



Detection of Iranian Foods in Images Using Deep Learning

ZAHRA HAJALIOGHLI¹ | SOLEIMAN HOSSEINPOUR² | SEYED SAEID MOHTASEBI³

1. Department of Agricultural Machinery Engineering, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran .E-mail: z.hajalioghli@ut.ac.ir
2. Department of Agricultural Machinery Engineering, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran E-mail: shosseinpour@ut.ac.ir
3. Department of Agricultural Machinery Engineering, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran E-mail: mohtaseb@ut.ac.ir

Article Info

Article type: Research Article

Article history:

Received: Oct. 12, 2023

Revised: Dec. 19, 2023

Accepted: Dec. 27, 2023

Published online: Autumn 2023

Keywords:

*Deep learning,
Food detection,
Instance segmentation,
YOLOv7*

ABSTRACT

Maintaining the well-being of individuals is greatly influenced by a healthy lifestyle and balanced diet. The identification and segmentation of food items can be improved by utilizing a mobile-based system in this era of rapid lifestyle changes and technology. This article introduces a novel system that, upon receiving input images, detects and segmentation the food items within the images. The system utilizes deep learning techniques and models, employing the YOLO algorithm. By incorporating regression-based simple methods, the system achieves the capability to detect and categorize food items in a single pass through the network, aiming to enhance accuracy and speed in the detection process. YOLOv7 was employed for food detection and YOLOv5, YOLOv7, and YOLOv8 was utilized for image segmentation. Based on the results, the accuracy, recall, and average precision values for YOLOv7 were 0.844, 0.924, and 0.932, respectively. Furthermore, the instance segmentation performance of YOLOv7 outperformed YOLOv5 and YOLOv8, with precision, recall, and mean average precision values of 0.959, 0.943, and 0.906, respectively. These findings underscore the high accuracy in detecting Iranian foods and the remarkable speed and precision in food image segmentation attainable through advanced deep-learning algorithms. Consequently, this study establishes that accurate detection of Iranian foods can be accomplished through the utilization of sophisticated deep-learning techniques. This research focuses on promoting a healthy lifestyle through intelligent technology and novel deep learning algorithms in Iran.

Cite this article: Hajalioghli, Zahra., Hosseinpour, Soleiman., & Mohtasebi, Seyed. saeid. (2023). Detection of Iranian Foods in Images Using Deep Learning., *Iranian Journal of Biosystem Engineering*, 54 (3), 19-41. <https://doi.org/10.22059/ijbse.2023.366560.665526>

© The Author(s).

Publisher: The University of Tehran Press.

DOI: <https://doi.org/10.22059/ijbse.2023.366560.665526>



شناسایی مواد غذایی ایرانی در تصاویر با استفاده از یادگیری عمیق

زهرا حاج علی اوغلی^۱ | سلیمان حسین پور^۲ | سید سعید محتسبی^۳

۱. گروه مهندسی ماشین‌های کشاورزی، دانشکده مهندسی و فناوری کشاورزی، دانشکدگان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران:

z.hajalioghli@ut.ac.ir

۲. نویسنده مسئول، گروه مهندسی ماشین‌های کشاورزی، دانشکده مهندسی و فناوری کشاورزی، دانشکدگان کشاورزی و منابع طبیعی، دانشگاه

تهران، کرج، ایران: shosseinpour@ut.ac.ir

۳. گروه مهندسی ماشین‌های کشاورزی، دانشکده مهندسی و فناوری کشاورزی، دانشکدگان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران:

mohhtaseb@ut.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۲/۷/۲۰

تاریخ بازنگری: ۱۴۰۲/۹/۲۸

تاریخ پذیرش: ۱۴۰۲/۱۰/۰۶

تاریخ انتشار: پاییز ۱۴۰۲

واژه‌های کلیدی:

تشخیص مواد غذایی،

بخش‌بندی نمونه‌ای،

یادگیری عمیق،

YOLO

سبک زندگی سالم و رژیم غذایی متعادل نقش حیاتی در حفظ سلامت انسان‌ها ایفا می‌کند. در این دوره از تغییر سریع سبک زندگی و فناوری، یک سیستم تشخیص و بخش‌بندی مواد غذایی مبتنی بر موبایل که مواد غذایی را شناسایی کند، می‌تواند بسیار مفید باشد و عادات غذایی را بهبود بخشد. در این مقاله یک سیستم جدید ارائه شده است که با دریافت تصویر ورودی، مواد غذایی داخل تصویر را تشخیص و بخش‌بندی می‌کند. این سیستم از تکنیک‌ها و مدل‌های یادگیری عمیق استفاده می‌کند. الگوریتم مورد استفاده YOLO است که با بهره‌مندی از روش‌های ساده مبتنی بر رگرسیون، توانایی تشخیص و بخش‌بندی مواد غذایی را با یک گذر از شبکه فراهم می‌آورد که با هدف بهبود دقت و سرعت در تشخیص ارائه شده است. این روش‌ها شامل استفاده از YOLOv7 برای تشخیص مواد غذایی و استفاده از بخش‌بندی نمونه‌ای YOLOv5، YOLOv7 و YOLOv8 برای بخش‌بندی تصاویر است. علاوه بر این، مجموعه داده‌ای از غذاهای ایرانی حاوی مواد غذایی مختلف تهیه و مورد استفاده قرار گرفت. بر اساس نتایج، مقادیر دقت، یادآوری و دقت متوسط میانگین YOLOv7 به ترتیب ۰/۸۴۴، ۰/۹۲۴ و ۰/۹۳۲ به دست آمد. همچنین، عملکرد بخش‌بندی نمونه‌ای YOLOv7 نسبت به YOLOv5 و YOLOv8 بهتر بود که مقادیر دقت بخش‌بندی، یادآوری و دقت متوسط میانگین ۰/۵ برای YOLOv7 به ترتیب ۰/۹۵۹، ۰/۹۴۳ و ۰/۹۰۶ است. نتایج حاکی از آن هستند که روش پیشنهاد شده در این مقاله دقت بالا در تشخیص مواد غذایی ایرانی و همچنین سرعت و دقت بالا در بخش‌بندی نمونه‌ای را فراهم می‌کند. بنابراین با استفاده از الگوریتم YOLO، می‌توان غذاهای ایرانی را با دقت بالا تشخیص داد و تصاویر آن‌ها را تقسیم بندی کرد. این پژوهش از طریق تکنولوژی هوشمند و الگوریتم‌های جدید یادگیری عمیق به ترویج سبک زندگی سالم از طریق تکنولوژی هوشمند در ایران می‌پردازد.

