Sahand University of Technology

DOR: 20.1001.1.23223146.1404.12.1.5.8

Journal of Nonlinear Systems in Electrical Engineering

Vol.12, No.1

Spring and Summer 2025

ISSN: 2322 – 3146

http:// jnsee.sut.ac.ir

Vehicle Detection from Images Using a Modified YOLOv4 Algorithm with High Accuracy in Deep Learning

Javad Sayyadi, Mahdi nangir *, Behzad Mozaffari Tazekand, Mahmood Mohassel Feghh, Hamid sayyadi

Faculty of Electrical & Computer Engineering,University of Tabriz,Tabriz, Iran
Emails: J.sayyadi1400@ms.tabrizu.ac.ir -Nangir@tabrizu.ac.ir- Mozaffari@tabrizu.ac.ir-Mohasselfeghhi@tabrizu.ac.ir -H.sayyadi1400@ms.tabrizu.ac.ir

Abstract

Vehicle detection from images is one of the key applications of computer vision in fields such as military systems, urban surveillance, transportation, and safety. In this study, an improved version of the YOLOv4 algorithm is proposed for vehicle detection. The main innovation lies in the combination of two optimizers, ADAM and SGDM, to enhance the training speed and stability, along with the design of a lightweight post-processing algorithm, called IASD, for eliminating redundant and noisy bounding boxes. The proposed method was evaluated on a large-scale dataset, and the results showed that, compared with the original YOLOv4, significant improvements were achieved in both detection accuracy and speed. This outcome indicates that the proposed version can be a suitable option for real-time and resource-constrained systems such as UAVs and surveillance platforms. Moreover, the introduction of a lightweight and hardware-friendly post-processing algorithm distinguishes this research from similar studies. The results of this work can serve as a foundation for developing safer and more efficient applications in intelligent transportation and autonomous systems.

Keywords

Vehicle detection, YOLOv4, ADAM, SGDM, IASD algorithm

1. Introduction

Vehicle detection is a critical task in computer vision for applications like surveillance and autonomous systems. While YOLOv4 offers a balance of speed and accuracy, newer versions demand significant computational resources. This paper aims to enhance YOLOv4's performance for resource-constrained, real-time systems by addressing its training stability and post-processing efficiency. The proposed modifications focus on optimizing the learning process and refining detection outputs to achieve higher accuracy without heavy hardware reliance.

2. Proposed Work and Methodology (including comprision, simulation/experimental results and discusion)

The core of the proposed work involves two key innovations integrated into the YOLOv4 framework. First, a novel hybrid optimizer combining ADAM and SGDM is employed to accelerate training convergence and improve stability, as evidenced by a faster and lower training loss curve compared to the standard model. Second, a lightweight post-processing algorithm named IASD is designed to filter out redundant, noisy, and low-confidence bounding boxes through steps like confidence thresholding, Non-Maximum Suppression (NMS), and box merging. This significantly enhances detection precision.The method was evaluated on a large-scale dataset. Results demonstrate that the proposed YOLOv4+IASD model achieves a notable improvement over the original YOLOv4, increasing mAP (IoU=0.5) from 83.3% to 90.1% and mAP (IoU=0.5:0.95) from 66.0% to 68.4%, while maintaining a practical speed of 70 FPS on CPU-limited hardware. This performance is competitive with newer models like YOLOv5/v7/v8, which require GPUs for their superior FPS. The discussion highlights that this makes the proposed model a suitable and efficient choice for embedded and real-time applications such as UAVs and urban surveillance systems.

3. Conclusion

This paper presented an enhanced YOLOv4 algorithm for vehicle detection, achieving a significant balance between accuracy and computational efficiency. The integration of a hybrid ADAM-SGDM optimizer and the novel IASD post-processing algorithm led to measurable improvements in both mAP and training convergence speed. While newer YOLO variants excel on GPU hardware, the proposed method's lower resource requirements make it exceptionally viable for real-time, embedded applications like drones and intelligent transportation systems, bridging the gap between high performance and practical deployability.

تشخیص وسیله نقلیه از تصاویر با بهره‌گیری از الگوریتم تغییریافته YOLOv4 با دقت بالا در یادگیری عمیق

جواد صیادی[1]، مهدی نان‌گیر[2]، بهزاد مظفری تازه کند[3]، محمود محصل فقهی[4]، حمید صیادی[5]

۱- فارغ التحصیل کارشناسی ارشد، دانشکده مهندسی برق و کامپیوتر- دانشگاه تبریز- ایران

j.sayyadi1400@ms.tabrizu.ac.ir

۲- دانشیار، دانشکده مهندسی برق و کامپیوتر- دانشگاه تبریز- ایران

nangir@tabrizu.ac.ir

۳- استاد، دانشکده مهندسی برق و کامپیوتر- دانشگاه تبریز- ایران

Mozaffary@tabrizu.ac.ir

۴- دانشیار، دانشکده مهندسی برق و کامپیوتر- دانشگاه تبریز- ایران

mohasselfeghhi@tabrizu.ac.ir

۵- فارغ التحصیل کارشناسی ارشد، دانشکده مهندسی برق و کامپیوتر- دانشگاه تبریز- ایران

h.sayyadi1400@ms.tabrizu.ac.ir

DOR: 20.1001.1.23223146.1404.12.1.5.8

نشریه سامانه‌های غیرخطی در مهندسی برق

دوره ۱۲ - شماره ۱

بهار و تابستان ۱۴۰۴

صفحات ۱۰۰ الی ۱۱۷

ISSN: 2322-3146

http:// jnsee.sut.ac.ir

تاریخ ارسال: ۱۴۰۴/۴/۲۷ تاریخ بازنگری: ۱۴۰۴/۷/۴ تاریخ پذیرش: ۱۴۰۴/۷/۱۳

چکیده

تشخیص وسایل نقلیه از تصاویر یکی از کاربردهای مهم بینایی ماشین در حوزه‌هایی چون نظامی، نظارت شهری، حمل‌ونقل و ایمنی است. در این پژوهش یک نسخه بهبودیافته از الگوریتم YOLOv4 برای تشخیص وسایل نقلیه ارائه می‌شود. نوآوری اصلی در ترکیب دو بهینه‌ساز ADAM و SGDM برای افزایش سرعت و پایداری آموزش و طراحی یک الگوریتم پس‌پردازش سبک‌وزن باعنوان IASD برای حذف جعبه‌های تکراری و نویزی است. روش پیشنهادی بر روی یک مجموعه‌داده بزرگ ارزیابی شد و نتایج نشان داد که در مقایسه با YOLOv4 اصلی، هم از نظر دقت و هم از نظر سرعت تشخیص، بهبود قابل توجهی حاصل شده است. این دستاورد نشان می‌دهد که نسخه پیشنهادی می‌تواند گزینه‌ای مناسب برای سامانه‌های بلادرنگ و کم منبع مانند پهپادها و سیستم‌های نظارتی باشد. از سوی دیگر، ارائه یک الگوریتم پس‌پردازش سبک و قابل پیاده‌سازی در سخت‌افزارهای محدود، این پژوهش را از کارهای مشابه متمایز می‌سازد. نتایج این مطالعه می‌تواند به‌عنوان مبنایی برای توسعه کاربردهای ایمن‌تر و کارآمدتر در حوزه حمل‌ونقل هوشمند و سامانه‌های خودکار مورد استفاده قرار گیرد.

