بزرگ این شبکه در شکل (۵) نشان داده شده است [۲۱]. یکی از ویژگی‌های برجسته YOLOv5، سرعت بالا است. این مدل قادر به تشخیص و برچسب‌گذاری اشیاء با سرعت بالا است که این ویژگی در کاربردهایی که نیاز به تشخیص در زمان واقعی دارند، بسیار مهم است. همچنین، YOLOv5 دقت بالایی در تشخیص اشیاء ارائه می‌دهد به دلیل استفاده از روش‌های بهبود یافته و بهینه شده است. علاوه بر این، YOLOv5 قابلیت آموزش و تنظیم برای کاربردهای خاص را نیز داراست. کاربران قادر خواهند بود این مدل را با استفاده از داده‌های خود آموزش دهند و به نیازهای خاص خود تنظیم کنند. این ویژگی انعطاف‌پذیری و قابلیت سفارشی‌سازی مدل را برای کاربران فراهم می‌کند.



شکل ۵. معماری نسخه بزرگ YOLOv5 که شامل ۴۶ میلیون پارامتر قابل یادگیری است [۲۱].

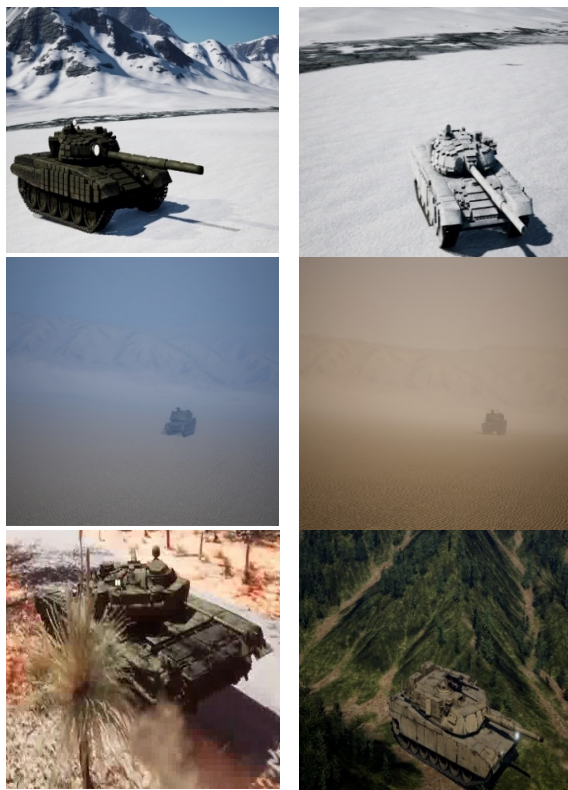
¹ Spatial Pyramid Pooling Fast

۵-۱. تولید تصاویر مصنوعی

طراحی و ساخت صحنه سه بعدی: در این مرحله، با استفاده از موتور بازی سازی Unreal Engine، صحنه های سه بعدی را طراحی خواهیم کرد. این شامل ساخت پس زمینه و اضافه کردن هدف در صحنه، نورپردازی و تنظیمات دیگر می شود.

تغییر پارامترها: در این بخش پارامترهای مختلفی را تغییر خواهیم داد؛ مانند زاویه دید دوربین، نوع نور، محیط های مختلف و غیره.

تولید تصویر: با استفاده از موتور Unreal Engine، قادر خواهیم بود تصاویر مصنوعی را باتوجه به صحنه و پارامترهای تنظیم شده تولید نماییم. در این مطالعه، شرایط احتمالی در زمان جنگ مانند قرارگیری تانک در محیط های کوهستانی، جنگلی، کویری، شهری، استتار، مه و گردوغبار مدل سازی گردید. با تنظیم پارامترهای مختلف مانند روشنایی، رنگ ها و جزئیات دیگر، تصاویر مصنوعی از این شرایط تولید شدند. نمونه هایی از این تصاویر در شکل (۷) به نمایش درآمده است.



