

Identification and Recognition of Objects in the Image Using the Optimized YOLO Algorithm

Amirhosein Zanganeh^{1*} , Ehsan Sharifi² , Pezhman Gholamnezhad³ 

¹ Assistant Professor, Shahid Sattari University of Aeronautical Sciences and Technology, Tehran, Iran
(*Correspondence: zangane.amirhosein@ssau.ac.ir)

² Assistant Professor, Shahid Sattari University of Aeronautical Sciences and Technology, Tehran, Iran

³ Assistant Professor, Shahid Sattari University of Aeronautical Sciences and Technology, Tehran, Iran

ARTICLE INFO

Article history:

Article Type: Research paper

Received: 12 March 2025

Revised: 10 May 2025

Accepted: 08 June 2025

Available online: 11 July 2025

Keywords:

Video Surveillance System

Object Recognition

YOLO Algorithm




Deep Neural Network

Activation Function

ABSTRACT

Today, video surveillance systems are widely used to monitor and control the environment. To prevent unwanted incidents and protect military and security sites, people, and their property, investment in video surveillance systems has increased greatly, and the goal is to make the most of all the technological achievements in this field for the development of video surveillance systems. The use of video surveillance systems in organizations, offices, factories, and work environments has resulted in careful monitoring and control of the environment, reducing violations, and increasing the ability to quickly detect incidents and order the work environment. The purpose of identifying and recognizing objects in video surveillance systems is to categorize and label objects and determine their exact position in the image or video.

Today, deep neural networks are used to solve various problems. In this article, we have used the optimized YOLO algorithm to identify and recognize objects in images. In the basic architecture of the YOLO network, the Relu activation function is used, which is not derivable at the zero point, and secondly, it leads to the zeroing of all negative values. As a result, we are trying to increase the accuracy of the basic neural network by changing the activation function. By making a change in the basic architecture of the Yolo algorithm in identifying and recognizing objects, we have increased the accuracy of basic neural networks in identifying and recognizing objects in images by mAP = 0.8%.

Cite this article: Zanganeh A. , Sharifi E. , Gholamnezhad P.  (2025). Identification and Recognition of Objects in the Image Using the Optimized YOLO Algorithm. Journal of Electronic and Cyber Defens. 2025; 13(2):1-8.


DOR: <https://dor.isc.ac/dor/20.1001.1.23224347.1404.13.2.1.8>

© Author(s) retain the copyright and full publishing rights

Publisher: Imam Hossein University.



شناسایی و تشخیص اشیا در تصویر با استفاده از الگوریتم یولو بهینه شده

امیرحسین زنگنه^{۱*}، احسان شریفی^۲، پژمان غلام نژاد^۳ 

^۱ استادیار، دانشگاه علوم و فنون هوایی شهید ستاری، تهران، ایران (نویسنده مسئول: zangane.amirhosein@ssau.ac.ir)

^۲ استادیار، دانشگاه علوم و فنون هوایی شهید ستاری، تهران، ایران (sharifi@ssau.ac.ir)

^۳ استادیار، دانشگاه علوم و فنون هوایی شهید ستاری، تهران، ایران (pezhman.gholamnezhad@gmail.com)

مشخصات مقاله

چکیده

تاریخچه مقاله:

نوع مقاله: علمی پژوهشی

دریافت: ۱۴۰۳/۱۲/۲۲

بازنگری: ۱۴۰۴/۲/۲۰

پذیرش: ۱۴۰۴/۳/۱۸

ارائه آنلاین: ۱۴۰۴/۴/۲۰

کلید واژه‌ها:

سیستم نظارت تصویری

تشخیص اشیا

الگوریتم یولو

شبکه عصبی عمیق

تابع فعال سازی

امروزه به صورت گسترده برای نظارت و کنترل محیط از سامانه‌های نظارت تصویری استفاده می‌شود. برای جلوگیری از حوادث ناخواسته و حفاظت از اماکن نظامی، امنیتی، مردم و اموال آن‌ها، سرمایه‌گذاری در سامانه‌های نظارت تصویری بسیار افزایش یافته و هدف، استفاده حداکثری از تمام دستاوردهای تکنولوژیکی موجود در این زمینه برای توسعه سامانه‌های نظارت تصویری است. استفاده از سامانه‌های نظارت تصویری در سازمان‌ها، ادارات، کارخانه‌ها و محیط‌های کاری موجب مراقبت و کنترل دقیق محیط، کاهش تخلفات، افزایش توانایی در آشکارسازی سریع حوادث و نظم دهی محیط کاری شده است. هدف از شناسایی و تشخیص اشیا در سامانه‌های نظارت تصویری، دسته‌بندی و برچسب‌گذاری اشیا و تعیین موقعیت دقیق آن‌ها در تصویر یا ویدئو است.