واژه‌های کلیدی

خیص وسیله نقلیه،

YOLOv،

ADAM،

SGDM،

گوریتم IASD

۱- مقدمه

در دهه اخیر، با پیشرفت فناوری و توسعه روش‌های یادگیری عمیق، تشخیص اشیا در تصاویر به یکی از مسائل مهم در حوزه هوش مصنوعی تبدیل شده است. تشخیص وسایل نقلیه از تصاویر، به‌خصوص از منابع هوایی که برای مانیتورینگ و نظارت در محیط‌های با دسترسی محدود استفاده می‌شوند، یکی از موضوعات مهم در این حوزه محسوب می‌شود[۱]. در این مقاله، به معرفی دو روش برای تشخیص وسایل نقلیه از تصاویر با استفاده از الگوریتم‌های تغییریافته YOLOv4 در یادگیری عمیق می‌پردازیم. YOLOv4 یکی از روش‌های معروف و قدرتمند در زمینه تشخیص اشیا است که با استفاده از شبکه‌های عصبی پیچشی، قادر است اشیا مختلف را با دقت و سرعت بالا تشخیص دهد[۲].

در این پژوهش، الگوریتم YOLOv4 با هدف تشخیص وسایل نقلیه در تصاویر بهبود یافته و نسخه‌ای ارتقاءیافته از آن ارائه شده است. این الگوریتم تغییریافته، با بهره‌گیری از تکنیک‌های پیشرفته‌تر در آموزش و بهینه‌سازی شبکه‌های عصبی، دقت و سرعت بیشتری را در تشخیص وسایل نقلیه ارائه می‌دهد. با وجود معرفی نسخه‌های جدید YOLO مانند YOLOv7, YOLOv8 و اخیراً YOLOv13، انتخابYOLOv4 به دلیل تعادل میان دقت، سرعت و نیازمندی سخت‌افزاری در سامانه‌های تعبیه‌شده همچنان رایج است.

در این مقاله، ابتدا مسئله تشخیص وسیله نقلیه از تصاویر معرفی می‌شود. سپس الگوریتم تغییریافته YOLOv4 توضیح داده شده و آزمایش‌ها و نتایج حاصل ارائه می‌گردد. در پایان نیز نتایج مقایسه و جمع‌بندی می‌شوند. این مقاله می‌تواند به عنوان یک مبنای قوی برای تحقیقات آتی در زمینه تشخیص وسایل نقلیه از تصاویر با استفاده از یادگیری عمیق مورد استفاده قرار گیرد.

در این الگوریتم تغییریافته، دو نوآوری ارائه شده است. اولین نوآوری مربوط به تابع پس‌پردازش است که به YOLOv4 اضافه شده است. این تابع پس از تشخیص اولیه وسایل نقلیه، اقدام به اصلاح و بهبود نتایج می‌کند. این تابع شامل مراحلی مانند حذف تشابه‌های متوالی، ادغام جعبه‌های مجاور و حذف برچسب‌های تکراری است. این تابع باعث افزایش دقت و سرعت تشخیص می‌شود و نتایج بهتری را ارائه می‌دهد[۳].

دومین نوآوری ارائه شده در این مقاله مربوط به تنظیمات الگوریتم‌های ADAM و SGDM است که سبب افزایش دقت و سرعت درYOLOv4 می‌شود. الگوریتم ADAM یک الگوریتم بهینه‌سازی است که برای بهبود سرعت یادگیری شبکه‌های عصبی مورد استفاده قرار می‌گیرد[۴]. همچنین، الگوریتم SGDM یک ترکیب از الگوریتم‌های استوانه‌ای و نزولی گرادیان است که باعث رشد چشمگیر سرعت و دقت الگوریتم می‌شود[۵].

پژوهش حاضر، نتایج آزمایش‌های خود را بر روی یک مجموعه داده بزرگ از تصاویر نشان می‌دهد. نتایج بیانگر آن است که الگوریتم تغییریافته باعث بهبود قابل ملاحظه دقت و سرعت تشخیص وسیله نقلیه می‌گردد. با توجه به این نکات، در این مقاله به بررسی و ارائه الگوریتم تغییریافته‌ای برای تشخیص وسایل نقلیه از تصاویر می‌پردازیم. این الگوریتم با استفاده از تقسیم تصویر به مناطق کوچک و اختصاص برچسب به هر منطقه، قادر است نوع و موقعیت وسیله نقلیه را با دقت بالا شناسایی نماید. مزایای استفاده از این الگوریتم شامل دقت و سرعت بالا، قابلیت تشخیص چند وسیله نقلیه همزمان و توانایی استفاده در برنامه‌های وسیعی از جمله نظامی، حمل‌ونقل و نظارت شهری است. الگوریتم تغییریافته در این پژوهش توانسته است نتایج قابل قبولی را در تشخیص وسایل نقلیه از تصاویر داشته باشد و همچنین می‌تواند به عنوان یک راه‌حل کارآمد در بسیاری از برنامه‌های کاربردی استفاده شود[۶].

با وجود معرفی نسخه‌های جدید YOLO مانند YOLOv7 و YOLOv8، انتخاب YOLOv4 همچنان در بسیاری از سامانه‌های تعبیه‌شده و کاربردهای صنعتی رایج است، زیرا تعادلی مناسب میان سرعت، دقت و نیازمندی سخت‌افزاری برقرار می‌سازد. نوآوری اصلی این پژوهش در دو بخش خلاصه می‌شود: نخست، ترکیب دو الگوریتم بهینه‌ساز ADAM و SGDM به‌گونه‌ای که مزایای هر دو در آموزش شبکه حفظ شود و هم‌زمان دقت و سرعت افزایش یابد. دوم، معرفی یک تابع پس‌پردازش سبک‌وزن با عنوان IASD که قادر است جعبه‌های محدودکننده تکراری و نویزی را حذف کرده و دقت تشخیص را بهبود دهد. این دو رویکرد، الگوریتم YOLOv4 را برای شرایط عملکرد بلادرنگ در سامانه‌های کم‌منبع بهینه می‌سازد و آن را از نسخه‌های استاندارد متمایز می‌کند.

۲- پیشینه پژوهش

مقاله[۶] به بهبود دقت شناسایی اشیا در شبکه‌های عصبی کانولوشنی می‌پردازد. هدف اصلی طراحی این مقاله، ایجاد یک سرعت عملیاتی بالا، برای یک تشخیص دهنده اشیا در سیستم‌های تولید و بهینه‌سازی برای محاسبات موازی است که با شاخص نظری حجم کم محاسبات در تضاد است. هدف این مطالعه همچنین بررسی تأثیر روش‌های جدید Bag-of-Freebies و-Bag-of Specialsدر طول آموزش شناسایی اشیا توسط تشخیص دهنده است.

مقاله[۷] یک الگوریتم تشخیص ماسک صورت و تشخیص استاندارد پوشیدن ماسک بر اساس YOLO-v4 بهبودیافته را پیشنهاد می‌دهد. نویسندگان یک CSPDarkNet53[1] بهبودیافته را در شبکه عصبی به عنوان ورودی اصلی معرفی می‌کنند که هزینه محاسباتی را کاهش می‌دهد و توانایی یادگیری را بهبود می‌بخشد. آن‌ها همچنین الگوریتم تغییر اندازه تصویر را قابل‌تطبیق معرفی می‌کنند تا به طور مؤثر محاسبات تکراری را کاهش دهد.

YOLOv4-tiny یک روش تشخیص سریع اشیا بر پایه YOLOv4-tiny است که ساختار شبکه را ساده‌تر می‌کند و پارامترهای محاسباتی را کاهش می‌دهد. این روش از دو ماژول ResBlock-D به‌جای دو ماژول CSPBlock در YOLOv4-tiny[2] استفاده می‌کند که پیچیدگی محاسباتی را کاهش می‌دهد. همچنین یک بلوک کمکی طراحی شده است تا اطلاعات بیشتری از اشیا استخراج کند و خطاهای تشخیص را کاهش دهد[۸].