شکل ۷. تصاویر تولید شده برای آموزش شبکه

۵-۲. آموزش و ارزیابی

دسته بندی تصاویر: به منظور ارزیابی عملکرد الگوریتم ها در دسته بندی تصاویر، مجموعه داده ها به سه بخش آموزشی، اعتبارسنجی و تست تقسیم می شود. در مجموعه داده آموزشی، تصاویری قرار می گیرند که برای آموزش الگوریتم ها به کار می رود. بخش اعتبارسنجی، به منظور ارزیابی و انتخاب بهترین مدل ها

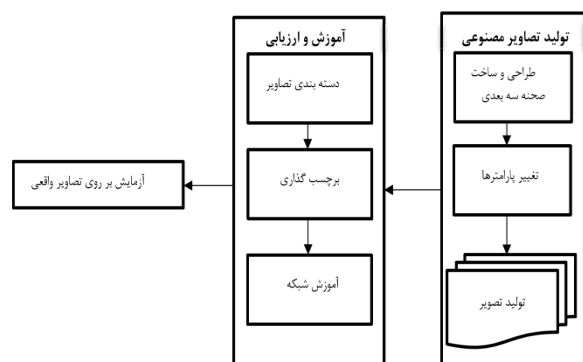
نرم افزارهای مدل سازی سه بعدی مانند Blender، به عنوان ابزارهای برتر در تحقیقات استفاده می شوند. آنها امکان ایجاد و استفاده آسان از محیط های مجازی سه بعدی را با تنظیمات پیش فرض و قابل تنظیم فراهم می کنند [۳].

موتور بازی سازی Unreal Engine، به دلیل قابلیت رندرینگ در زمان واقعی و برخورداری از محیط برنامه نویسی جداگانه [۲۳]، در بسیاری از موارد برای تولید تصاویر آموزشی در فرآیند آموزش شبکه های عصبی استفاده می شود. این موتور با برتری خود در زمینه رندرینگ در زمان واقعی و قابلیت ایجاد محیط برنامه نویسی جداگانه، امکان راه اندازی سریع و با کیفیت بالا برای محیط های مجازی را فراهم می کند و به ما امکان بهره برداری از این محیط ها را می دهد. از این رو، Unreal Engine به عنوان یک ابزار قوی و کارآمد در فرآیند آموزش شبکه های عصبی با تصاویر آموزشی مورد استفاده قرار می گیرد. همچنین، این موتور امکان استفاده از منابع موجود در آنها و ضبط صحنه ها با استفاده از دوربین های مجازی را نیز فراهم می کنند. با استفاده از نور مصنوعی و ویژگی های دیگر محیط های مجازی، به سرعت می توان به اهداف محیط های مجازی دست یافت.

قبلاً، موتورهای بازی بیشتر برای توسعه دنیای مجازی طراحی شده بودند. اما امروزه با توسعه تکنولوژی و استفاده از داده های مصنوعی، این موتورها قادر به تولید و جمع آوری داده های پیچیده و شبیه سازی های واقع گرا شده اند. با استفاده از این داده ها، می توان در زمینه های مختلفی مانند آموزش، تحقیقات علمی، توسعه بازی های ویدیویی و بسیاری موارد دیگر از مزایای استفاده از داده های مصنوعی بهره برداری کرد.

۵. روش پیشنهادی

این روش شامل طراحی صحنه های سه بعدی با استفاده از Unreal Engine و تولید تصاویر مصنوعی است. سپس مدل های تشخیص الگو با استفاده از این تصاویر آموزش می بینند و مورد آزمایش بر روی تصاویر واقعی قرار بگیرند. این طرح به صورت مراحل زیر می باشد و هم چنین در شکل (۶) به صورت گرافیکی نمایش داده شده است.