امروزه شبکه‌های عصبی عمیق برای حل مسائل مختلف مورد بررسی و استفاده قرار می‌گیرند. در این مقاله ما به منظور شناسایی و تشخیص اشیا در تصاویر از الگوریتم بهینه شده یولو استفاده نموده‌ایم. در معماری پایه شبکه یولو از تابع فعال‌سازی ریلو استفاده شده است که این تابع فعال‌سازی در نقطه صفر مشتق پذیر نیست و دوم منجر به صفر شدن تمام مقادیر منفی می‌گردد در نتیجه ما با تغییر تابع فعال‌سازی سعی در افزایش دقت شبکه عصبی پایه داریم. ما با ایجاد تغییر در معماری پایه الگوریتم یولو در شناسایی و تشخیص اشیا، دقت شبکه‌های عصبی پایه را در شناسایی و تشخیص اشیا در تصاویر را به میزان $mAP = 0.8$ درصد افزایش داده‌ایم.

استناد: زنگنه، امیرحسین^۱، شریفی، احسان^۲، غلام نژاد، پژمان^۳. شناسایی و تشخیص اشیا در تصویر با استفاده از الگوریتم یولو بهینه شده. پدافند الکترونیک و سایبری. (۱۴۰۴)؛ ۱۳ (۲): ۸-۱.

DOR: <https://dor.isc.ac/dor/20.1001.1.23224347.1404.13.2.1.8>

© نویسنده(گان) حق نشر و حقوق کامل انتشار را برای خود محفوظ می‌دارند.



ناشر: دانشگاه جامع امام حسین (ع).

OPEN ACCESS

۱- مقدمه

شدن تمام مقادیر منفی نمی‌گردد. همچنین ما در پروسه یادگیری آن از فن ادغام تصاویر بهره گرفته‌ایم. با استفاده از این فن فرایند آموزش با تصاویر پیچیده‌تر انجام شده و نهایتاً بردار ویژگی استخراج شده جامع‌تر خواهد بود. ادامه مقاله به شرح زیر سازمان‌دهی شده است: در بخش ۲، روش تحقیق شامل: کارهای انجام شده در زمینه شناسایی و تشخیص اشیا در تصویر و روش پیشنهادی شرح داده می‌شود، در بخش ۳ نتایج تجربی ارائه شده است و در نهایت، نتیجه‌گیری در بخش ۴ بیان شده است.

۲- روش تحقیق

شبکه عصبی کانولوشنال^۱ یکی از روش‌های موفق و پیشرو در دسته‌بندی و طبقه‌بندی تصاویر است. در مدل‌های یادگیری عمیق، هر تصویر ورودی از یک سری لایه‌های کانولوشن با فیلترها، لایه پولینگ و لایه‌های کاملاً متصل عبور می‌کند، و در نهایتاً از تابع سافت‌مکس^۲ برای دسته‌بندی و تعیین کلاس اشیا با مقادیر احتمالی بین ۰ و ۱ استفاده می‌شود. الگوریتم‌های یادگیری عمیق به‌طور گسترده‌ای برای طبقه‌بندی اشیا مورد استفاده قرار می‌گیرد. توانایی پردازش سریع خوشه‌های بزرگ تصاویر، مدل یادگیری عمیق را به‌عنوان مهم‌ترین روش برای طبقه‌بندی اشیا در تصاویر معرفی می‌کند [۹]. امروزه، عملکرد الگوریتم‌های شناسایی و تشخیص اشیا عمدتاً در استخراج ویژگی‌ها [۱۰]، با استفاده از شبکه‌های عصبی عمیق بهبود یافته‌اند و می‌توانند با دقت خوبی فرآیند شناسایی و تشخیص اشیا را انجام دهند [۱۱].

گیرشیک^۳ و همکاران [۱۲] از شبکه عصبی کانولوشن مبتنی بر ناحیه برای شناسایی و تشخیص اشیا استفاده کرده‌اند. این شبکه عصبی، روشی برای حل مشکل انتخاب تعداد زیادی از مناطق پیشنهاد شده است. این روش از جستجوی انتخابی برای استخراج تنها ۲۰۰۰ منطقه از تصویر استفاده می‌کند و آن‌ها را ناحیه‌های پیشنهادی می‌نامد. این روش را نمی‌توان به‌صورت در لحظه^۴ پیاده‌سازی کرد. آن‌ها با استفاده از شبکه‌ی عصبی کانولوشنال مبتنی بر ناحیه به $mAP = 62.4$ درصد در شناسایی و تشخیص اشیا دست یافته‌اند.

رن^۵ و همکاران [۱۳] یک شبکه عصبی کانولوشن مبتنی بر ناحیه سریع را برای شناسایی و تشخیص اشیا معرفی کرده‌اند. شبکه عصبی کانولوشن مبتنی بر منطقه سریع مدلی برای تشخیص اشیا است که برای عملکرد بالا در زمان واقعی طراحی شده است. شبکه کل تصویر را پردازش می‌کند تا یک نقشه ویژگی تولید کند. سپس برای هر شیء پیشنهادی، یک بردار ویژگی با طول