در پژوهش حاضر [۹] روش تشخیص کلاه ایمنی بر اساس YOLOv4 بهبودیافته بررسی می‌شود. این روش با استفاده از الگوریتم K-means[3]، داده‌ها را خوشه‌بندی کرده و اطلاعات لبه‌ای هدف را به دست می‌آورد. در فرایند آموزش شبکه از استراتژی آموزش چند مقیاسی استفاده می‌شود تا قابلیت سازگاری مدل با مقیاس‌های مختلف تشخیص، افزایش یابد. نتایج آزمایش‌ها نشان می‌دهند که مقدار MAP[4] مدل به ۹۲.۸۹٪ می‌رسد، سرعت تشخیص به ۱۵ فریم بر ثانیه می‌رسد که دقت و سرعت تشخیص آن نسبت به YOLOv4 بهبود یافته‌اند.

[1] Cross-Stage Partial DarkNet 53
[2] You Only Look Once version 4-tiny
[3] Kluster Means
[4] mean Average Precision

در این مقاله[۱۰] یک روش تشخیص برای شکاف‌های پل بر اساس YOLOv4-FPM با ماژول Feature Pyramid پیشنهاد می‌شود تا به طور مؤثر شکاف‌ها در سازه‌های پل شناسایی شوند. YOLOv4-FPM از چارچوب تشخیص اشیا YOLOv4 به همراه ماژولFeature Pyramid برای افزایش عملکرد تشخیص شکاف‌های کوچک و بزرگ استفاده می‌کند.

این روش از مجموعه دادهٔ بزرگ از تصاویر پل، حاوی برچسب‌های شکاف برای آموزش شبکه YOLOv4-FPM استفاده می‌کند. شبکه با استفاده از فرایند دو مرحله‌ای، ابتدا در یک مجموعه داده بزرگ آموزش می‌بیند و سپس در مجموعه داده شکاف پل، به طور دقیق تنظیم می‌شود. این فرایند به شبکه کمک می‌کند تا ویژگی‌های شکاف را یاد بگیرد و دقت تشخیص را بهبود بخشد. برای مدیریت تغییرات مقیاس شکاف‌های پل، ماژول Feature Pyramid به معماری YOLOv4 اضافه شده است. این ماژول ویژگی‌های چند مقیاس را از سطوح مختلف استخراج می‌کند و تشخیص شکاف‌ها در اندازه‌ها و رزولوشن‌های مختلف را ممکن می‌سازد. این ماژول توانایی شبکه در ثبت اطلاعات شکاف‌های محلی و سراسری را تقویت می‌کند[۱۰].

این پژوهش[۱۱] بر روی استفاده از شبکه‌های تولیدی مبارزه‌ای[1]برای بهبود تشخیص علائم ترافیک تمرکز دارد. این مطالعه تصاویر مصنوعی را با تصاویر اصلی ترکیب کرده و از این روش برای افزایش مجموعه داده‌ها و ارزیابی اثر بخشی مجموعه داده‌های مصنوعی استفاده کرده است. پژوهشگران از تعداد و اندازه‌های مختلفی از تصاویر برای آموزش استفاده کرده‌اند و از شاخص شباهت ساختاری (SSIM[2]) و خطای میانگین مربعات (MSE[3]) برای ارزیابی کیفیت تصویر استفاده می‌کنند. پژوهشگران مجموعه داده تصاویر اصلی را با تصاویر مصنوعی گسترش داده و مدل تصویر اصلی را با مدل تصویر سنتز شده مقایسه کردند.

پس از ترکیب تصویر واقعی با تصویر سنتز شده توسط LSGAN[4]، عملکرد تشخیص به صورت قابل ملاحظه‌ای بهبود یافت و دقت آن در YOLOv3 به ٪ ۸۴.۹ و درYOLOv4 به ٪ ۸۹.۳۳ رسید[۱۱].

این مقاله [۱۲] یک چارچوب تشخیص‌دهنده اشیا یک مرحله‌ای برای رانندگی خودکار ارائه می‌دهد که بر YOLOv4 استوار است. این چارچوب با استفاده از شبکهٔ پایه (P[7])CSP[5]Darknet53_dcn[6]، دقت تشخیص را بهبود می‌بخشد. لایهٔ آخر خروجی در CSPDarknet53 با کانولوشن تغییرپذیر جایگزین می‌شود تا دقت تشخیص را بهبود دهد. الگوریتم YOLOV4-5D باعث افزایش میانگین دقت تشخیص در مجموعه‌داده‌های BDD[8] به میزان ٪۴.۲۳ و در مجموعه داده‌های KITTI[9] به میزان ٪۱.۶۸ می‌شود. مدل سرعت استنتاج را ٪ ۳۱.۳ افزایش و استفاده از حافظه را به ۹۸.۱ مگابایت کاهش می‌دهد. الگوریتم پیشنهادی قادر به تشخیص با سرعت بیشتر از ۶۶ فریم در ثانیه است و دقت بالاتری نسبت به روش‌های قبلی ارائه می‌دهد[۱۲].

این مقاله[۱۳] یک شبکه عصبی کانولوشن عمیق را برای بهبود تشخیص علائم ترافیکی با استفاده از YOLOv4 ارائه می‌دهد. در این روش، شبکه عصبی عمیق برای تشخیص علامت‌های ترافیکی طراحی شده است که با استفاده از معماری مبتنی بر YOLOv4، این روش قادر است علامت‌های ترافیکی را با دقت و سرعت بالا تشخیص دهد.

[1] Generative Adversarial Networks
[2] Structural Similarity Index
[3] Mean Squared Error
[4] Least Squares Generative Adversarial Networks
[5] Cross-Stage Partial connections
[6] Deformable Convolutional Networks
[7] Parallel processing
[8] Behavior Driven Development
[9] Karlsruhe Institute of Technology and Toyota Technological Institute

شبکه عصبی در این روش با استفاده از لایه‌های عمیق کانولوشنی و ادغام، ویژگی‌های مهم را استخراج می‌کند. سپس با استفاده از لایه‌های کاملاً متصل و لایه‌های تشخیص، علامت‌های ترافیکی را تشخیص می‌دهد. این روش از مزایای معماری YOLOv4 استفاده می‌کند که شامل عملکرد و سرعت بالا و قدرت تشخیص عالی است. با استفاده از این روش، دقت تشخیص علامت‌های ترافیکی به طور قابل‌توجهی افزایش می‌یابد. این شبکه عصبی عمیق، قابلیت یادگیری و بهبود پذیری دارد.

این مقاله[۱۴] یک مدل تشخیص شیء با جزئیات دقیق و سریع را بر اساس شبکه عصبی عمیق YOLOv4 ارائه می‌دهد. این مدل برای تشخیص شیء با جزئیات دقیق طراحی شده است و قادر است با دقت و سرعت بالا اقدام به تشخیص اشیا کند. در این مدل از شبکه عصبی عمیق YOLOv4 استفاده شده است تا با بهره‌گیری از لایه‌های کانولوشنی و ادغام، ویژگی‌های کلیدی و جزئیات دقیق اشیا استخراج شود. سپس لایه‌های خروجی مدل، عملیات تشخیص را با دقت و سرعت بالا انجام می‌دهند. به این ترتیب، مدل قادر است اشیا را با جزئیات کامل و در زمان کوتاه شناسایی کند[۱۴].