شکل ۶. نمایی از مراحل روش پیشنهادی

- تعداد تصاویر اعتبارسنجی: ۳۵۰ تصویر
- تعداد تصاویر تست: ۱۵۰ تصویر
- بهینه‌ساز: GSD
- تعداد دوره: ۱۵۰ (Epochs)
- اندازه دسته (Batch Size): ۱۶

۷. ارزیابی روش پیشنهادی و تحلیل نتایج

باتوجه به اینکه از تمامی نسخه‌های شبکه YOLOv5 استفاده شده است اطلاعات این نسخه‌ها در جدول (۱) به نمایش گذاشته شده است.

جدول ۱. اطلاعات معماری نسخه‌های استفاده شده از شبکه YOLO

mAPval 50	mAPval 50-95	تعداد پارامترها	اندازه (پیکسل)	نوع معماری
۵۶/۸	۳۷/۴	۷/۲ میلیون	۶۴۰	YOLOv5s
۶۴/۱	۴۵/۴	۲۱/۲ میلیون	۶۴۰	YOLOv5m
۶۷/۳	۴۹	۴۶/۵ میلیون	۶۴۰	YOLOv5l
۶۸/۹	۵۰/۷	۸۶/۷ میلیون	۶۴۰	YOLOv5X

محاسبه دقت در آستانه‌های مختلف برای تمامی دسته‌ها در نظر گرفته می‌شود که هر کدام از این معیارها با استفاده از معادلات (۱) تا (۵) محاسبه می‌شوند [۲۴].

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (۱)$$

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

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

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (۴)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (۵)$$

حد آستانه در معیار mAP به مقداری اشاره دارد که برای تصمیم‌گیری درباره تشخیص اشیا توسط شبکه عصبی مورد استفاده قرار می‌گیرد. در این بخش ابتدا برای حد آستانه از مقدار ۰/۵ استفاده شده است. این به این معنی است که شبکه عصبی تنها زمانی اشیا را تشخیص می‌دهد که احتمال تشخیص صحیح آنها بیشتر از این حد باشد. با تعیین حد آستانه ۰/۵ و ۰/۵۰/۹۵، هر شیء که احتمال تشخیص آن توسط شبکه بیشتر از این مقادیر باشد، به‌عنوان شیء تشخیص داده می‌شود.

پس از آموزش شبکه با ۱۵۰ دوره، نتایج حاصل برای نسخه‌های مختلف YOLOv5، یعنی YOLOv5s، YOLOv5m، YOLOv5l و YOLOv5X با استفاده از معیارهای ارزیابی Precision، Recall، mAP0.5:0.95 در شکل‌های (۸) تا (۱۱) به ترتیب برای هر یک از نسخه‌های YOLOv5 نشان داده شده است.

استفاده می‌شود و تصاویری که در آن قرار می‌گیرند، تاحدامکان نماینده تنوع و تممیم‌پذیری داده‌ها هستند. نهایتاً، بخش تست برای ارزیابی نهایی الگوریتم‌ها و اندازه‌گیری عملکرد آن‌ها در شرایط واقعی استفاده می‌شود.

برچسب‌گذاری: در این بخش، تصاویر پس از دسته‌بندی با استفاده از نرم‌افزارهای برچسب‌گذاری و با مدل برچسب‌گذاری بر حسب الگوریتم YOLO توسط کاربر انسانی برچسب‌گذاری می‌شوند. برچسب‌ها شامل موقعیت و گونه اشیا در تصویر هستند. این فرایند مهم است تا بتوان از اطلاعات موجود در تصاویر برای دسته‌بندی صحیح اشیا استفاده کرد و مدل‌های الگوریتمی را برای تشخیص و دسته‌بندی بهتر آموزش داد.

آموزش شبکه: پس از دسته‌بندی و برچسب‌گذاری تصاویر موردنیاز برای آموزش شبکه، مرحله بعدی شامل آموزش خود شبکه است. در این مرحله، تصاویر مصنوعی و برچسب‌های متناظر به‌عنوان داده‌های آموزشی برای شبکه مورد استفاده قرار می‌گیرند. آموزش شبکه یک فرایند مهم است که هدف آن ارتقای عملکرد شبکه و بهبود توانایی آن در تشخیص تصاویر است.