استناد: حاج علی اوغلی، زهرا؛ حسین پور، سلیمان؛ و محتسبی، سید سعید (۱۴۰۲). شناسایی مواد غذایی ایرانی در تصاویر با استفاده از یادگیری عمیق، *مجله مهندسی*

بیوسیستم ایران، ۵۴ (۳)، ۴۱-۱۹. <https://doi.org/10.22059/ijbse.2023.366560.665526>



© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijbse.2023.366560.665526>

مقدمه

در دنیای امروز، با توسعه صنعت فرآوری مواد غذایی و افزایش جمعیت جهان، نیاز به تغذیه سالم و مدیریت بهینه منابع اهمیت بیش‌تری پیدا کرده‌است. یکی از راهکارهای مؤثر در این زمینه، بهره‌گیری از تکنولوژی‌های هوش مصنوعی و یادگیری ماشین است. از جمله کاربردهای یادگیری ماشین در این حوزه می‌توان به بهینه‌سازی زنجیره تامین، انتخاب محصول، پیش‌بینی نگهداری، تشخیص بیماری محصول، پیش‌بینی عملکرد محصول و ردیابی مواد غذایی اشاره کرد. این تکنولوژی امکان تحلیل دقیق داده‌ها و پیش‌بینی رویدادهای آینده را فراهم می‌سازد که منجر به افزایش بهره‌وری، کاهش ضایعات و ارتقاء کیفیت محصولات می‌شود (Pallathadka et al., 2022). با توجه به اهمیت تغذیه سالم و نیاز به توسعه پایدار در حوزه صنایع غذایی، ایجاد یک سیستم هوشمند برای شناسایی و بخش‌بندی مواد غذایی ایرانی تبدیل به یک ضرورت مهم شده است. این سیستم پیشنهادی نقش اساسی در افزایش آگاهی افراد و ترویج تغذیه سالم بازی خواهد کرد.

موضوع شناسایی و بخش‌بندی مواد غذایی ایرانی از اهمیت ویژه‌ای برخوردار است، زیرا تشخیص دقیق و سریع مواد غذایی در تصاویر می‌تواند به بهبود سلامت عمومی و کیفیت محصولات غذایی کمک کند. در سال‌های اخیر، علی‌رغم پیشرفت‌های پزشکی، بیماری‌های مزمن ناشی از عادات غذایی نامنظم و سبک زندگی ناسالم هنوز چالش بزرگی برای سلامت جامعه هستند. چاقی، فشارخون بالا، چربی خون، بیماری‌های قلبی عروقی، قند خون و انواع سرطان‌ها از جمله بیماری‌های مزمن شایع هستند که رابطه مستقیمی با رژیم غذایی دارند (Tahir & Loo, 2021). برای پیشگیری یا کنترل این بیماری‌ها، لازم است که افراد مصرف مواد مغذی خود را ثبت و ارزیابی کنند. اما بسیاری از برنامه‌های کاربردی رژیم غذایی که در دسترس هستند، نیاز به وارد کردن دستی اطلاعات غذایی مصرف شده دارند که این کار هم زمان‌بر است و هم خطای زیاد دارد. به همین دلیل، کاربران معمولاً از استفاده طولانی‌مدت از این برنامه‌ها منصرف می‌شوند و به گزارش‌های نادقیق خود درباره دریافت کالری تکیه می‌کنند.

راه‌حل این مسئله استفاده از الگوریتم‌ها و سامانه‌های تشخیص مواد غذایی مبتنی بر دید است. با بهره‌گیری از تکنیک‌های تصویربرداری دیجیتال و یادگیری عمیق، امکان شناسایی و بخش‌بندی مواد غذایی با دقت بالا و به‌دست‌آوردن اطلاعات مفید مرتبط با آن‌ها فراهم می‌شود. با توجه به پیشرفت‌های اخیر در یادگیری عمیق، می‌توان از این رویکرد برای تشخیص و بخش‌بندی مواد غذایی بهره‌برداری کرد. این سیستم‌ها علاوه بر کاهش بار کاربران، دقت و کارایی آن‌ها را نیز بهبود می‌بخشند (Matsuda & Yanai, 2012). تشخیص غذا یک وظیفه پیچیده در حوزه بینایی ماشین است که با چالش‌هایی همچون تغییرات در ساختار و منبع غذا، شباهت ترکیبات مختلف به یکدیگر و تأثیر روش‌هایی مختلف پخت روبرو است. در نتیجه، تشخیص و بخش‌بندی غذا به عنوان یک موضوع تحقیقاتی جذاب در بینایی ماشین مورد توجه قرار گرفته است (Jiang et al., 2020; Kaur et al., 2019). بینایی ماشین نیز به عنوان یک زیرشاخه از یادگیری ماشین، با توسعه الگوریتم‌های یادگیری عمیق مانند شبکه‌های عصبی کانولوشنال (CNN)، رشد چشمگیری داشته است (Sezer & Altan, 2021; Sezer et al., 2021; Sun et al., 2020). بینایی کامپیوتر شامل وظایف متنوع و متعددی است که از جمله آن‌ها می‌توان به طبقه‌بندی تصویر، تشخیص اشیاء، بخش‌بندی معنایی و بخش‌بندی نمونه‌ای اشاره کرد (Shima et al., 2018). در این راستا، شبکه‌های عصبی کانولوشنال عمیق مانند AlexNet و EfficientNet به عنوان رایج‌ترین روش برای استخراج ویژگی‌های تصویر استفاده می‌شوند (Krizhevsky et al., 2012; Tan & Le, 2019).

در این پژوهش، بر روی دو تلاش عمده تحقیقاتی تمرکز شده است: (۱) توسعه الگوریتم‌های جدید تشخیص مواد غذایی، از جمله الگوریتم‌های شناسایی تصویر مواد غذایی بر اساس یادگیری عمیق و (۲) الگوریتم‌های پیش‌پردازش و بخش‌بندی تصویر.

با توجه به بررسی‌های صورت گرفته، این اولین مطالعه در مورد سیستم بخش‌بندی و شناسایی مواد غذایی ایرانی است که نه تنها از ۱۰۰۰۰ تصویر استفاده می‌کند، بلکه تحت شرایط مختلف مانند استفاده از نور و زاویه‌های مختلف نیز استفاده می‌شود. همچنین از انواع مواد غذایی مانند غذای جامد یا مایع و غذای مخلوط یا غیرمخلوط استفاده شده است. اکثر کارهای موجود از مجموعه داده‌های غذایی آماده استفاده می‌کنند و همچنین تغییرات شرایط فوق را در نظر نمی‌گیرند. برای این منظور استفاده از شبکه‌های عصبی با یادگیری عمیق را به عنوان ابزاری برای بهبود دقت در تشخیص مواد غذایی ایرانی پیشنهاد می‌کنیم. این سیستم پیشنهادی قصد دارد گامی به سوی تناسب اندام و ایجاد آگاهی در بین افراد برای تغذیه سالم باشد. برنامه پیشنهادی به شما کمک می‌کند کاربر بتواند مواد غذایی را به راحتی

تشخیص دهد.