این مقاله یک چارچوب تشخیص بصری مبتنی بر یادگیری عمیق را با سرعت و کارایی بالا برای بازرسی اجزای راه‌آهن ارائه می‌دهد. شبکه عصبی کانولوشنی پیشرفته YOLOv4 بر اساس تصاویر موجود در پایگاه داده تصاویر، اجزای راه‌آهن را ارزیابی می‌کند. این شبکه با استفاده از لایه‌های کانولوشنی و لایه‌های دیگر، ویژگی‌های مهم اجزای ریل‌ها را استخراج کرده و با استفاده از لایه‌های مذکور، تشخیص در سریع‌ترین زمان را انجام می‌دهد. با استفاده از این روش، می‌توان به طور دقیق اجزای ریل‌های راه‌آهن مانند استوانه‌ها و اتصالات را تشخیص داد و به‌صورت هم‌زمان بررسی نمود. این روش از دقت و سرعت بالا برخوردار است و قابلیت تشخیص در شرایط واقعی را داراست. از این روش می‌توان در حوزه‌هایی مانند نگهداری و تعمیرات راه‌آهن، امنیت و بررسی اجزای ریل‌ها استفاده کرد[۱۵].

این مقاله[۱۶] یک مدل جدید برای تشخیص هدف در تصاویر پهپادی با نام YOLOv4_Drone ارائه می‌دهد که بر اساس نسخه بهبودیافته الگوریتم YOLO بنا شده است. نتایج آزمایشی نشان می‌دهند که مدل پیشنهادی برای تشخیص هدف در تصاویر پهپادی YOLOv4_Drone نسبت به الگوریتم YOLOv4 بهبود داشته است که نشانگر کارآمدی روش است.

۳- نسخه بهبود یافته YOLOv4 با استفاده از الگوریتم IASD[1]

مدل شبیه‌سازی شده این الگوریتم شامل فرایندهایی است که موجب می‌شود YOLOv4 از نظر دقت و سرعت در تشخیص به نسخهٔ بهبودیافته تبدیل شود. YOLOv4 یک الگوریتم تشخیص اشیا در تصاویر و ویدئوها است که بر اساس شبکه عصبی عمیق ساخته شده است. در مرحله پیش‌پردازش تصاویر در YOLOv4، عملیاتی اعمال می‌شود تا تصاویر به صورتی مناسب برای ورود به شبکه عصبی آماده شوند. این عملیات شامل تغییر اندازه تصاویر، تبدیل آن‌ها به فضای رنگی مشخص، نرمال‌سازی و استانداردسازی داده‌ها و حذف نویز می‌شود.

پس از انجام عملیات پیش‌پردازش، تصاویر آماده به عنوان ورودی به شبکه عصبی YOLOv4 داده می‌شوند. در این مرحله، شبکه عصبی عمیق با استفاده از لایه‌های پیچشی[2] و لایه‌های تماماً متصل[3]، تشخیص اشیا را در تصاویر انجام می‌دهد. پس از خروج نتایج از شبکه، کار تمام می‌شود؛ اما در این مقاله سعی بر این است که از پس‌پردازش برای بهبود و تفسیر نتایج استفاده شود.

[1] Improving Accuracy and Speed in Detection
[2] Convolutional Layers
[3] Fully Connected Layers

خروجی YOLOv4 شامل مختصات و اطلاعات مربوط به اشیا تشخیص داده شده است. این داده‌ها موقعیت مکانی مستطیلی شیء، میزان اطمینان از تشخیص و برچسب کلاس مربوطه را در بر می‌گیرند. با استفاده از پس‌پردازش، می‌توان این خروجی‌ها را تجزیه و تحلیل کرده و به شیوه‌های مختلفی نمایش داده و اطلاعات مورد نیاز را استخراج کرد[۱۷].

همچنین، پس‌پردازش در YOLOv4 می‌تواند شامل تنظیم پارامترهای شبکه باشد. این شامل تنظیماتی مانند آستانه‌ها برای تصمیم‌گیری در مورد وجود یا عدم وجود شیء، تنظیمات مربوط به تشخیص شیء در حالت‌های خاص و سایر تنظیمات مربوط به عملکرد و دقت[1] شبکه است.

نقش پس‌پردازش در نسخه بهبودیافته YOLOv4، بهبود دقت در تشخیص است. YOLOv4 از یک شبکه عصبی کانولوشنی برای پیش‌بینی کادرهای محدودکننده اشیا در یک تصویر استفاده می‌کند. با این حال، پیش‌بینی‌های اولیه شبکه عصبی ممکن است دقیق نباشند. پس‌پردازش با استفاده از تکنیک‌های مختلفی مانند فیلترکردن و حدس‌زدن، دقت تشخیص را بهبود می‌بخشد.

الگوریتم IASD در پنج گام اجرا می‌شود:

گام ۱ : خروجی YOLOv4 شامل چندین جعبه محدودکننده است که ممکن است تکراری، نویزی یا با اطمینان کم باشند.

گام ۲ : ابتدا جعبه‌هایی که امتیاز اطمینان‌شان کمتر از یک آستانه مشخص (θ) است، حذف می‌شوند.

گام ۳ : با استفاده از Non-Maximum Suppression، جعبه‌های همپوشان حذف شده و تنها جعبه‌ای که بالاترین امتیاز اطمینان را دارد باقی می‌ماند.

گام ۴ : جعبه‌های باقی‌مانده که بسیار نزدیک هستند با یکدیگر ادغام می‌شوند، سپس ابعادشان باز تنظیم می‌گردد تا مرز شیء دقیق‌تر شود.

گام ۵ : جعبه‌های نهایی معتبر، در لیست خروجی قرار می‌گیرند.

[1] Accuracy

شبه کد الگوریتم IASD

ورودی‌ها:
تکنیک پس‌پردازش.
predictions: یک ساختار حاوی Bounding Boxes پیش‌بینی‌شده، کلاس‌ها، و امتیازات.
ds: یک مجموعه‌داده حاوی Bounding Boxes و کلاس‌های واقعی.
خروجی:
Bounding Boxes : جعبه‌های فیلترشده
classes: کلاس‌های فیلترشده
scores: امتیازات فیلترشده
۱ فیلتر کردن Bounding Boxes پیش‌بینی‌شده:
برای هر Bounding Box پیش‌بینی‌شده:
امتیاز آن را محاسبه کنید.
اگر
امتیاز Bounding Box برابر یا بیشتر از یک آستانه خاص باشد، آن را نگه دارید.
پایان
پایان
۲ اختصاص کلاس‌ها به Bounding Boxes فیلترشده:
برای هر Bounding Box فیلتر شده:
کلاس با بیشترین احتمال را پیدا کنید.
احتمالات بر اساس امتیازات Bounding Boxes محاسبه می‌شوند.
پایان

حذف کادرهای محدودکننده ضعیف و تکراری: پس‌پردازش از اشتراک بر اجتماع[1] برای حذف کادرهای محدودکننده ضعیف و تکراری استفاده می‌کند. کادرهای محدودکننده‌ای که اشتراک بر اجتماع آن‌ها کمتر از آستانه باشد، حذف می‌شوند. این امر به بهبود دقت تشخیص کمک می‌کند؛ زیرا کادرهای محدودکننده ضعیف و تکراری به احتمال زیاد برای شناسایی شیء اشتباه می‌کنند[۱۸].

بهبود دقت تشخیص اشیا کوچک و اشیا با همپوشانی: پس‌پردازش از الگوریتم‌های حدس‌زدن برای بهبود دقت تشخیص اشیا کوچک و اشیا با همپوشانی استفاده می‌کند. این الگوریتم‌ها از ویژگی‌های کادرهای محدودکننده و امتیازات شناسایی برای حدس‌زدن کلاس صحیح شیء استفاده می‌کنند. این امر به بهبود دقت تشخیص اشیا کوچک و اشیا با همپوشانی کمک می‌کند؛ زیرا این اشیا ممکن است به طور دقیق توسط شبکه عصبی اولیه شناسایی نشوند.