برای جلوگیری از حوادث ناخواسته و حفاظت از اماکن نظامی و امنیتی، مردم و اموال آن‌ها، سرمایه‌گذاری در سامانه‌های نظارت تصویری بسیار قابل توجه است و هدف، استفاده حداکثری از تمام دستاوردهای تکنولوژیکی موجود در این زمینه برای توسعه سامانه‌های نظارت تصویری یا سامانه‌های پایش تصویری است. این دستگاه‌ها برای ایجاد امنیت و پیشگیری از وقوع جرم هستند و از یک یا چند دوربین ویدئویی در شبکه تشکیل شده که ویدئو و اطلاعات صوتی را به مکانی خاص ارسال می‌کنند. دوربین‌های با ویژگی‌های مختلف برحسب نوع نیاز، در حال حاضر در سامانه‌های نظارت تصویری که در اماکن مختلف نصب می‌شوند، مورد استفاده قرار می‌گیرند. استفاده از سامانه‌های نظارت تصویری در سازمان‌ها، ادارات، کارخانه‌ها و محیط‌های کاری موجب مراقبت و کنترل دقیق محیط، کاهش تخلفات، افزایش توانایی در آشکارسازی سریع حوادث و نظم دهی محیط کاری شده است. هدف نهایی استفاده از سامانه‌های نظارت تصویری مکانیزه کردن فرایند پایش محیطی هست و هدف از شناسایی و تشخیص اشیا در سامانه‌های نظارت تصویری، دسته‌بندی و برچسب‌گذاری اشیا و تعیین موقعیت دقیق آن‌ها در تصویر یا ویدئو است.

امروزه شبکه‌های عصبی عمیق برای حل مسائل مختلف مورد بررسی و استفاده قرار می‌گیرند [۱۱]، [۱۲]، [۱۳]. با توسعه سریع فناوری بینایی کامپیوتری، تحقیقات در مورد الگوریتم‌های تشخیص اشیا توجه گسترده‌ای را برانگیخته و به کانون تحقیقاتی تبدیل شده‌اند [۱۴].

روش سنتی تشخیص شیء در سامانه‌های پایش تصویری را می‌توان به سه مرحله تقسیم کرد: ۱- انتخاب ناحیه: ناحیه مورد نظر را می‌توان با استفاده از فن پنجره کشویی با مقیاس‌های مختلف انتخاب کرد. ۲- استخراج ویژگی: ویژگی‌ها را می‌توان از ناحیه‌های برچسب‌گذاری شده استخراج کرد. در روش‌های استخراج ویژگی سنتی، تبدیل ویژگی تغییر مقیاس [۵]، هیستوگرام گرادیان جهت‌دار [۶]، یادگیرنده‌های مطلوب و توابع چگالی احتمال [۷] و هار [۸]، معمولاً برای استخراج ویژگی‌ها استفاده می‌شود. روش‌های سنتی با تغییر شکل و تغییر شرایط نور نمی‌توانند همه ویژگی‌ها را استخراج کنند و در نتیجه قابل اتکا و مطمئن نیستند. ۳- طبقه‌بندی و رگرسیون: در این مرحله بر اساس ویژگی‌های استخراج شده، فرآیند طبقه‌بندی انجام می‌شود و باکس پیش‌بینی شده برای موقعیت هر کلاس تعیین می‌شود.

ما در این مقاله با ایجاد تغییر در معماری پایه شبکه یولو از تابع فعال‌سازی لیکیریلو استفاده کرده‌ایم. این تابع فعال‌سازی دو مزیت مهم نسبت به تابع ریلو دارا است، نخست اینکه برخلاف تابع ریلو در نقطه صفر مشتق‌پذیر است و دوم اینکه منجر به صفر

¹ Convolutional Neural Network

² Soft-max

³ R. Girshick

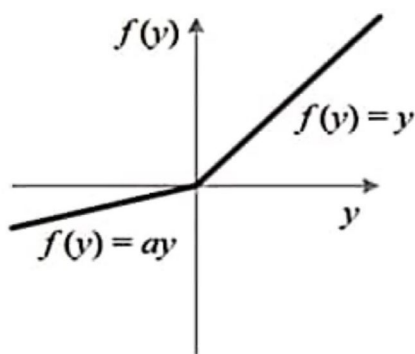
⁴ REAL-time

⁵ S. Ren

ریلو^۹ است. این تابع فعال‌سازی دو مزیت مهم نسبت به تابع ریلو دارا است. نخست اینکه برخلاف تابع ریلو در نقطه صفر مشتق پذیر است و دوم اینکه همانند تابع ریلو منجر به صفر شدن تمام مقادیر منفی (که باعث ازدست رفتن برخی از نورون‌ها در ساختار شبکه عصبی می‌شود) نمی‌گردد.