ساختار این مقاله به شرح زیر است: در ادامه این بخش، کارهای پژوهشی مرتبط با تشخیص و بخش‌بندی مواد غذایی مورد بررسی قرار می‌گیرد. در بخش سوم، روش‌های یادگیری عمیق پیشرفته که در این مقاله به کار گرفته شده‌اند شرح داده می‌شود. در بخش چهارم، نتایج آزمایش‌ها و ارزیابی عملکرد روش‌های پیشنهادی نشان داده می‌شود. در بخش پنجم، جمع‌بندی و نتیجه‌گیری از کار انجام شده و پیشنهادهایی برای کارهای آتی ارائه می‌گردد.

پیشینه پژوهش

در فرایند تشخیص، هدف شناسایی طبقه و مکان شیء موردنظر در تصویر است که با یک جعبه محدودکننده نشان داده می‌شود. مدل‌های تشخیص شیء به دو دسته یک مرحله‌ای و دو مرحله‌ای تقسیم می‌شوند (Afzaal et al., 2021). مدل‌های یک مرحله‌ای همانند YOLO، Efficientdet و CenterNet با استفاده از یک شبکه عصبی کانولوشنال عمیق، هم‌زمان کلاس و جعبه محدودکننده شیء را پیش‌بینی می‌کنند (Redmon et al., 2016a; Tan et al., 2020; Zhou et al., 2019). مدل‌های دو مرحله‌ای همچون R-CNN، Fast R-CNN و Faster R-CNN و شبکه هرمی ویژگی (FPN¹) ابتدا با استفاده از یک الگوریتم پیشنهاد منطقه، ناحیه‌های حاوی شیء را در تصویر شناسایی کرده و سپس با استفاده از یک شبکه عصبی کانولوشنال عمیق، کلاس و جعبه محدودکننده آن‌ها را پردازش می‌کنند (Lin et al., 2017; Girshick et al., 2014). در پژوهشی یک سیستم نظارت خودکار رژیم غذایی را برای مشتریان غذاخوری بر اساس تشخیص چند غذا ارائه شد. برای کاهش ناحیه‌ای که برای تشخیص نیاز به پردازش دارد، یک روش محلی‌سازی را به کار بردند. سپس با استفاده از چندین توصیفگر بصری از قبیل توصیفگر رنگی، آماری، فضایی فرکانس یا طیفی، ساختاری و ترکیبی، نوع و طبقه غذا را تعیین کردند. با توجه به منوهای ثابت هفتگی و شرایط کنترل شده در ثبت تصویر غذا، صحنه غذاخوری دارای پیچیدگی کم است. مقدار ثابتی غذا ارائه می‌شود و سیستم می‌تواند با اندازه‌گیری باقیمانده غذا، میزان غذای مصرفی را تخمین بزند (Ciocca et al., 2015). Pouladzadeh & Shirmohammadi (2017) سیستم خود را بهبود بخشیدند و از جستجوی انتخابی برای تولید مناطق تشخیص برای ۳ ماده غذایی نان، سیب و موز استفاده کردند و دقتی برابر ۹۹ درصد به دست آوردند. اما این روش نیاز به زمان و منابع محاسباتی زیادی دارد. برای حل این مشکل، بسیاری از تحقیقات از Faster R-CNN برای بهینه‌سازی پیشنهادی منطقه استفاده کرده‌اند (Ren et al., 2017).

در یک بررسی از مدل Faster R-CNN برای استخراج مناطق مورد علاقه (ROIs²) و بهبود کارایی مدل تشخیص استفاده شد. سپس با استفاده از یک شبکه عصبی کانولوشنال عمیق، اقدام به استخراج نقشه‌های ویژگی جهت تشخیص غذا در مجموعه داده‌های UEC-Food100، UEC-Food256 و Food20-With-BBX شد. نتایج به دست آمده به شرح زیر است: دقت متوسط برای مجموعه داده UEC-Food100 برابر با ۱۷/۵ درصد، دقت متوسط برای مجموعه داده UEC-Food256 برابر با ۱۰/۵ درصد و دقت برای مجموعه داده Food20-With-BBX برابر با ۷۱/۷ درصد به دست آمد. برای هر مورد شناسایی شده در تصویر غذا، وزن پایه هر ماده غذایی را ۴۰۰ گرم فرض می‌کند و سپس گزارش ارزیابی رژیم غذایی را برای کاربر ارائه می‌کند (L. Jiang et al., 2020). علاوه بر این، با استفاده از Faster R-CNN، مرحله محلی‌سازی غذا با انتخاب مناطق حاوی اقلام غذایی در یک تصویر ورودی، آموزش داده شد. مناطق غذایی انتخاب شده به یک اندازه ثابت تغییر اندازه داده شد و به یک شبکه عصبی کانولوشنال عمیق برای طبقه‌بندی مواد غذایی وارد شده‌اند. بر اساس ویژگی‌های استخراج شده از شبکه عصبی کانولوشنال عمیق، بخش‌بندی دسته‌های غذایی مشابه برای ایجاد یک ساختار سلسله‌مراتبی انجام شده است. دقت ۶۵ درصدی حاصل شده و خروجی نهایی شامل جعبه‌های محدودکننده و برچسب‌های مواد غذایی برای هر غذا است (Mao et al., 2021).

با توسعه یادگیری عمیق، مدل Mask R-CNN که یک توسعه از Faster R-CNN است، پیشنهاد شد (He et al., 2017). این مدل علاوه بر تشخیص و مکان‌یابی شیء، قابلیت بخش‌بندی نمونه‌ای را نیز دارد. با الهام از این روش، Qiu et al. (2019) مصرف رژیم غذایی را بر اساس ویدئوهای صحنه‌های غذا خوردن، تخمین زده‌اند. در ویدئوهای ضبط شده آنها، ۲-۳ کاربر با هم غذا به اشتراک می‌گذاشتند. آنها تشخیص غذا را با فناوری تشخیص چهره و ردیابی دست ادغام کردند و از Mask R-CNN با تنظیم دقیق برای شناسایی

۱۴ دسته غذا موجود در مجموعه داده COCO، Food-101 و افراد استفاده کردند. در پژوهشی، مقایسه‌ای بر اساس کارایی بین چندین الگوریتم مانند ResNet-50، VGG-16، ImageNet و Inception مورد استفاده در توسعه طبقه‌بندی غذا صورت گرفت. نتایج نشان داد که الگوریتم YOLO عملکرد بهتری نسبت به سایر تکنیک‌ها دارد (Jamnekar et al., 2023).