[1] Intersection over Union

کاهش نویز در خروجی الگوریتم: پس‌پردازش می‌تواند به کاهش نویز در خروجی الگوریتم کمک کند. این امر به بهبود دقت تشخیص کمک می‌کند؛ زیرا نویز می‌تواند منجر به شناسایی شیء اشتباه شود. در مجموع، پس‌پردازش نقش مهمی در بهبود دقت تشخیص YOLOv4 ایفا می‌کند.

٤- تنظیم الگوریتم‌های ADAM[1] و SGDM[2]

در مدل YOLOv4، الگوریتم‌های بهینه‌سازی مانند ADAM و SGDM می‌توانند برای آموزش شبکه عصبی استفاده شوند. این الگوریتم‌ها به‌منظور بهبود فرایند یادگیری و کاهش تابع هدف (تابع هزینه) مورد استفاده قرار می‌گیرند. در YOLOv4، این الگوریتم‌ها می‌توانند به‌عنوان بهینه‌سازی استفاده شوند تا شبکه عصبی در فرایند آموزش بهبود یابد. اما برای انتخاب مناسب‌ترین الگوریتم بهینه‌سازی و تنظیم پارامترهای آن، نیاز به ارزیابی و آزمایش بر روی مسئله و داده‌های خاصی داریم. پارامترهای الگوریتم بهینه‌سازی وابسته به مسئله، مجموعه داده و محدودیت‌های مربوطه باید انجام شود.

SGDM نیز یک الگوریتم بهینه‌سازی است که بر پایه گرادیان کاهشی تصادفی[3] با عامل سرعت است. عامل سرعت به منظور حرکت سریع‌تر در فضای جستجو و از بین بردن اثر نویز و عدم قطعیت در گرادیان‌ها استفاده می‌شود. این عامل با افزودن یک مقدار تنظیمی به گرادیان‌ها، موجب کاهش نوسانات و جلوگیری از تغییرات ناخواسته در جهت گرادیان می‌شود. در SGDM نیز می‌توان نرخ یادگیری و عامل سرعت را تنظیم کرد. در این فرمول:

$$V(t)=\alpha\times v(t-1)-\eta\times\nabla J(w(t-1))$$
$$w(t)=W(t-1)+V(t) \qquad (۱)$$

نشان‌دهنده مقدار وزن‌های مورد استفاده در مرحله بعدی است. این مقدار با استفاده از مقدار وزن‌های قبلی و نرخ یادگیری (η) و گرادیان تابع هزینه $(\nabla J(w(t-1)))$ محاسبه می‌شود.

α نشان‌دهنده ضریب تکانه است که معمولاً در بازه [۰, ۱] قرار می‌گیرد. این ضریب مقداری بین ۰ و ۱ است که نشان می‌دهد چه میزان از وزن‌های قبلی باید در محاسبه وزن‌های جدید مؤثر باشند. η نرخ یادگیری است که معمولاً مقدار ثابتی در بازه (۰, ۱) است. این مقدار نشان می‌دهد چقدر باید وزن‌ها در هر مرحله به سمت گرادیان کاهش یابند. $\nabla J(w(t-1))$ گرادیان تابع هزینه نسبت به وزن‌ها در مرحله قبلی است. $w(t)$ وزن‌های جدید در مرحله جاری هستند که با استفاده از مقدار وزن‌های قبلی و مقدار $v(t)$ محاسبه می‌شوند.

SGDM با استفاده از ضریب تکانه می‌تواند بهبود قابل‌توجهی در سرعت و کیفیت یادگیری شبکه‌های عصبی داشته باشد، زیرا از گرادیان‌های گذشته استفاده می‌کند تا به سرعت تغییرات و جهت بهینه را به شبکه برساند.

ADAM یک الگوریتم بهینه‌سازی است که برای تطبیق خودکار نرخ یادگیری با چندین پارامتر، کارآمد است. این الگوریتم با استفاده از گرادیان‌های انتقالی[4] از طولانی‌مدت و میانگین مربعات گرادیان‌ها، نرخ یادگیری را تطبیق می‌دهد. ADAM دارای دو

[1] Adaptive Moment Estimation
[2] Stochastic Gradient Descent with Momentum
[3] Stochastic Gradient Descent
[4] Moving Average

پارامتر اصلی است که قابل‌تنظیم هستند: نرخ یادگیری و بتا[1] که مقادیر بهینه آن‌ها وابسته به مسئله و مجموعه داده مورد استفاده قرار می‌گیرد.

$$m(t) = \beta 1 \times m(t-1) + (1-\beta 1) \times \nabla J(w(t-1))) \quad (۲)$$

$$v(t) = \beta 2 \times v(t-1) + (1-\beta 2) \times (\nabla J(w(t-1)))^2 \quad (۳)$$

$$\hat{m}(t) = \frac{m(t)}{(1-\beta 1^t)} \quad (۴)$$

$$\hat{v}(t) = \frac{v(t)}{(1-\beta 2^t)} \quad (۵)$$

$$w(t) = w(t-1) - \eta \times \frac{\hat{m}(t)}{(\sqrt{\hat{v}_1(t)} + \mathrm{E}} \quad (۶)$$

m(t) نشان‌دهنده یک تخمین اولیه از میانگین گرادیان است. این مقدار با استفاده از مقدار m(t-1) و گرادیان تابع هزینه (∇J(w(t-1))) و ضریب β1 محاسبه می‌شود. v(t) نشان‌دهنده یک تخمین اولیه از میانگین مربعات گرادیان است. این مقدار با استفاده از مقدار v(t-1) و گرادیان تابع هزینه $(\nabla J(w(t-1)))^2$ و ضریب β2 محاسبه می‌شود. مقادیر m(t) و v(t) تخمینی از میانگین گرادیان و میانگین مربعات گرادیان‌اند که در ابتدای فرایند آموزش به منظور کاهش اثر تمایل m(t) و v(t) به سمت صفر به کار گرفته می‌شوند. β1 و β2 ضرایبی هستند که معمولاً در بازه (۰, ۱) قرار می‌گیرند. این ضرایب نشان‌دهنده نسبت اهمیت گرادیان و مربعات گرادیان به تخمین‌های قبلی است.η نرخ یادگیری است که معمولاً مقدار ثابتی در بازه (۰, ۱) است.

۴-۱- پارامترهای تنظیم کننده

نرخ یادگیری[2] نشان‌دهنده میزان تغییراتی است که در هر مرحله به وزن‌ها اعمال می‌شود. مقدار این پارامتر باید به‌گونه‌ای تنظیم شود که بهبودی پیوسته در طول آموزش حاصل شود. در صورت انتخاب نرخ یادگیری بسیار بزرگ، الگوریتم ممکن است به‌جای همگرایی در نقطه بهینه، پیرامون آن نوسان کرده یا از مسیر بهینه منحرف شود.

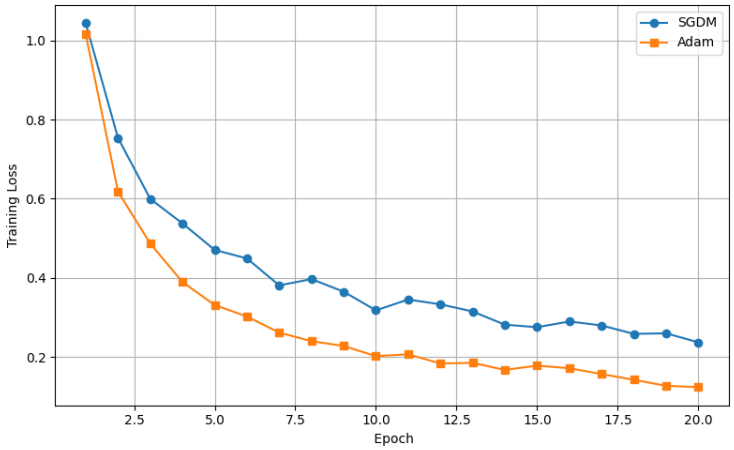
اگر نرخ یادگیری بیش از حد کوچک باشد، ممکن است همگرایی به جواب بهینه بسیار کند شود یا حتی همگرایی به جواب بهینه اصلاً رخ ندهد؛ بنابراین، تنظیم مناسب نرخ یادگیری یکی از مهم‌ترین چالش‌ها در آموزش الگوریتم‌های یادگیری عمیق است.