۵-۳. آزمایش بر روی تصاویر واقعی

هدف نهایی از آموزش شبکه عصبی با استفاده از داده‌های مصنوعی، توانایی تممیم‌پذیری شبکه بر روی داده‌های واقعی و نمایش عملکرد مطلوب در مقابل آن‌ها است. در این آزمایش، شبکه با استفاده از داده‌های مصنوعی آموزش داده می‌شود تا بتواند الگوها و ویژگی‌های مهم موجود در داده‌های واقعی را یاد بگیرد. هدف اصلی این آموزش، اطمینان حاصل کردن از قابلیت شبکه در تعامل با داده‌های واقعی و نمایش عملکرد بهینه در مقابل آن‌ها است. این آزمایش به ما این امکان را می‌دهد تا صحت و قابلیت استنتاج شبکه در مقابل داده‌های واقعی را بررسی کنیم و از عملکرد مطلوب آن اطمینان حاصل کنیم.

۶. جزئیات پیاده‌سازی

سخت‌افزار استفاده شده جهت تولید تصاویر مصنوعی با موتور بازی‌سازی Unreal Engine عبارت‌اند از:

- پردازنده: Intel Core(TM) i5-11400
- حافظه دسترسی تصادفی: 32GB
- پردازنده گرافیکی: Nvidia GeForce GTX 1070

برای آموزش و آزمایش شبکه از سخت‌افزار زیر استفاده شده است:

- پردازنده: Intel(R) Core(TM) i5-1135G7
- حافظه دسترسی تصادفی: 8GB
- پردازنده گرافیکی: Nvidia GeForce MX330

همچنین، پارامترهای آموزش شبکه به شرح زیر است:

- تعداد تصاویر آموزش: ۱۵۰۰ تصویر

در پایان، نتایج حاصل از آموزش شبکه YOLOv5 با ۱۵۰ دوره آموزش بر روی نسخه‌های مختلف این شبکه، به صورت خلاصه در جدول (۲) قابل مشاهده است.

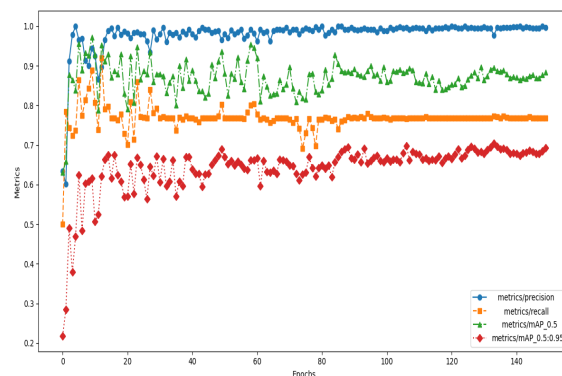
در شکل (۱۲)، نتایج عملکرد الگوریتم YOLOv5 بر روی تصاویر آزمایش به نمایش گذاشته شده است. این شکل نشان می‌دهد که الگوریتم YOLOv5 با امتیاز بالا توانسته با داده‌های مصنوعی آموزش دیده و تانک واقعی را با دقت شناسایی کند.

جدول ۲. نتایج ارزیابی مربوط به آموزش نسخه‌های مختلف شبکه YOLO نسخه ۵ بر روی مجموعه داده