Kaur et al. (2023) در تحقیقی برای تشخیص غذا در تصاویر از مدل YOLOV5 استفاده نمودند. مجموعه‌داده‌ای با ۳ کلاس غذایی (۵۰۰ تصویر غذا) که از FOOD-101 مشتق شده است، استفاده شده است. این مجموعه‌داده در قالب YOLOV5 با حاشیه‌نویسی جعبه محدود^۱ تهیه شده است. دقت^۲، یادآوری^۳ و دقت متوسط میانگین^۴ برای همه کلاس‌ها به ترتیب ۶۷، ۶۳ و ۶۶ درصد به دست آمد. بخش‌بندی تصویر، یک فرایند مهم در تشخیص و شناسایی مواد غذایی است که هر پیکسل تصویر را با یک برچسب خاص مشخص می‌کند تا پیکسل‌هایی با ویژگی‌های مشابه و برچسب‌های مشترک، در یک منطقه بخش‌بندی شده جمع‌آوری شوند. بخش‌بندی، نقش مهم‌تری نسبت به تشخیص غذاها ایفا می‌کند (Ando et al., 2019; Dehais et al., 2015). در مطالعات گذشته، بخش‌بندی مواد غذایی بر اساس فرض‌های خاصی انجام شده است که باعث ایجاد تفاوت‌هایی در دقت و کارایی روش‌ها می‌شود. به‌عنوان مثال، برخی از مطالعات اولیه فرض کرده‌اند که بشقاب‌ها دارای رنگ یا شکل خاصی هستند و غذاها روی هم قرار نمی‌گیرند. این فرض‌ها می‌توانند به دقت بخش‌بندی غذا و پس‌زمینه کمک کنند، اما محدودیت‌های آنها از لحاظ تعمیم‌پذیری و عملکرد در شرایط مختلف باعث شد تا روش‌های بر اساس آستانه‌گذاری و فضای رنگ YCbCr مورد استفاده قرار گیرند.

به‌علاوه، روش‌های پیشرفته‌تر و مبتنی بر یادگیری عمیق نیز برای بخش‌بندی مواد غذایی مورد استفاده قرار گرفته‌اند. استفاده از الگوریتم‌های رشد منطقه و ادغام و مدل‌های مبتنی بر CNN مانند U-Net و Mask-RCNN از جمله این روش‌ها است. این روش‌ها به‌طور مؤثری نتایج بخش‌بندی و دقت در اندازه‌گیری کالری و مواد مغذی غذا را بهبود داده‌اند (Lo et al., 2020; Ando et al., 2019; He et al., 2017; Okamoto & Yanai, 2016; Pouladzadeh et al., 2016; Weng & Zhu, 2015).

به همین دلیل، تحقیقات جدید به‌منظور بهبود در بخش‌بندی مواد غذایی ارائه شده‌اند، به‌ویژه با بهره‌گیری از چارچوب Mask-RCNN و رویکردهای بخش‌بندی بافت و بازسازی سه‌بعدی. سیستم پیشنهادی با نام "goFOOD TM" توانایی تخمین مقدار کالری و مواد مغذی ماکرو یک وعده غذایی را از تصاویر یا ویدئوهای گرفته‌شده توسط گوشی‌های هوشمند دارد. روش ارائه شده با استفاده از بخش‌بندی مواد غذایی، شناسایی مواد غذایی و تخمین حجم غذا را انجام می‌دهد. استفاده از پایگاه داده‌های ترکیبات غذایی و چارچوب Mask-RCNN در این روش به نتایج قابل قبولی در پردازش تصاویر غذا منجر می‌شود (Lu et al., 2020). همچنین، برخی از مطالعات اخیر از روش‌های پیچیده‌تر مبتنی بر یادگیری عمیق بهره‌مند شده‌اند. به‌عنوان مثال، در تحقیقی جهت مطالعه میزان تغذیه بیماران بستری در بیمارستان، الگوریتم‌های هوش مصنوعی جدیدی توسعه داده شد و شبکه‌های چند وظیفه‌ای برای بخش‌بندی هم‌زمان بشقاب‌ها و غذاها پیشنهاد شد. نتایج این تحقیق نشان‌دهنده همبستگی بالایی ($< 0/91$) با حقیقت زمین و خطای نسبی کمتر از ۲۰ درصد در تخمین مصرف مواد مغذی بوده‌اند (Lu et al., 2021).

به‌طور جداگانه، الگوریتم‌های مبتنی بر یادگیری عمیق برای تعیین محل غذا و دسته‌بندی معنایی غذا ارزیابی شدند. این مطالعه از روش‌های بخش‌بندی غذا/بدون غذا و بخش‌بندی معنایی استفاده کرده است و یک مجموعه‌داده جدید با ۵۰۰۰ تصویر از ۵۰ دسته غذایی ایجاد شده است. نتایج نشان می‌دهند که الگوریتم‌های مختلف از جمله ENet، GUN، ERFNet، SSNet، Edanet، ESPNet و SegNet، DeepLab-v2 و Pixel Objectnes عملکرد متفاوتی در برابر تغییرات نور، فشرده‌سازی JPEG، نویز گاوسی و مات شده گاوسی دارند. به‌طور خاص، بهترین عملکرد برای هر دو وظیفه تعیین محل و بخش‌بندی معنایی به GUN با دقت ۹۵/۳ درصد برای تصاویر اصلی، ۹۴ درصد بر روی تصاویر با تغییر نور، ۹۴/۴ درصد بر روی تصاویر با فشرده‌سازی JPEG و ۹۳/۳ درصد بر روی تصاویر با نویز گاوسی و ۹۳/۶ درصد مات شده گاوسی تعلق دارد (Aslan et al., 2020).

با توجه به بررسی‌های صورت گرفته، برخی از مطالعات هنوز دقت بخش‌بندی را به طور جامع ارزیابی نکرده‌اند و فقط از آن به‌عنوان بخشی از روش کلی خود استفاده می‌کنند. همچنین، بعضی از روش‌ها تنها بر روی مواد غذایی خاصی اعمال شده‌اند و در موارد گسترده‌تری که شامل انواع مختلف مواد غذایی هستند، مورد استفاده قرار نگرفته‌اند. این محدودیت‌ها باعث می‌شود تا اندازه‌گیری دقیق مواد غذایی با

1 . Bounding Box
2 . Precision
3 . Recall
4 . mean Average Precision

چالش‌های بیشتری مواجه شود.