اندازه دسته[3] نشان‌دهنده تعداد نمونه‌ها است که در هر مرحله برای محاسبه گرادیان استفاده می‌شود. مقدار مناسب برای اندازه دسته باید به‌گونه‌ای انتخاب شود که مصرف منابع محاسباتی از جمله حافظه و توان پردازشی در سطحی بهینه و معقول حفظ شود.

[1] Beta
[2] Learning Rate
[3] Batch Size

اندازه دسته بزرگ‌تر معمولاً به معنای استفاده از بیشترین توان محاسباتی است و می‌تواند سرعت آموزش را افزایش دهد، اما ممکن است بهبود کیفی در نتایج دسته‌بندی را کاهش دهد. از طرف دیگر، اندازه دسته کوچک‌تر می‌تواند دقت را افزایش دهد، اما آموزش را به طور کلی کندتر کند.

حداکثر تعداد دوره‌های آموزشی[1] تعیین می‌کند که فرایند آموزش تا چه مدت ادامه یابد. این پارامتر معمولاً بر اساس تعداد دوره‌هایی تنظیم می‌شود که طی آن، بهبود قابل‌توجهی در عملکرد مدل مشاهده می‌گردد. در مراحل ابتدایی، معمولاً از تعداد دوره‌های کمتر استفاده می‌شود و سپس با ارزیابی روند پیشرفت مدل، این مقدار به‌گونه‌ای تنظیم می‌گردد که بهبود تدریجی و همگرایی به سمت جواب بهینه حاصل شود. با این حال، افزایش بیش از حد تعداد دوره‌های آموزشی می‌تواند منجر به بروز پدیده‌ی بیش‌برازش شود. در این حالت، مدل به جای یادگیری الگوهای کلی داده‌ها، ویژگی‌های مجموعه‌ی آموزشی را به‌صورت خاص حفظ می‌کند و در نتیجه تنها بر روی همان داده‌ها عملکرد خوبی دارد، اما در مجموعه‌های جدید دقت خود را از دست می‌دهد.



شکل (۱) مقایسه روند کاهش خطا بین دو الگوریتم بهینه‌ساز SGDM و ADAM

شکل (۱) روند کاهش خطا[2] را در طول دوره‌های آموزشی برای دو الگوریتم بهینه‌ساز SGDM و ADAM نشان می‌دهد. همان‌طور که دیده می‌شود، SGDM گرچه روندی پایدار دارد، اما سرعت همگرایی آن کمتر از ADAM است. در مقابل، ADAMدر مراحل اولیه با سرعت بیشتری بهینه می‌شود، هرچند ممکن است نوسانات بیشتری داشته باشد. ترکیب این دو الگوریتم در پژوهش حاضر باعث شده است که هم سرعت همگرایی و هم پایداری آموزش بهبود یابد.

[1] Maximum Number of Training Epochs
[2] Training Loss

شبه کد الگوریتم بهینه سازی ترکیبی ADAM و SGDM

ورودی‌ها:

مجموعه‌داده‌های آموزشی: مجموعه داده، شامل نمونه‌هایی است که مدل بر اساس آن‌ها آموزش می‌بیند.

۱ تنظیم نرخ یادگیری:

برای افزایش دقت:

نرخ یادگیری را کاهش می‌دهیم.

اگر

افزایش سرعت تشخیص را می‌خواهیم:

نیازمند بالا بردن نرخ یادگیری است

پایان

پایان

۲ تنظیم اندازه دسته:

برای افزایش دقت:

اندازه دسته را کاهش می‌دهیم.

اگر

افزایش سرعت تشخیص را می‌خواهیم:

نیازمند بالا بردن اندازه دسته است

پایان

پایان

۳ تنظیم حداکثر تعداد دوره‌های آموزشی:

برای افزایش دقت:

حداکثر تعداد دوره‌های آموزشی را افزایش می‌دهیم.

اگر

افزایش سرعت تشخیص را می‌خواهیم:

نیازمند کاهش حداکثر تعداد دوره‌های آموزشی است

پایان

پایان

۴- ۲- تشخیص دهنده و شبکه عصبی مورد استفاده

با توجه به داده‌های جدول (۱) تشخیص‌دهنده با استفاده از شبکه عصبی و جعبه‌های کدگذار، قادر به تشخیص اشیا موجود در تصاویر است. جعبه‌های کدگذار یک مجموعه از مستطیل‌های ثابت هستند که برای تشخیص اشیا به کار می‌روند. هر جعبه کدگذار شامل اطلاعات مکانی و ابعادی است که به شبکه عصبی کمک می‌کند تا اشیا را در تصویر تشخیص دهد. نام‌های کلاس نیز مربوط به

دسته‌بندی مختلف اشیا هستند که تشخیص‌دهنده می‌تواند تشخیص دهد. این نام‌ها معمولاً شامل اشیای معروف مانند ماشین و غیره هستند.

جدول (۱) مشخصات تشخیص دهنده

[1×1 dl network]	شبکه عصبی
[2×1 cell]	جعبه‌های پایه[1]
[80×1 cell]	نام‌های کلاس
[416 416 3]	اندازه ورودی
tiny-yolov4-coco	نام مدل

با ورودی‌هایی به اندازه [۴۱۶ ۴۱۶ ۳]، این آشکارساز قادر است اشیا را در تصاویر تشخیص داده و نتایج را ارائه کند. با استفاده از این تشخیص‌دهنده و ویژگی‌ها، می‌توانید تصاویر را به عنوان ورودی داده و نتایج تشخیص که شامل مکان اشیا و برچسب آن‌ها را دریافت کنید. در این مثال، آشکارساز، تشخیص وسایل نقلیه از تصاویر را انجام می‌دهد.

نتایج دقیق و کیفیت آن‌ها بستگی به عواملی مانند دقت مدل، تنظیمات پیش‌پردازش، کیفیت تصویر و مقدار داده آموزشی دارد. همچنین، مدل‌های مختلف می‌توانند دقت و نتایج متفاوتی داشته باشند. معمولاً مدل‌هایی با دقت بالا و آموزش بر روی مجموعه داده‌های گسترده، نتایج بهتری ارائه می‌دهند.

شبکه عصبی شامل لایه CNN با ابعاد [n net CNN Layer ۱×۷۴] است. با توجه به جدول (۲) این لایه شامل ۷۴ فیلتر با اندازه n×۱ است که برای استخراج ویژگی‌های مختلف از ورودی استفاده می‌شوند.

جدول (۲) مشخصات شبکه عصبی عمیق

[74×1 n net CNN Layer]	لایه‌ها
[80×2 table]	اتصالات
[80×3 table]	پارامترهای قابل یادگیری[2]
[38×3 table]	وضعیت
['input_1']	نام‌های ورودی
['conv_31' 'conv_38']	نام‌های خروجی
1	مقداردهی اولیه

شبکه عصبی دارای اتصالاتی با ابعاد [table ۲×۸۰] است. این اتصالات نشان‌دهنده ارتباطات بین لایه‌های مختلف شبکه و نحوه جریان اطلاعات است.

[1] Anchor Boxes
[2] Learnable

شبکه عصبی دارای پارامترهای قابل یادگیری با ابعاد [table ۳×۸۰] است. این پارامترها معمولاً به صورت وزن‌ها و بایاس‌ها[1] بین لایه‌ها استفاده می‌شوند و در هنگام آموزش شبکه تغییر می‌کنند. در شبکه عصبی، جدولی با ابعاد [۳×۳۸] برای نمایش وضعیت وزن‌ها و پارامترهای داخلی در هر لحظه در نظر گرفته شده است.