Type	Filters Size	Output
Convolutional	32 3 × 3	256 × 256
Convolutional	64 3 × 3 / 2	128 × 128
Convolutional	32 1 × 1	128 × 128
Convolutional	64 3 × 3	
Residual		128 × 128
Convolutional	128 3 × 3 / 2	64 × 64
Convolutional	64 1 × 1	64 × 64
Convolutional	128 3 × 3	
Residual		64 × 64
Convolutional	256 3 × 3 / 2	32 × 32
Convolutional	128 1 × 1	32 × 32
Convolutional	256 3 × 3	
Residual		32 × 32
Convolutional	512 3 × 3 / 2	16 × 16
Convolutional	256 1 × 1	16 × 16
Convolutional	512 3 × 3	
Residual		16 × 16
Convolutional	1024 3 × 3 / 2	8 × 8
Convolutional	512 1 × 1	8 × 8
Convolutional	1024 3 × 3	
Residual		8 × 8
Avgpool	Global	
Connected	1000	
Softmax		

شکل ۱: ساختار شبکه دارک نت ۵۳ [۱۸]



شکل ۲: تابع فعال‌سازی leaky relu

همچنین ما در پروسه یادگیری آن از تکنیک ادغام تصاویر بهره گرفته‌ایم. (به صورت رندوم به برخی از تصاویر، تصویر رندومی اضافه می‌شود.) همان‌طور که در شکل شماره ۳ نشان داده شده است ما با استفاده از این تکنیک با ایجاد تغییر در تصاویر اولیه، افزودن برخی اشیا به تصویر اولیه و غیره تصاویر ورودی به شبکه عصبی را پیچیده‌تر کرده‌ایم تا بتوانیم فرایند آموزش شبکه را به صورت کامل انجام دهیم.

ثابت را از نقشه ویژگی استخراج می‌کند. هر بردار ویژگی وارد دنباله‌هایی از لایه‌های کاملاً متصل می‌شود. این الگوریتم از قدرت پردازش چی‌پی‌یو^۱ و سی‌پی‌یو^۲ استفاده می‌کند.

فو^۳ و همکاران [۱۴] پیشنهاد کردند که از یک شبکه رزنت ۱۰۱ لایه‌ی عمیق‌تر برای استخراج ویژگی‌ها و یک لایه‌ی دکانولوشنال^۴ برای جایگزینی لایه کانولوشن استفاده شود تا دقت مدل را به بهای اضافه کردن پارامترها و محاسبات بیش‌ازحد افزایش دهد.

شن^۵ و همکاران [۱۵] روشی را برای بهبود دقت شناسایی و تشخیص اشیا با استفاده از شبکه عصبی عمیق دنس‌نت [۱۶] پیشنهاد کردند. آن‌ها یک آشکارساز عمیق شی با نظارت^۶ را که می‌تواند با داده‌های آموزشی محدود آشکارسازی شی را از ابتدا یاد بگیرد، ارائه کرده‌اند.

۱-۲- روش پیشنهادی

الگوریتم YOLO توسط ردمن^۷ و همکاران [۱۷] ارائه شده است. این الگوریتم توانست سرعت شناسایی و تشخیص اشیا را به طرز چشمگیری افزایش دهد. الگوریتم یولو^۳ [۱۸]، برخلاف الگوریتم یولو^۱ و الگوریتم یولو^۲ [۱۹]، با حفظ مزیت دو الگوریتم یولو^۱ و ^۲، سرعت بالایی و شناسایی و تشخیص اشیا، دقت شناسایی و تشخیص اشیا را تا حد زیادی بهبود بخشید. ساختار شبکه چند مقیاسی معرفی شده در الگوریتم یولو موجب عملکرد عالی این الگوریتم در شناسایی و تشخیص اشیا شده است. بدون شک ساختار شبکه در یولو عمیق است و در عین حال به نتایج بهتری هم‌دست می‌یابد. در الگوریتم ورژن ۳ مدل یولو از ساختار دارک-نت^۸ ۵۳ (شکل ۱) به منظور استخراج ویژگی‌ها استفاده شده است. این شبکه به‌طور کلی از کنار هم قرار گرفتن تعدادی بلوک با دو ساختار رژیوال و کانولوشن ساخته شده است.

در این مقاله و در قسمت مصورسازی تصویر خروجی، هر مرحله از نمودار شکل ۱ به‌عنوان یک لایه نام‌گذاری شده است. ابعاد تصاویر خروجی از لایه‌ها به‌صورت زیر است. مؤلفه اول هر یک از تنسورهای زیر مشخص‌کننده بچ‌سایز است، مؤلفه دوم بیانگر تعداد کانال و دو مؤلفه انتهایی طول و عرض تصاویر را مشخص می‌کنند. قابل توجه است که ابعاد مشخص شده برای لایه‌های مختلف برحسب تصویر ورودی ۴۱۶*۴۱۶ مشخص شده‌اند.

ما با ایجاد تغییر در معماری پایه شبکه یولو از تابع فعال‌سازی لیکی‌ریلو استفاده کرده‌ایم، به عبارتی تابع فعال‌سازی به‌کارگرفته شده در تمام کانولوشن‌های ساختار دارک‌نت لیکی-

^۱ GPU

^۲ CPU

^۳ C.-Y. Fu

^۴ Deconvolutional

^۵ Z. Shen

^۶ Deeply Supervised Object Detector

^۷ Redmon

^۸ darknet 52

^۹ Leaky Relu

معیارهای بازیابی^۲ (رابطه ۱) و وضوح^۳ (رابطه ۲) است، و معیار - اف^۴ (رابطه ۳) استفاده کرده‌ایم. معیار mAP یا میانگین متوسط دقت، یکی از مهم‌ترین معیارهایی است که برای ارزیابی دقت مدل‌های شناسایی و تشخیص اشیا استفاده می‌شود. در این ارزیابی‌ها هدف شناسایی و تشخیص انسان در تصاویر حرارتی است.