از این رو، تحقیقات اخیر نشان می‌دهند که تشخیص و بخش‌بندی مواد غذایی موضوع حائز اهمیت است و کاربردهای مفید آن از جمله مدیریت رژیم غذایی و نظارت بر تغذیه بسیار گسترده است. با پیشرفت یادگیری عمیق، معماری‌های تشخیص اشیاء مبتنی بر شبکه‌های عصبی کانولوشنال به‌عنوان روش‌های پیشرو در این زمینه شناخته شده‌اند. در مطالعه‌ای، شبکه عصبی یادگیری عمیق به‌عنوان روشی غیرمخرب و برخط جهت تشخیص چهار گونه مهم و اقتصادی خانواده کپور ماهیان شامل کپور معمولی، کپور علفخوار، کپور سرگنده و کپور نقره‌ای ایجاد و مورد استفاده قرار گرفت (طاهری گراوند و همکاران، ۱۴۰۰). در تحقیقی دیگر، یک روش ساده و خودکار با استفاده از بخش‌بندی تصویر توسعه داده شد که با استفاده از ترکیب ادغام منطقه پویا و الگوریتم ژنتیک، شاخص منطقه برای هر پروفیل خاک محاسبه شد (Mahmoodi-Eshkaftaki et al., 2020). مدل‌های YOLO به‌عنوان آشکارسازهای شناخته‌شده با قابلیت بلادرنگ و دقت بالا در کاربردهای تشخیص مواد غذایی بخصوص در محیط‌های پرسرعت مورد استفاده قرار می‌گیرند. نهایتاً، با توجه به اهمیت تشخیص و بخش‌بندی مواد غذایی، تحقیقات بیشتری در این زمینه نیاز است تا بتوان به روش‌ها و الگوریتم‌های جدیدی دست یافت که دقت کارایی این فرایند را بهبود دهند و بتوانند با چالش‌های موجود در تصاویر مختلف غذاها مواجه شوند (Redmon et al., 2016a).

با توجه به مطالعات پیشین تاکنون، تحقیقاتی بر روی تشخیص و بخش‌بندی مواد غذایی ایرانی انجام نشده است و همچنین مطالعات قبلی به مدل YOLOv7 برای این منظور توجه نداشته‌اند. بنابراین تحقیق حاضر در جهت ارزیابی مواد غذایی، اقدام به ساخت مجموعه داده‌ای از غذاهای ایرانی کرده و با بررسی و تحلیل الگوهای الگوریتم‌های جدید در نمونه‌های مواد غذایی ایرانی، راهکارهای نوینی برای تشخیص و بخش‌بندی آن‌ها ارائه می‌دهد. از مدل YOLOv7 در مجموعه داده‌ای شامل مواد غذایی ایرانی برای این هدف استفاده شده است. همچنین، روش‌های بخش‌بندی نمونه‌های YOLOv5، YOLOv7، و YOLOv8 برای بخش‌بندی مواد غذایی استفاده شده‌اند.

مواد و روش‌ها

در این پژوهش، یک سیستم تشخیص و بخش‌بندی مواد غذایی در تصاویر ارائه شده است که با استفاده از شبکه‌های عصبی کانولوشنال (CNN)، نام و مکان هر ماده غذایی را در تصویر تعیین می‌کند. برای این منظور، از معماری YOLO که دارای دقت و سرعت بالایی در تشخیص شیء است، بهره گرفته شده است. همچنین، از روش‌های بخش‌بندی نمونه‌های YOLOv5، YOLOv7، و YOLOv8 که قادر به شناسایی و محلی‌سازی دقیق مواد غذایی در تصاویر هستند، استفاده شده است. خروجی سیستم شامل نام و ماسک پیش‌بینی شده برای هر ماده غذایی در تصویر است.

فرایند پژوهش شامل چند مرحله است: جمع‌آوری و پیش‌پردازش داده‌ها، آموزش و ارزیابی مدل. جمع‌آوری داده‌ها شامل ایجاد یک مجموعه داده از عکس‌های مواد غذایی با برچسب‌گذاری مناسب است. پیش‌پردازش داده‌ها شامل اعمال تکنیک‌های تقویت داده^۱ بر روی تصاویر غذا برای جلوگیری از بیش‌برازش و بهبود عملکرد مدل است. آموزش مدل شامل استفاده از مدل‌های CNN از پیش آموزش‌دیده بر روی مجموعه داده جمع‌آوری شده برای چارچوب YOLO است. این مدل قادر به استخراج ویژگی‌های قابل توجه از تصاویر با استفاده از لایه‌های کانولوشن و لایه‌های ادغام است. ارزیابی مدل شامل بررسی و آزمایش عملکرد و دقت مدل بر روی مجموعه داده‌های مختلفی از عکس‌های مواد غذایی است. پس از پردازش خروجی‌های پیش‌بینی شده، ماسک‌های بخش‌بندی به منظور بهبود دقت تشخیص و بخش‌بندی مواد غذایی به کار گرفته می‌شود. نتیجه نهایی مدل شامل مجموعه‌ای از ماسک‌های بخش‌بندی است که با مواد غذایی موجود در تصویر مطابقت دارند. در کل، روش‌های بخش‌بندی نمونه‌های YOLOv5، YOLOv7، و YOLOv8 ابزارهای قدرتمند و سازگاری برای تشخیص و بخش‌بندی مواد غذایی در تصاویر هستند (Achanta et al., 2012; Rajayogi et al., 2019).

ساخت مجموعه داده تصاویر

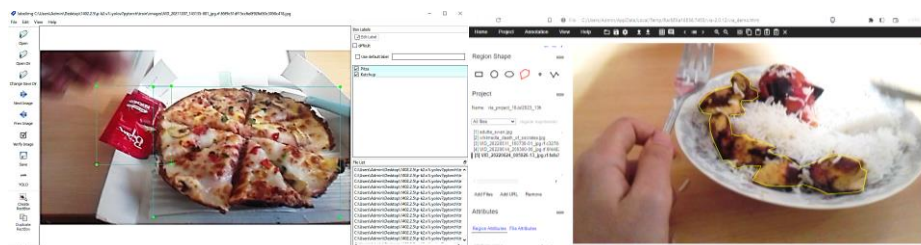
در این پژوهش برای ساخت مجموعه داده تصاویر غذایی از غذاهای رستوران دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران به‌علاوه چند غذای ایرانی دیگر استفاده شده است. تصاویر با دوربین گوشی Huawei Ascend P7 در زمان‌ها، نورها و فواصل مختلف عکس‌برداری گرفته شده‌اند تا مدل شبکه را با شرایط متنوع آموزش دهند. مجموعه داده شامل ۱۲۰۰۰ تصویر با اندازه پیکسل ۷۲۰*۱۲۸۰ است که بعد

از حذف تصاویر نامعتبر، ۱۰۰۰۰ تصویر برای مدل تشخیص مواد غذایی و ۱۴۱۳ تصویر برای مدل بخش‌بندی مواد غذایی استفاده شد. تعداد تصاویر برای مدل تشخیص و مدل بخش‌بندی در جدول ۱ آورده شده است.