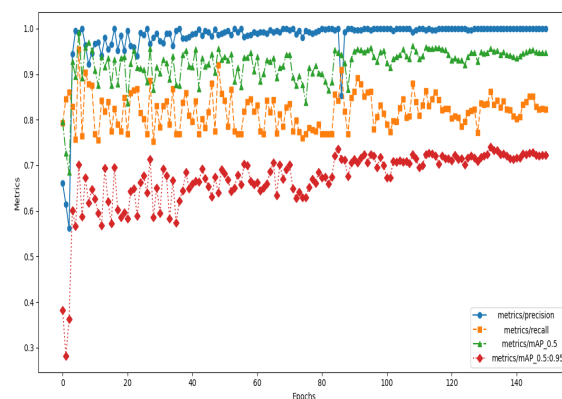
YOLOv5X	YOLOv5l	YOLOv5m	YOLOv5s	معیار ارزیابی
۰/۸۴	۰/۸۵	۰/۸۲	۰/۷۷	دقت
۰/۹۸۱	۰/۹۷۶	۱	۰/۹۷۷	درستی
۰/۸۸۶	۰/۸۸۴	۰/۸۶۲	۰/۷۷۳	فراخوانی
۰/۹۵۳	۰/۹۶۳	۰/۹۵۴	۰/۸۹۵	متوسط میانگین دقت ۰/۵۰
۰/۷۱۹	۰/۶۸۳	۰/۷۴	۰/۷۰۵	متوسط میانگین دقت ۰/۵۰ : ۰/۹۵
۰/۹۱۳	۰/۹۱۸	۰/۹	۰/۸۷	F1



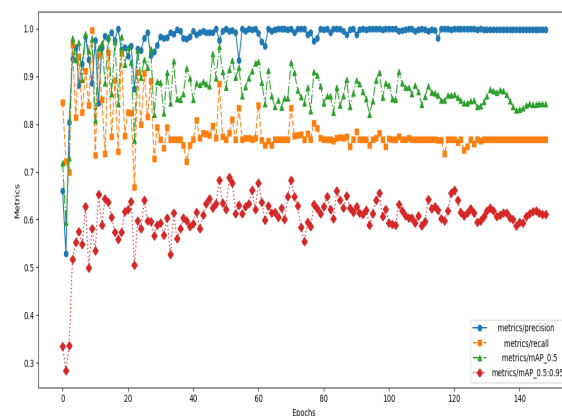
شکل ۱۲. عملکرد الگوریتم YOLOv5 بر روی تصاویر آزمایش



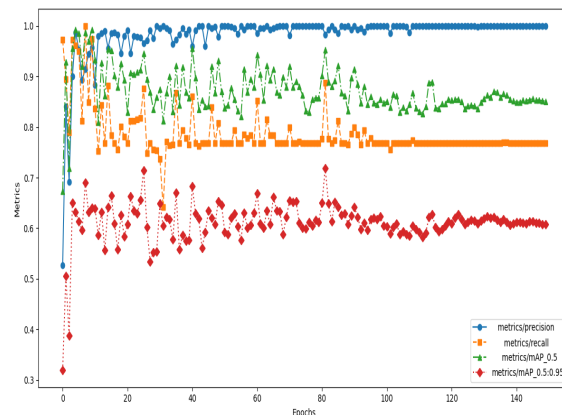
شکل ۸. نمودار معیارهای ارزیابی برحسب تعداد دوره برای YOLOv5s