$$Recall = \frac{TP}{TP+FP} \quad (1)$$

$$Precision = \frac{TP}{TP+FN} \quad (2)$$

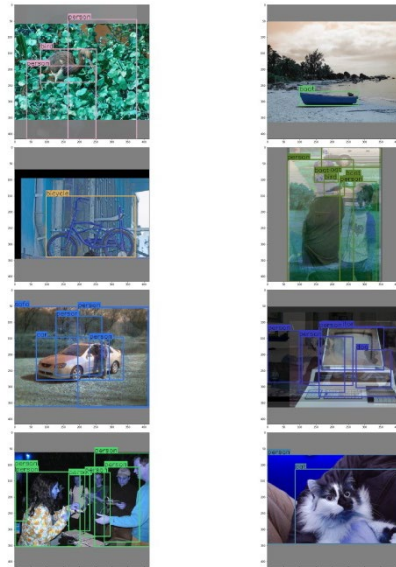
$$f - measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (3)$$

که در آن TP^۵ تعداد نمونه‌های مثبتی است که به درستی مثبت شناسایی شده‌اند (نرخ مثبت واقعی)، TN^۶ تعداد نمونه‌های منفی که به درستی منفی شناسایی شده‌اند، FP^۷ تعداد شناسایی‌های مثبت کاذب و FN^۸ تعداد شناسایی‌های منفی کاذب هستند.

در طول پروسه یادگیری پس از طی شدن هر ۴ اپیاک بر روی دیتای آموزشی، عملکرد مدل با استفاده از دیتای آزمودن مورد ارزیابی قرار می‌گیرد و در نهایت مدلی که به بیشینه مقدار mAP برسد به عنوان مدل نهایی ذخیره می‌شود. پروسه یادگیری در ۸۰ اپیاک صورت پذیرفته و در پروسه یادگیری آن از تکنیک ادغام تصاویر که در بخش ۲-۱ به آن اشاره کردیم، بهره گرفته شده است. پیاده‌سازی مدل با استفاده از چارچوب پای‌تورچ^۹ صورت پذیرفته است. در تعریف ساختار الگوریتم یولو ۳، در قسمت backbone از مدل دارک نت ۵۲ آموزش داده شده بر روی imagenet استفاده شده است تا پروسه یادگیری سریع‌تر طی شود. همچنین در پروسه یادگیری از روش گرادین کاهشی تصادفی^{۱۰} به منظور بهینه‌سازی ضرایب ساختار مدل استفاده شده است. میزان مومنتوم مربوط به این روش برابر با ۰٫۹، در نظر گرفته شده است و نرخ یادگیری^{۱۱} آن با استفاده از روش cosine_lr_scheduler تنظیم می‌شود. در به کارگیری روش گرادین کاهشی تصادفی از weight decay برابر با ۰٫۰۰۰۵ نیز استفاده شده است.

۳-۳- ارزیابی و مقایسه روش ارائه شده

در این بخش، نتایج ارزیابی روش پیشنهادی در شناسایی و تشخیص اشیا در تصویر با استفاده از روش پیشنهادی ارائه شده است. به منظور بررسی صحت عملکرد مدل می‌بایست دو پارامتر مهم NMS_THRESH و CONF_THRESH در شروع کار تعیین شوند. پارامتر نخست بیان می‌کند که چه مقدار iou^{۱۲} بین تصویر



شکل ۳: تکنیک ادغام تصویر

۳- نتایج و بحث

در این بخش به ارزیابی و تحلیل روش پیشنهادی می‌پردازیم. در ابتدا مشخصات پایگاه داده تصاویر مورد استفاده را معرفی کرده و سپس روش پیشنهادی را مورد ارزیابی قرار می‌دهیم.

۳-۱- پایگاه داده تصاویر

در این مقاله از مجموعه دیتاست پاسکال^۱ ۲۰۱۲ و ۲۰۰۷ استفاده شده است (جدول ۱). این دیتاست شامل تصاویر مربوط به اجسام مختلف در ۲۰ دسته است. تعداد تصاویر مورد استفاده آموزشی در این دیتاست برابر تصاویر موجود در trainval موجود در این دو دیتاست، که برابر با ۱۶۵۵۱ هست. همچنین در مجموع از ۴۹۵۲ تصویر برای آزمودن مدل استفاده شده است. (این تصاویر بر اساس دیتای آزمودن پاسکال ۲۰۰۷ انتخاب شده‌اند).