هر ماده غذایی در هر تصویر با نرم‌افزار LabelImg به صورت دستی برایش تشخیص برچسب گذاری شده است. همچنین، برای بخش‌بندی، با استفاده از نرم‌افزار LabelMe، هر ماده غذایی با یک شکل چند ضلعی انتخاب شده و ماسک حقیقت زمین^۱ ساخته شده است. شکل ۱ فرایند برچسب‌گذاری و ساخت ماسک را برای یک تصویر نمونه نشان می‌دهد. سپس، مجموعه داده به سه بخش آموزش (۷۰ درصد)، اعتبارسنجی (۲۰ درصد) و آزمایش (۱۰ درصد) تقسیم شده است. عکس نهایی با اندازه ۴۱۶*۴۱۶ به ورودی مدل داده شده است. شکل ۲ نمونه‌ای از تصاویر مجموعه داده بخش‌بندی را در شرایط مختلف نشان می‌دهد (Dutta & Zisserman, 2019).

جدول ۱: تعداد تصاویر مربوط به مدل‌ها

تعداد کل تصاویر	آموزش	اعتبارسنجی	آزمایش
۱۰۰۰۰	۷۰۰۰	۲۰۰۰	۱۰۰۰
۱۴۱۳	۹۹۰	۲۸۲	۱۴۱



شکل ۱: فرایند برچسب‌گذاری با استفاده از نرم‌افزارهای LabelMe و LabelImg



شکل ۲: تعدادی از تصاویر برچسب‌گذاری شده مجموعه داده بخش‌بندی

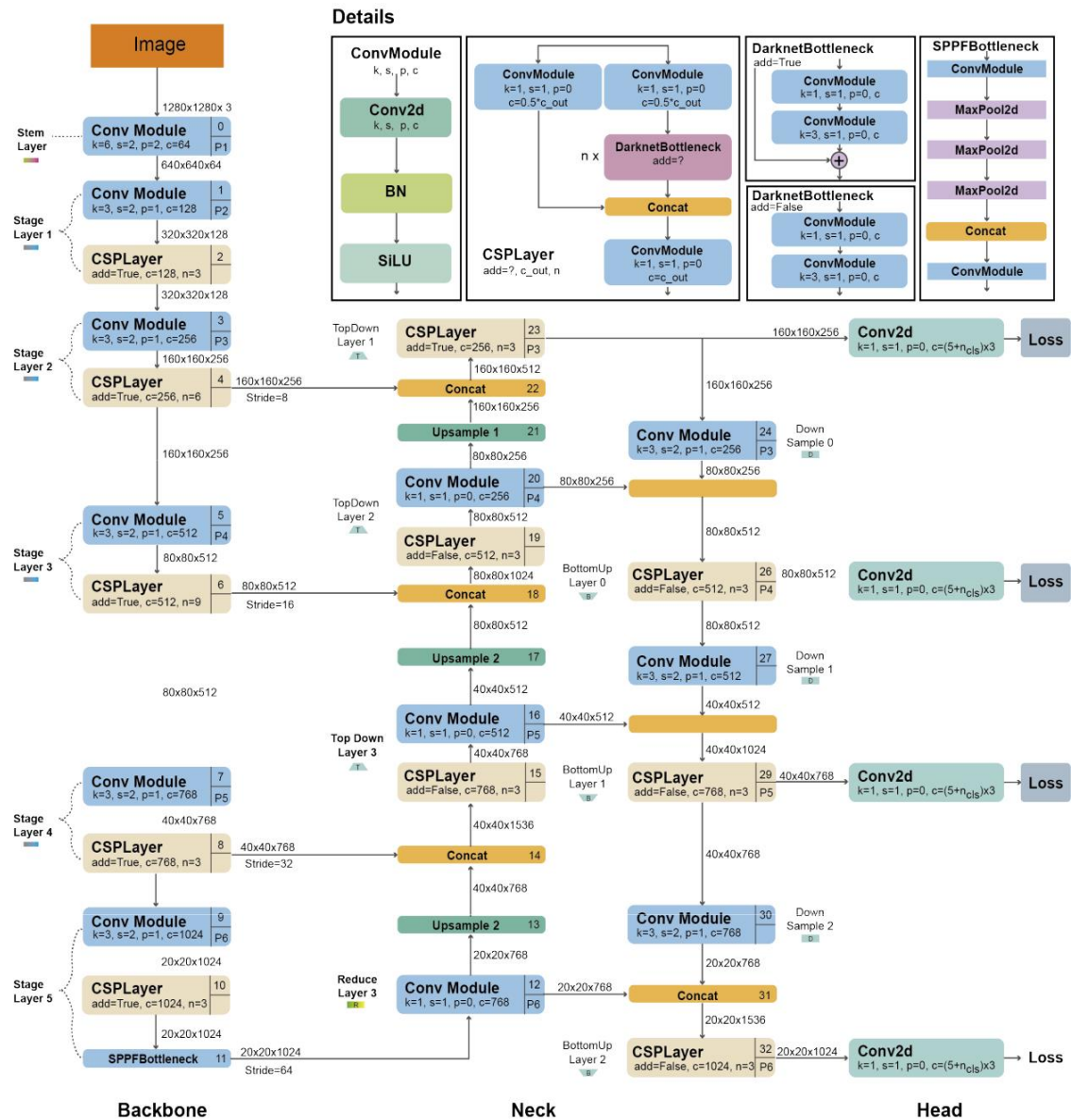
معماری مدل‌ها

YOLO توسط جوزف ردمن و همکاران در CVPR 2016 معرفی شد. این روش برای اولین بار رویکرد زمان واقعی و انتها به انتها برای تشخیص اشیا ارائه داد. نام YOLO مخفف "You Only Look Once" است و به این واقعیت اشاره دارد که بر خلاف روش‌های قبلی که از پنجره‌های کشویی و یک طبقه‌بندی‌کننده استفاده می‌کردند و نیاز به صدها یا هزاران اجرا داشتند، YOLO توانایی انجام وظیفه تشخیص تنها با یک گذر از شبکه را داشت. این معماری با استفاده از روش ساده مبتنی بر رگرسیون برای پیش‌بینی خروجی‌های تشخیص عمل می‌کند، در مقابل روش Fast R-CNN که از دو خروجی جداگانه، از طبقه‌بندی برای احتمالات و رگرسیون برای مختصات جعبه محدود استفاده می‌کند (Ren et al., 2015; Redmon et al., 2016).