ورودی شبکه عصبی با نام 'input_1' و خروجی با نام‌های 'conv_31' و 'conv_38' تعیین شده است. این نام‌ها معمولاً برای ارجاع به لایه‌ها و تصاویر میانی در شبکه استفاده می‌شوند.

۵- نتایج شبیه‌سازی

در این مقاله، به بهبود دقت و سرعت YOLOv4 تمرکز شده است. با استفاده از تکنیک پس‌پردازش بر روی خروجی شبکه و انجام عملیاتی مانند حذف داده‌های تکراری و تصحیح موقعیت اشیا، دقت تشخیص YOLOv4 بهبود یافت.

علاوه بر این، با بهینه‌سازی نرخ یادگیری و تنظیم مناسب آن با استفاده از الگوریتم‌های بهینه‌سازی مانند ADAM یا SGDM، می‌توان عملکرد آموزشی YOLOv4 را بهبود بخشید. همچنین، با تنظیم مناسب اندازه دسته و تعیین حداکثر تعداد دوره‌های آموزشی، می‌توان سرعت آموزش YOLOv4 را افزایش داد و به عملکرد بهتری دست‌یافت.

همان‌گونه که جدول (۳) نشان می‌دهد، نسخه‌های جدید YOLO، در برخی معیارها از YOLOv4 عملکرد بهتری دارند، به‌ویژه در mAP[1] (IoU[2]=0.5:0.95) و سرعت پردازش بر پایه GPU[3]. با این حال، الگوریتم پیشنهادی (YOLOv4+IASD) توانسته است نسبت به YOLOv4 اصلی، هم در mAP (IoU=0.5) و هم در mAP (IoU=0.5:0.95) بهبود قابل توجهی داشته باشد، بدون آنکه نیازمند منابع سخت‌افزاری سنگین باشد. این مزیت برای کاربردهای بلادرنگ در سامانه‌های کم‌منبع مانند پهپادها یا دوربین‌های نظارتی تعبیه‌شده بسیار کلیدی است، چرا که نسخه‌های جدید YOLO عمدتاً برای پردازنده‌های گرافیکی قدرتمند طراحی شده‌اند.

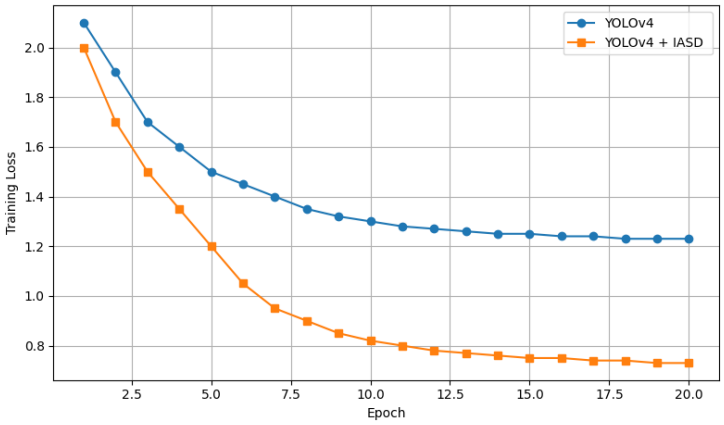
جدول (۳) مقایسه روش پیشنهادی با نسخه‌های جدیدYOLO

نیاز به GPU/CPU	FPS	mAP (IoU=0.5:0.95) (%)	mAP (IoU=0.5) (%)	الگوریتم
متوسط	65	66.0	89.3	YOLOv4
بالا(GPU)	140	69.5	91.2	YOLOv5
بالا(GPU)	160	71.3	92.5	YOLOv7
بالا(GPU)	180	72.4	93.1	YOLOv8
کم (CPU هم قابل اجرا)	70	68.4	90.1	YOLOv4 + IASD

شکل (۲) روند کاهش تابع هزینه را در طول ۲۰ دوره آموزشی برای دو مدل YOLOv4 اصلی و نسخه پیشنهادی (YOLOv4+IASD) نشان می‌دهد. همان‌طور که مشاهده می‌شود، مدل پیشنهادی در مراحل اولیه آموزش با سرعت بیشتری به

[1] Bias

همگرایی می‌رسد و مقدار نهایی تابع هزینه نیز در آن پایین‌تر از YOLOv4 اصلی است. این موضوع بیانگر آن است که ترکیب بهینه‌سازهایADAM و SGDM به‌همراه تابع پس‌پردازش IASD موجب بهبود کارایی فرایند آموزش و در نهایت افزایش دقت مدل در مرحله آزمون شده است.



شکل (۲) روند کاهش تابع هزینه در طول ۲۰ دوره آموزشی برای دو مدل **YOLOv4** اصلی و نسخه پیشنهادی

شکل (۳) تشخیص وسایل نقلیه با استفاده از الگوریتم بهبود یافته **YOLOv4**

شکل (۴) تشخیص وسایل نقلیه با استفاده از الگوریتم اصلی **YOLOv4**

جدول (۴) پارامترهای کلیدی مورد استفاده در آموزش مدل پیشنهادی را نشان می‌دهد. انتخاب این مقادیر بر اساس مطالعات پیشین و آزمایش‌های تجربی صورت گرفته است تا تعادلی میان سرعت همگرایی، پایداری آموزش و دقت نهایی حاصل شود. به‌ویژه، استفاده هم‌زمان از دو بهینه‌ساز ADAM و SGDM با نرخ یادگیری مناسب و ضریب تکانه بهینه، باعث کاهش نوسانات گرادیان و دستیابی به دقت بالاتر را در مدت آموزش کوتاه‌تر فراهم کرده است.

جدول (۴) پارامترهای کلیدی مورد استفاده در آموزش مدل پیشنهادی

توضیح	مقدار انتخاب‌شده	پارامتر
مقدار اولیه برای شروع آموزش	0.001	نرخ یادگیری
تعادل بین سرعت و مصرف حافظه	32	اندازه دسته
تا رسیدن به همگرایی مدل	50	تعداد دوره
ضریب تکانه[۴] در ADAM	0.9	β1 (ADAM)
ضریب کاهش واریانس گرادیان	0.999	β2 (ADAM)
سرعت همگرایی در SGDM	0.9	تکانه (SGDM)
برای افزایش دقت و سرعت هم‌زمان	ADAM + SGDM	بهینه سازی ترکیبی
مطابق استاندارد YOLOv4	416×416	اندازه ورودی تصاویر
سنجش جامع دقت و سرعت	FPS ،mAP (IoU=0.5:0.95) ،mAP (IoU=0.5)	معیارهای ارزیابی

در این پژوهش یک نسخه بهبودیافته از الگوریتم YOLOv4 برای تشخیص وسایل نقلیه معرفی شد. نوآوری اصلی در ترکیب دو بهینه‌سازADAM و SGDM و نیز ارائه یک تابع پس‌پردازش سبک‌وزن با عنوان IASD بود که موجب بهبود هم‌زمان دقت و سرعت تشخیص گردید. نتایج تجربی نشان داد که روش پیشنهادی در مقایسه با YOLOv4 اصلی، هم در معیار mAP (IoU=0.5) و هم در معیار mAP (IoU=0.5:0.95) بهبود قابل توجهی دارد و از نظر سرعت آموزش نیز سریع‌تر به همگرایی می‌رسد. همچنین مقایسه با نسخه‌های جدیدترYOLO نشان داد که هرچند آن‌ها از نظر سخت‌افزاری قدرتمندتر عمل می‌کنند، اما الگوریتم پیشنهادی به دلیل نیاز کمتر به منابع محاسباتی، گزینه‌ای مناسب برای سامانه‌های تعبیه‌شده و کاربردهای بلادرنگ مانند پهپادها، سیستم‌های نظارت شهری و خودروهای هوشمند است. در نهایت می‌توان گفت روش ارائه‌شده پلی میان سادگی و کارایی فراهم کرده و مسیر را برای تحقیقات آینده در جهت بهبود معماری‌های سبک‌وزن و بهینه‌سازی الگوریتم‌های تشخیص شیء هموار می‌سازد.