جدول ۱: پارامترهای پایگاه داده

پاسکال (۲۰۱۲+۲۰۰۷)	نام پایگاه داده
۲۰	تعداد کلاس‌ها
۱۶۵۵۱	تعداد تصاویر آموزشی
۴۹۵۲	تعداد تصاویر آزمودن
۵۲۰۹۰	تعداد باکس‌های نمونه
۲٫۴	تعداد کل تصاویر/ تعداد باکس‌های نمونه

۳-۲- معیارهای ارزیابی

ما به منظور ارزیابی عملکرد روش پیشنهادی از معیار اصلی سنجش عملکرد الگوریتم تشخیص اجسام mAP که برحسب

² Recall

³ Precision

⁴ F-measure

⁵ True Positive

⁶ True Negative

⁷ False Positive

⁸ False Negative

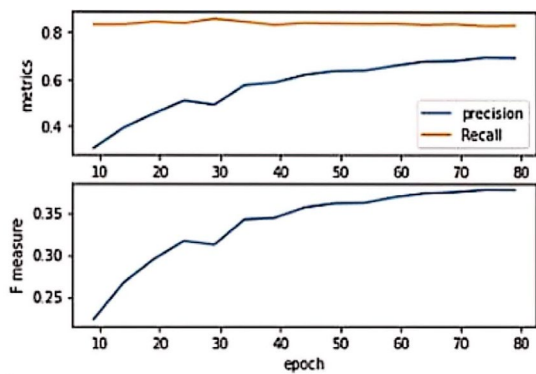
⁹ Pytorch Framework

¹⁰ Stochastic Gradient Descent

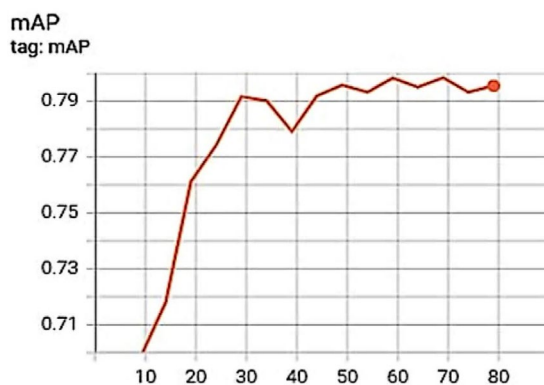
¹¹ Learning Rate

¹² Intersection over Union

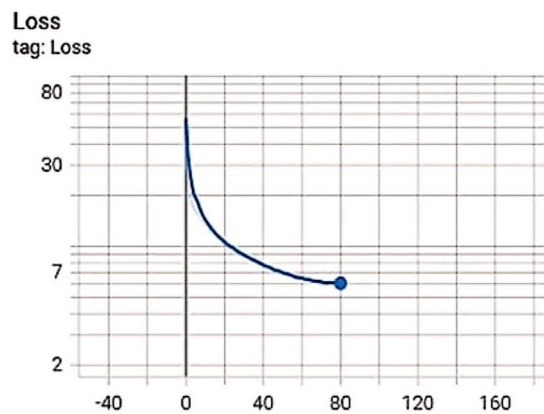
¹ <http://host.robots.ox.ac.uk/pascal/VOC/voc/2012>



شکل ۵: ارزیابی نتایج حاصل از اجرای مدل. معیار وضوح، ریکال و f-measure برحسب اپیاک



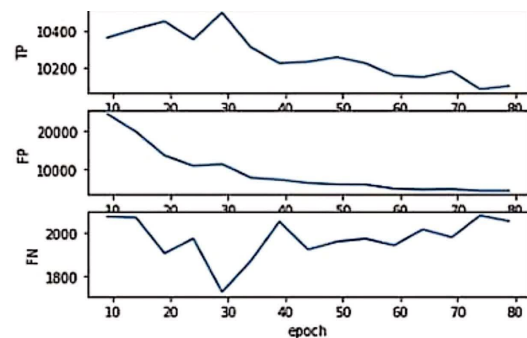
شکل ۶: مقدار میانگین متوسط دقت برحسب اپیاک



شکل ۷: مقدار تابع هزینه بر حسب اپیاک

در جدول شماره ۲ میزان دقت شناسایی و تشخیص گروه‌های مختلف موجود در دیتاست پاسکال ارائه شده است. توجه به این اعداد می‌تواند در تفسیر مدل و اینکه در چه مواردی می‌توان مدل را بهبود داد، موثر باشد. همان‌طور که از نتایج مشخص است مدل در تشخیص دوچرخه و موتورسیکلت با وجود شباهت نسبی این دو به یکدیگر عملکرد بسیار مناسبی را از خود نشان داده است. این در حالی است که در بعضی از گروه‌ها همانند گیاهان گلدانی مدل در عمل نتوانسته به میانگین متوسط دقت قابل توجهی دست یابد. شاید علت این موضوع تشابه تصویر

پیش‌بینی شده و تصویر برچسب زده شده به‌عنوان درصد قابل قبول در نظر گرفته شود. در اجرای نهایی این مقدار برابر با ۰٫۴۵ در نظر گرفته شده است. پارامتر دوم مشخص می‌کند که درصد مورد اعتماد مدل برای تأیید پیش‌بینی یک لیبل چه میزان است که این پارامتر برابر با ۰٫۲ در نظر گرفته شده است. انتخاب عدد کوچک‌تر برای این معیار منجر به رسیدن به میانگین متوسط دقت بالاتری می‌شود. زیرا در این وضعیت نرخ مثبت واقعی بیشتری توسط مدل در تصاویر دیده خواهد شد. با این حال از طرف دیگر کم کردن این عدد منجر به افزایش تعداد نمونه‌های مثبت کاذب خروجی نیز می‌شود که در نتیجه آن معیار وضوح به‌شدت کاهش خواهد یافت.