YOLOv5

YOLOv5 چند ماه پس از YOLOv4 در سال ۲۰۲۰ توسط Glen Jocher، بنیان‌گذار و مدیرعامل شرکت Ultralytics منتشر شد.

YOLOv5 بر اساس نسخه‌های قبلی خانواده YOLO ساخته شده است، اما به جای استفاده از Darknet به‌عنوان موتور اصلی از PyTorch استفاده می‌کند (YOLOv1- YOLOv4).



شکل ۳: معماری YOLOv5 (Jocher, 2020).

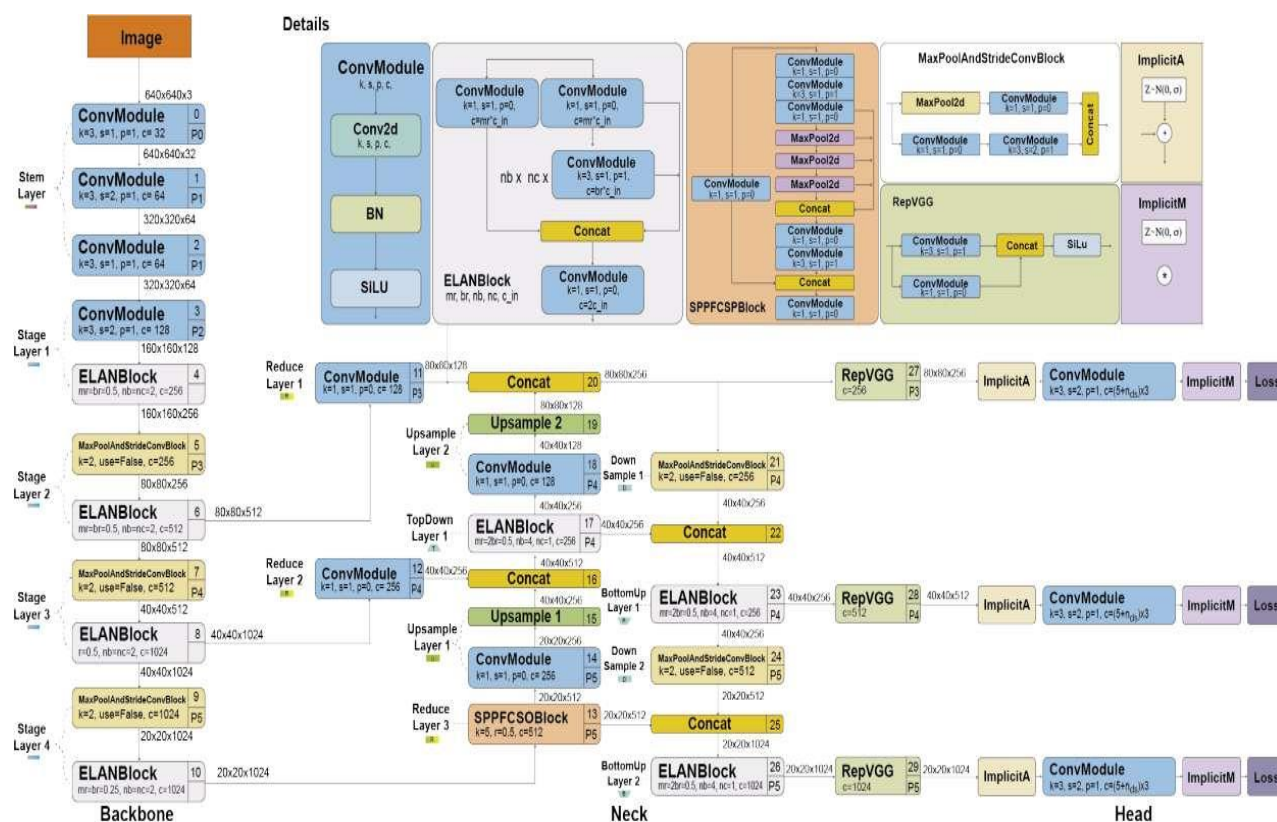
YOLOv5 از الگوریتم Ultralytics به نام AutoAnchor برای بهبود جعبه‌های لنگر استفاده می‌کند. این ابزار پیش‌آموزشی، جعبه‌های لنگر را در صورتی که اندازه تصویر مناسب برای مجموعه داده‌ها و تنظیمات آموزشی نیستند، بررسی و تنظیم می‌کند. برای ایجاد شرایط اولیه برای الگوریتم تکامل ژنتیکی (GE¹), ابتدا تابع k-means برای برجسب‌های مجموعه داده اعمال می‌شود و سپس GE این لنگرها را تکامل می‌دهد و از تابع تناسب بهترین فراخوان و اتلاف CIoU استفاده می‌کند. معماری مدل YOLOv5 شامل سه قسمت، یعنی ستون فقرات، گردن و سر است. YOLOv5 از CSP-Darknet53 به عنوان ستون فقرات خود استفاده می‌کند. این CSP-Darknet53، نسخه اصلاح شده‌ای از شبکه کانولوشن Darknet53 است که تکنیک شبکه چند مرحله‌ای جزئی (CSPⁿ) را در خود دارد. همچنین، YOLOv5 تکنیک CSPNet را پیاده‌سازی می‌کند که نقشه ویژگی لایه پایه را به دو قسمت تقسیم می‌کند و سپس با استفاده

1 . Genetic Evolution
2 . Cross-Stage Partial

از سلسله مراتب مرحله‌ای ترکیب می‌شوند. برای افزایش جریان اطلاعات و کمک به مکان‌یابی دقیق پیکسل‌ها در تابع پیش‌بینی ماسک، از PANet که یک شبکه هرمی و ویژگی استفاده می‌کند، بهره‌برداری می‌کند. همچنین، از بلوک SPP^۱ نیز استفاده می‌شود که اطلاعات دریافتی از ورودی‌ها را جمع‌آوری کرده و یک خروجی با طول ثابت ارائه می‌دهد. در نهایت، از YOLOv5head از سه لایه کانولوشن با اندازه‌های مختلف تشکیل شده است که موقعیت جعبه‌های محدود کننده، امتیازها و طبقه‌بندی اشیاء را تعیین می‌کند. شکل ۳ معماری دقیق YOLOv5 را نشان می‌دهد. به طور خلاصه، YOLOv5 یک الگوریتم موفق است که با انجام تغییرات و بهبودات در معماری تنظیمات مدل، دقت و سرعت تشخیص اشیاء را بهبود می‌بخشد (Islam et al., 2022; Xu et al., 2021; G. Jocher, 2020; C. Y. (Wang et al., 2020; Szegedy et al., 2015; He et al., 2014