شکل ۹. نمودار معیارهای ارزیابی برحسب تعداد دوره برای YOLOv5m



شکل ۱۰. نمودار معیارهای ارزیابی برحسب تعداد دوره برای YOLOv5l



شکل ۱۱. نمودار معیارهای ارزیابی برحسب تعداد دوره برای YOLOv5x

۸. نتیجه‌گیری

باتوجه به افزایش نیاز شبکه‌های عصبی تشخیص اشیا به تصاویر آموزشی، چالش‌هایی در این زمینه به وجود آمده است، به خصوص در حوزه نظامی که محدودیت‌های جدی در به دست آوردن تصاویر از اهداف وجود دارد. باتوجه به نیاز به تصاویر آموزشی بزرگ‌تر برای آموزش شبکه‌ها، استفاده از تصاویر مصنوعی تولید شده توسط موتور بازی‌سازی Unreal Engine که به واقعیت نزدیک هستند، می‌تواند به عنوان یک راه حل مناسب مورد استفاده قرار گیرد. با استفاده از این تصاویر، می‌توان شبکه‌های تشخیص اشیا را آموزش داد و نتایج حاصل از آموزش این شبکه‌ها با استفاده از داده مصنوعی را می‌توان با آزمایش بر روی داده‌های واقعی تأیید کرد. این پژوهش نشان می‌دهد که آموزش شبکه‌های تشخیص شی مانند YOLO با استفاده از تصاویر تولید شده توسط موتور بازی‌سازی Unreal Engine، بهبود قابل توجهی در عملکرد آنها در تشخیص اشیا واقعی ایجاد می‌کند. این روش نشان می‌دهد که با استفاده از داده‌های مصنوعی برای آموزش شبکه‌ها، می‌توان دقت و کارایی آنها در تشخیص اشیا واقعی را ارتقا داد. نوآوری این مقاله، استفاده از تصاویر تولید شده توسط موتور بازی‌سازی Unreal Engine به عنوان منبع آموزشی برای شبکه‌های تشخیص اشیا است. این رویکرد بهبود و پیشرفت قابل توجهی در عملکرد این شبکه‌ها در تشخیص اشیا واقعی به همراه دارد. این روش نوآورانه در تأمین داده‌های آموزشی برای شبکه‌های عصبی در حوزه تشخیص اشیا و سایر حوزه‌های مرتبط، قابلیت کارآمدی را ارائه می‌دهد و می‌تواند به توسعه و پیشرفت این حوزه‌ها کمک کند. با استفاده از این روش، می‌توان بانکی از انواع اهدافی که داده‌های واقعی از آنها نداریم یا برای دسترسی به این داده‌ها مسیر دشواری در پیش داریم، تولید کرد و در موارد شناسایی و یا طبقه‌بندی اهداف مورد استفاده قرار داد. این پژوهش همچنین نتایج کمی قابل توجهی را ارائه می‌دهد. به عنوان نمونه، معیار ارزیابی متوسط میانگین دقت ۰٫۵۰،۹۵ برای مدل‌های مختلف YOLO به ترتیب برابر با ۰٫۷۰، ۰٫۷۴، YOLOv5s، ۰٫۶۸، YOLOv5l و ۰٫۷۱ برای YOLOv5X است. این نتایج نشان‌دهنده بهبود عملکرد شبکه‌های تشخیص اشیا با استفاده از تصاویر تولید شده توسط موتور بازی‌سازی Unreal Engine است.