٦- نتیجه‌گیری

در این پژوهش یک نسخه بهبودیافته از الگوریتم YOLOv4 برای تشخیص وسایل نقلیه ارائه گردید. نوآوری اصلی کار در دو محور متمرکز بود: نخست، طراحی یک الگوریتم پس‌پردازش سبک‌وزن با عنوان IASD که قادر است جعبه‌های تکراری و نویزی

را حذف کرده و دقت نهایی مدل را به‌طور محسوسی افزایش دهد؛ دوم، بهینه‌سازهای ADAM و SGDM به‌صورت ترکیبی استفاده شده‌اند تا با بهره‌گیری از مزایای هر دو، همگرایی مدل سریع‌تر و پایدارتر گردد. نتایج آزمایش‌ها بر روی مجموعه داده گسترده نشان داد که روش پیشنهادی، در مقایسه با YOLOv4 اصلی، هم از نظر میانگین دقت (mAP) و هم از نظر سرعت تشخیص (FPS) بهبود چشمگیری داشته است. همچنین نمودارِ روند کاهش خطا نشان داد که مدل پیشنهادی در مراحل اولیه آموزش با سرعت بیشتری به سمت همگرایی حرکت می‌کند. علاوه بر این، مقایسه با نسخه‌های جدیدتر YOLO بیانگر آن است که هرچند آن‌ها در برخی معیارها عملکرد بهتری دارند، اما روش پیشنهادی به دلیل نیاز سخت‌افزاری کمتر و قابلیت پیاده‌سازی بر روی سامانه‌های کم منبع، مزیتی عملی و قابل توجه برای کاربردهای بلادرنگ نظیر پهپادها، سیستم‌های نظارت شهری و سامانه‌های هوشمند حمل‌ونقل دارد. بدین ترتیب، این پژوهش پلی میان دقت بالا و سادگی محاسباتی فراهم می‌آورد. به‌عنوان چشم‌انداز آینده، توسعه و ارزیابی الگوریتم IASD در کنار سایر نسخه‌های جدید YOLO و همچنین آزمایش بر روی مجموعه‌داده‌های متنوع‌تر می‌تواند به ارتقای بیشتر دقت و پایداری در سامانه‌های هوشمند منجر شود.

مراجع

[1] E. Ong et al. "Video Object Segmentation." Synthesis Lectures on Computer Vision (2024). https://doi.org/10.4018/978-1-59904-845-1.CH106.

[2] Myeongjae Jang et al. "Zero and Narrow-Width Value-Aware Compression for Quantized Convolutional Neural Networks." IEEE Transactions on Computers, 73 (2024): 249-262. https://doi.org/10.1109/TC.2023.3315051.

[3] Miaomiao Liu et al. "An Improved Adam Optimization Algorithm Combining Adaptive Coefficients and Composite Gradients Based on Randomized Block Coordinate Descent." Computational Intelligence and Neuroscience,(2023). https://doi.org/10.1155/2023/4765891.

[4] Gonzales, S. "Enhancing Image Classification Performance: A Comparative Analysis of Optimization Algorithms." International Journal of Advanced Research in Science, Communication and Technology, 2023.

[5] Zheng, Z., Zhang, T., Liu, Z., Liu, B., Li, Y., & Sun, C. "Arbitrarily Oriented Object Detection in Remote Sensing Images Based on Improved YOLOv4-CSP." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 9355-9368, 2022.

[6] Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. "YOLOv4: Optimal Speed and Accuracy of Object Detection." *Submitted*, 23 April 2020.

[7] Yu, J., & Zhang, W. "Face Mask Wearing Detection Algorithm Based on Improved YOLO v4." *Sensors*, vol. 21, no. 9, pp. 3263, 2021. https://doi.org/10.3390/s21093263

[8] Jiang, Z., Zhao, L., Li, S., & Jia, Y. "Real-time Object Detection Method Based on Improved YOLOv4-Tiny." *Submitted*, 9 November 2020 (v1), last revised 2 December 2020 (v2).

[9] H. Sayyadi, M. Mohassel Feghhi, M. Atashbar, M. Nangir, and J. Sayyadi, "Efficient Four-Dimensional Routing Planning of Heterogeneous Drones Utilizing the Fuzzy Clustering Algorithm in Complex Entropy Environments," Tabriz Journal of Electrical Engineering,2024.[Online].Available:https://doi.org/10.22034/tjee.2024.60800.4816.

[10] Zhenwei Yu, Yonggang Shen, Chenkai Shen, A real-time detection approach for bridge cracks based on YOLOv4-FPM, Automation in Construction, Volume 122, 2021.

[11] C. Dewi, R. -C. Chen, Y. -T. Liu, X. Jiang and K. D. Hartomo, "Yolo V4 for Advanced Traffic Sign Recognition With Synthetic Training Data Generated by Various GAN," in IEEE Access, vol. 9, pp. 97228-97242, 2021, doi: 10.1109/ACCESS.2021.3094201.

[12] Y. Caiet al., "YOLOv4-5D: An Effective and Efficient Object Detector for Autonomous Driving," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-13, 2021, Art no. 4503613, doi: 10.1109/TIM.2021.3065438.

[13] Dewi, C., Chen, RC., Jiang, X. et al. Deep convolutional neural network for enhancing traffic sign recognition developed on Yolo V4. Multimed Tools Appl 81, 37821–37845 (2022). https://doi.org/10.1007/s11042-022-12962-5.

[14] Roy, A.M., Bose, R. & Bhaduri, J. A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. Neural Comput & Applic34, 3895–3921 (2022). https://doi.org/10.1007/s00521-021-06651-x.

[15] Feng Guo, Yu Qian, Yuefeng Shi,Real-time railroad track components inspection based on the improved YOLOv4 framework,2021,(https://www.sciencedirect.com/science/article/pii/S0926580521000479).

[16] Li Tan, Xinyue Lv, Xiaofeng Lian, Ge Wang,YOLOv4_Drone: UAV image target detection based on an improved YOLOv4 algorithm,Computers & Electrical Engineering,Volume 93,2021.

[17] Gibson, I., Rosen, D., Stucker, B., Khorasani, M. (2021). Post-Processing. In: Additive Manufacturing Technologies. Springer, Cham. https://doi.org/10.1007/978-3-030-56127-7_16.

[18] H. Sayyadi, M. M. Feghhi, M. Nangir and J. Sayyadi, "Improving UAV-Based Monitoring of Solar Power Plants Using Coverage Path Planning Model with Adaptive Learning Algorithm," *2025 10th International Conference on Technology and Energy Management (ICTEM)*, Tabriz, Iran, Islamic Republic of, 2025, pp. 1-7, doi: 10.1109/ICTEM66196.2025.11063624.

[19] J. Sayyadi, M. Nangir, M. M. Feghhi and H. Sayyadi, "Optimal Management of Solar Energy Generation Using Derivatives of Irradiation Angle Data with the Gradient Boosting Algorithm," *2025 10th International Conference on Technology and Energy Management (ICTEM)*, Tabriz, Iran, Islamic Republic of, 2025, pp. 1-6, doi: 10.1109/ICTEM66196.2025.11063646.