شکل ۸: مقادیر مثبت حقیقی، مثبت کاذب و منفی کاذب نسبت به اپیاک ترین شده (سنجیده شده با استفاده از دیتای آزمودن)

نتایج به‌دست‌آمده در شکل ۴ نشان می‌دهد که عملکرد مدل به‌مرور و برحسب اپیاک‌های طی شده در تشخیص مقادیر مثبت کاذب عملکرد بسیار بهتری را از خود نشان می‌دهد. هرچند از اپیاک ۳۰ به بعد عملکرد مدل در تشخیص مقادیر مثبت حقیقی روند منفی را شروع می‌کند، با این حال تغییر این عدد به نسبت مقادیر مربوط به منفی کاذب قابل توجه نیست و همان‌طور که در نمودارهای شکل ۵ مشاهده می‌شود این موضوع منجر به افزایش قابل توجه مقدار Recall می‌شود. بر اساس شکل شماره ۶، مقادیر به‌دست‌آمده برای میانگین متوسط دقت که متریک اصلی مربوط به ارزیابی مدل است نیز مؤید بهبود قابل توجه عملکرد مدل در ۵۰ اپیاک ابتدایی است. بعد از این بازه روند بهبود این مقدار این شاخص کند شده و تقریباً در اپیاک ۷۰ رویه مثبت آن معکوس می‌شود. در مجموع مدلی که بهترین عملکرد را از نظر شاخص میانگین متوسط دقت دارد به‌عنوان بهترین مدل ذخیره می‌شود. به‌منظور اطمینان از صحت پروسه طی شده در یادگیری الگوریتم توجه به نمودار تغییرات تابع هزینه می‌تواند اطلاعات مفیدی را در اختیار ما قرار دهد. در شکل شماره ۷ نمودار تابع هزینه الگوریتم یولو۳ برحسب اپیاک ترسیم شده است. همان‌طوری که این نمودار ترسیم شده نشان می‌دهد روش بهینه‌سازی به کار گرفته شده در حل مسئله به‌خوبی نتوانسته است تابع هزینه را در پروسه یادگیری مدل به سمت رسیدن به نقطه کمینه هدایت کند.

۴- نتیجه‌گیری

امروزه برای نظارت، بررسی و کنترل سازمان‌ها، ادارات، کارخانه‌ها و محیط‌های کاری و غیره کاری از سامانه‌های نظارت یا پایش تصویری استفاده می‌شود. استفاده از این سیستم‌ها موجب مراقبت و کنترل دقیق محیط، کاهش تخلفات، افزایش توانایی در آشکارسازی سریع حوادث و نظم دهی محیط‌های کاری شده است. هدف نهایی این سیستم‌ها شناسایی و تشخیص اشیا در سامانه‌های نظارت تصویری، دسته‌بندی و برجسب‌گذاری اشیا و تعیین موقعیت دقیق آن‌ها به صورت خودکار و هوشمند در تصویر یا ویدئو است.

در این مقاله یک روش شناسایی و تشخیص اشیا با استفاده از الگوریتم یولو ۳ بهینه‌شده ارائه شده است. نتایج به دست آمده نشان می‌دهد که این الگوریتم کارایی قابل‌قبولی داشته و در مقایسه با سایر کارهای انجام‌شده در این حوزه، از دقت بالاتری برخوردار است و ما در این مقاله دقت شناسایی و تشخیص اشیا در تصویر را به میزان $mAP = 0.8$ درصد افزایش داده‌ایم.

۱۰- مراجع

- [1] A. Abdullah, G. Mohammad Amin, and Sh. F. Mohammad Hassan, "Diagnosis of road slippage using road camera images based on convolutional neural networks and transfer learning," 2022, Accessed: Mar. 12, 2024. <https://www.sid.ir/paper/1082685/fa> (In Persian)
- [2] M. Asadi, S. Parsa, M. Jamali, and V. Majidinjad, "Peer-to-Peer Bot Network Detection Using Deep Learning Method," *Electronic and Cyber Defense*, vol. 8, no. 2, pp. 1–14, 2020. (In Persian). <https://dor.isc.ac/dor/20.1001.1.23224347.1399.8.2.1.3>
- [3] A. Niloufer, M. M. Seyyed Mohammad Reza, and T. A. Sadat, "Reducing the effects of deception attacks in GPS receivers of phasor measurement units using neural networks," 2023, Accessed: Mar. 12, 2024. [Online]. Available: <https://www.sid.ir/paper/1083027/fa>. (In Persian)
- [4] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," *Advances in neural information processing systems*, vol. 26, 2013, Accessed: Nov. 13, 2023. [Online]. Available: <https://proceedings.neurips.cc/paper/5207-deep-neural-networks-for-object-detection>
- [5] E. N. Mortensen, H. Deng, and L. Shapiro, "A SIFT descriptor with global context," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, IEEE, 2005, pp. 184–190. Accessed: Nov. 13, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/1467266>
- [6] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *2009 IEEE 12th international conference on computer vision*, IEEE, 2009, pp. 32–39. Accessed: Nov. 13, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/5459207>