۹. مراجع‌ها

- [2] Jang, J.; Lee, H.; Kim, J.-C. "CarFree: Hassle-Free Object Detection Dataset Generation Using Carla Autonomous Driving Simulator"; Applied Sciences, 2021, DOI:10.3390/app12010281.
- [3] Man, K.; Chahl, J. "A Review of Synthetic Image Data and Its Use in Computer Vision"; Journal of Imaging, 2022, DOI:10.3390/jimaging8110310.
- [4] Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. "Object Detection in 20 Years: A Survey"; Proceedings of the IEEE, 2023, DOI:10.1109/jproc.2023.3238524.
- [5] Rasmussen, I.; Kvalsvik, S.; Andersen, P.-A.; Aune, T. N.; Hagen, D. "Development of a Novel Object Detection System Based on Synthetic Data Generated from Unreal Game Engine"; Applied Sciences, 2022, DOI:10.3390/app12178534.
- [6] Jaipuria, N.; et al. "Deflating Dataset Bias Using Synthetic Data Augmentation"; IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, DOI:10.1109/cvprw50498.2020.00394.
- [7] Aljundi, R.; Tuytelaars, T. "Lightweight Unsupervised Domain Adaptation by Convolutional Filter Reconstruction"; Computer Vision – ECCV 2016 Workshops, 2016, DOI:10.1007/978-3-319-49409-8_43.
- [8] Tsirikoglou, A.; Eilertsen, G.; Unger, J. "A Survey of Image Synthesis Methods for Visual Machine Learning"; Computer Graphics Forum, 2020, DOI:10.1111/cgf.14047.
- [9] Fayzi, A.; Dadashabadi, K.; Tohidi Far, A. "Detection and Tracking of Firearms and Cold Weapons Using Deep Learning Networks"; Journal of Advanced Defense Science & Technology, 2023. <https://dor.isc.ac/dor/20.1001.1.26762935.1401.13.4.2.1>
- [10] Farajollahi, A.; Rostami, M.; Parvin, H.; Nazarpour, B.; Lak, M. "Face Detection and Identification in Drones with Deep Learning"; Journal of Advanced Defense Science & Technology, 2022.
- [11] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. "You Only Look Once: Unified, Real-Time Object Detection"; 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, DOI:10.1109/cvpr.2016.91.
- [12] Sarda, A.; Dixit, S.; Bhan, A. "Object Detection for Autonomous Driving Using YOLO Algorithm"; International Conference on Intelligent Engineering and Management (ICIEM), 2021, DOI:10.1109/iciem51511.2021.9445365.
- [13] Redmon, J.; Farhadi, A. "YOLO9000: Better, Faster, Stronger"; 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, DOI:10.1109/cvpr.2017.690.
- [14] Liu, Z.; Chen, Z.; Li, Z.; Hu, W. "An Efficient Pedestrian Detection Method Based on YOLOv2"; Mathematical Problems in Engineering, 2018, DOI:10.1155/2018/3518959.
- [15] Kivrak, O.; Gürbüz, M. Z. "Performance Comparison of YOLOv3, YOLOv4, and YOLOv5 Algorithms: A Case Study for Livestock Recognition"; European Journal of Science and Technology, 2022, DOI:10.31590/ejosat.1111288.
- [16] Francies, M. L.; Ata, M. M.; Mohamed, M. A. "A Robust Multiclass 3D Object Recognition Based on Modern YOLO Deep Learning
- [1] Ali, S.; Abdullah, A.; Athar, A.; Ali, M.; Hussain, A.; Kim, H.-C. "Computer Vision-Based Military Tank Recognition Using Object Detection Technique: An Application of the YOLO Framework"; International Conference on Advanced Innovations in Smart Cities (ICAISC), 2023, DOI:10.1109/icaisc56366.2023.10085552.

- Algorithms”; Concurrency and Computation: Practice and Experience, 2021, DOI:10.1002/cpe.6517.
- [17] Yao, G.; Sun, Y.; Wong, M.; Lv, X. “A Real-Time Detection Method for Concrete Surface Cracks Based on Improved YOLOv4”; Symmetry, 2021, DOI:10.3390/sym13091716.
- [18] [Bahrami, M.; Asghari, A.; Binesh Marvasti, M.; Ansariyan, S. “Providing an Improved Unmanned Aerial Vehicle Detection System to Increase Detection Speed Using Deep Learning”; Electronic and Cyber Defense, 2023,41,81-96 <https://doi.isc.ac/dor/20.1001.1.23224347.1402.11.1.7.8>](#)
- [19] Qian, S.; Dai, S. “Identification of High-Speed Railway Tracksides Equipments Based on YOLOv4”; China Automation Congress (CAC), 2022, DOI:10.1109/cac57257.2022.10055473.
- [20] <https://github.com/ultralytics/yolov5>, Nov 2022.
- [21] <https://github.com/open-mmlab/mmyolo/tree/ma>, Aug 2023.
- [22] Shorten, C.; Khoshgoftaar, T. M. “A Survey on Image Data Augmentation for Deep Learning”; Journal of Big Data, 2019, DOI:10.1186/s40537-019-0197-0.
- [23] Freedman, E. “Game Engine Culture”; The Persistence of Code in Game Engine Culture, 2020, DOI:10.4324/9780429434242-1.
- [24] [Akbarian, H.; Sedaaghi, M. H. “Recognition of Acoustic Emitted from Surface Vessels Using MobileNet Convolutional Algorithm”; Journal of Advanced Defense Science & Technology, 2023, 51,39-50 <https://doi.isc.ac/dor/20.1001.1.26762935.1402.14.1.4.4>](#)