مربوط به گیاهان با پس‌زمینه تصاویری از محیط باشد که در پس‌زمینه آن‌ها گیاهان و رنگ سبز حضور دارند. یکی از راهکارها برای بهبود مدل ممکن است تقویت دیتاهای موجود در دسته‌هایی باشد که شاهد عملکرد ضعیف‌تری هستیم. این موضوع می‌تواند با استفاده از تأکید بیشتر بر تصاویر این گروه‌ها در مرحله ترکیب تصاویر و یا افزودن دیتای خارجی صورت پذیرد، البته با قطعیت نمی‌توان در مورد اینکه انجام این امور در نهایت منجر به بهتر شدن مدل نهایی شود، نظر داد.

جدول ۲: میزان میانگین متوسط دقت روش پیشنهادی بر حسب کلاس‌های موجود در دیتاست پاسکال

کلاس	هواپیما	دوچرخه	پرنده	قایق
mAP	۰.۶۶	۰.۸	۰.۶۶	۰.۵۷
کلاس	اتوبوس	ماشین	گربه	صندلی
mAP	۰.۷۴	۰.۸۴	۰.۸۱	۰.۵۷
کلاس	میز ناهارخوری	گاو	سگ	اسب
mAP	۰.۶۴	۰.۶۹	۰.۷۲	۰.۷۶
کلاس	موتورسیکلت	انسان	مبل	گوسفند
mAP	۰.۸۱	۰.۸۱	۰.۶۴	۰.۶۹
کلاس	گیاه گلدانی	قطار	بطری	مانیتور
mAP	۰.۴۳	۰.۷۳	۰.۶	۰.۷

در جدول شماره ۳ نتایج روش پیشنهادی در مقایسه با کارهای انجام‌شده قبلی که همگی بر اساس دیتاست پاسکال ۲۰۱۲ و ۲۰۰۷ هستند، ارائه شده است.

بر اساس نتایج ارائه‌شده، روش پیشنهادی کارایی قابل‌قبولی داشته و در مقایسه با سایر کارهای انجام‌شده در این حوزه، از دقت بالاتری برخوردار است و دقت شناسایی و تشخیص اشیا در تصویر را به میزان $mAP = 0.8$ درصد بهبود داده است.

جدول ۳: مقایسه نتایج روش پیشنهادی با کارهای انجام‌شده قبلی

Method	Backbone	Train Data	mAP
Faster R-CNN[13]	VGG16	07+12	73.2
Faster R-CNN[13]	ResNet101	07+12	76.4
RetinaNet300[20]	ResNet101	07+12	62.9
SSD300[21]	VGG16	07+12	77.1
YOLOv3[18]	Darknet53	07+12	78.6
Proposed Method	Darknet53	07+12	79.4

- [15] Z. Shen, Z. Liu, J. Li, Y.-G. Jiang, Y. Chen, and X. Xue, "Dsoc: Learning deeply supervised object detectors from scratch," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1919–1927. Accessed: Nov. 13, 2023. [Online]. Available: http://openaccess.thecvf.com/content_iccv_2017/html/Shen_DSOD_Learning_Deeply_ICCV_2017_paper.html
- [16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [17] [17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [18] [18] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement." arXiv, Apr. 08, 2018. Accessed: Nov. 13, 2023. [Online]. Available: <http://arxiv.org/abs/1804.02767>
- [19] [19] 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.
- [20] [20] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988. Accessed: Dec. 13, 2023. [Online]. Available: http://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html
- [21] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016*, vol. 9905, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., in Lecture Notes in Computer Science, vol. 9905. , Cham: Springer International Publishing, 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.
- [7] A. Zanganeh and M. Jampour, "Automatic Weak Learners Selection for Pattern Recognition and its application in Soccer Goal Recognition," in *2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, IEEE, 2019, pp. 240–245.
- [8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, Ieee, 2001, p. I–I. Accessed: Nov. 13, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/990517/>
- [9] W. A. Ezat, M. M. Dessouky, and N. A. Ismail, "Evaluation of deep learning yolov3 algorithm for object detection and classification," *Menoufia Journal of Electronic Engineering Research*, vol. 30, no. 1, pp. 52–57, 2021.
- [10] K. Parviz, "Malware classification method using visualization features and word embedding based on deep learning," 2023, Accessed: Mar. 12, 2024. [Online]. Available: <https://www.sid.ir/paper/1083015>. (In Persian)
- [11] A. Zanganeh, M. Jampour, and K. Layeghi, "IAUFD: A 100k images dataset for automatic football image/video analysis," *IET Image Processing*, vol. 16, no. 12, pp. 3133–3142, Oct. 2022, doi: 10.1049/ipr2.12543.
- [12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 142–158, 2015.
- [13] 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.
- [14] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: Deconvolutional Single Shot Detector." arXiv, Jan. 23, 2017. Accessed: Nov. 13, 2023. [Online]. Available: <http://arxiv.org/abs/1701.06659>