YOLOv7

YOLOv7 یک نسخه بهبود یافته از آشکارسازهای شناخته شده YOLOR و YOLOv4، در سال ۲۰۲۲ ارائه شد و در بازه سرعت ۵ تا ۱۶۰ فریم بر ثانیه (FPS^۲) عملکرد بسیار بهتری از نسخه‌های قبلی این الگوریتم ارائه داد (Bochkovskiy et al., 2022). شکل ۴ معماری دقیق YOLOv7 را نشان می‌دهد. این نسخه از YOLOv7 دو تغییر اساسی در معماری دارد که ارتقا دقت تشخیص بدون کاهش سرعت استنتاج را ایجاد می‌کند. اولین تغییر معماری به شبکه تجمیع لایه کارآمد توسعه یافته یا E-ELAN اختصاص دارد. E-ELAN یک استراتژی هوش مصنوعی است که به مدل‌های عمیق اجازه می‌دهد تا با کنترل کوتاه‌ترین - طولانی‌ترین مسیر گردان، به صورت کارآمدتر همگرا شوند. YOLOv7 از این تکنیک برای مدل‌هایی با بلوک‌های محاسباتی انباشته نامحدود استفاده می‌کند. E-ELAN با ادغام ویژگی‌های گروه‌های مختلف و ادغام کاردیالیته^۳، به شبکه این امکان را می‌دهد تا یادگیری را بدون از بین بردن مسیر گردان اصلی، افزایش دهد.



شکل ۴: معماری YOLOv7.

تغییر معماری دیگر YOLOv7 مربوط به مقیاس‌بندی مدل است. در این معماری مبتنی بر الحاق، مدل‌هایی با اندازه‌های مختلف

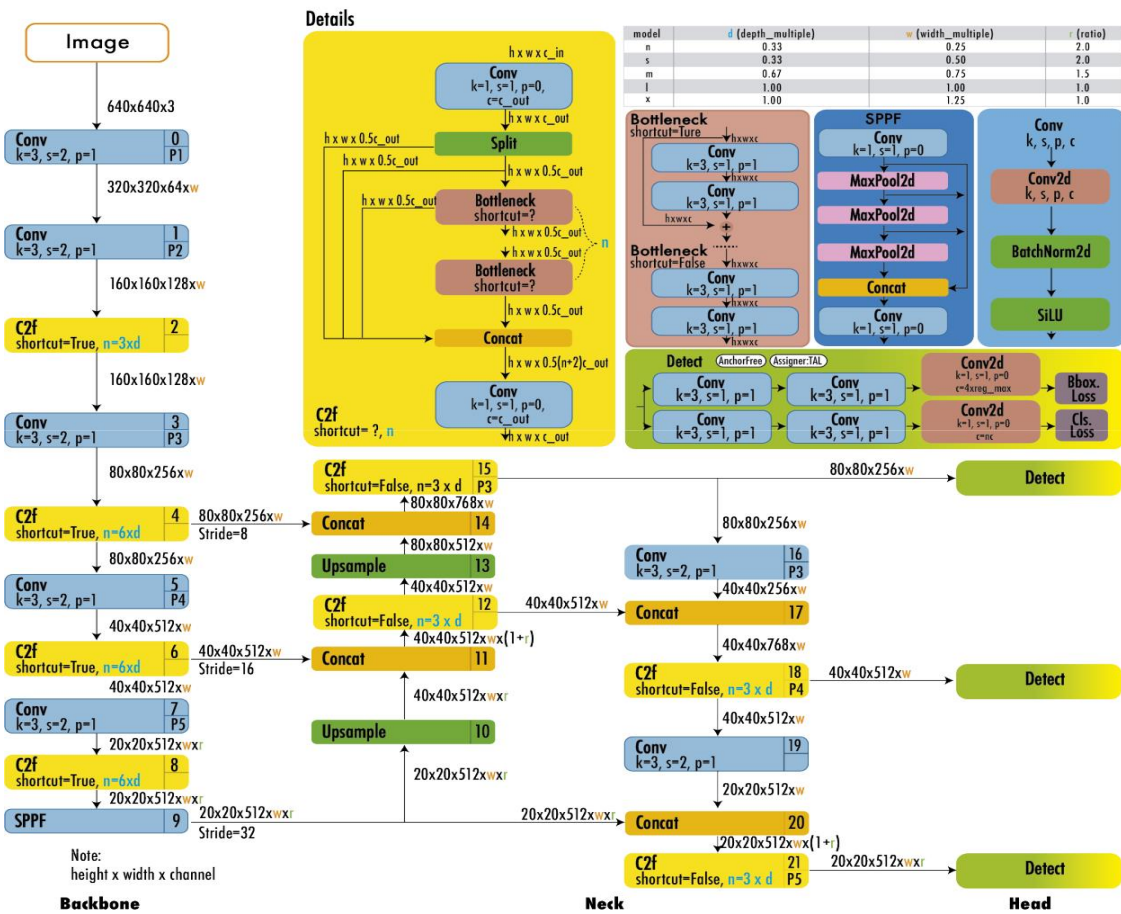
- 1 . Spatial Pyramid Pooling
- 2 . Frames Per Second
- 3 . Cardinality

تولید می‌شوند. این معماری از تکنیک‌های مقیاس‌بندی استاندارد مانند مقیاس‌گذاری عمقی استفاده نمی‌کند، بلکه یک استراتژی جدید برای مقیاس‌بندی مدل‌های مبتنی بر الحاق به ارمغان می‌آورد. در این استراتژی، عمق و عرض بلوک‌ها با همان فاکتور مقیاس‌بندی تغییر می‌کند که ساختار بهینه مدل را حفظ می‌کند (Wang et al., 2022).

در نتیجه، با این تغییرات YOLOv7 توانسته است دقت بیشتری در تشخیص اشیاء داشته باشد و در عین حال سرعت استنتاج آن کاهش نیافته و به عنوان یکی از بهترین آشکارسازهای شیء شناخته شود.

YOLOv8

YOLOv8 در ژانویه ۲۰۲۳ توسط Ultralytics، شرکت سازنده YOLOv5 منتشر شد (Jocher, 2023). این مدل پنج نسخه کوچک‌شده YOLOv8n (نانو)، YOLOv8s (کوچک)، YOLOv8m (متوسط)، YOLOv8l (بزرگ) و YOLOv8x (بسیار بزرگ) را ارائه می‌کند. YOLOv8 از چندین کار بینایی مانند تشخیص اشیاء، بخش‌بندی، تخمین وضعیت، ردیابی و طبقه‌بندی پشتیبانی می‌کند. شکل ۵ معماری دقیق YOLOv8 را نشان می‌دهد. این مدل از ستون فقرات مشابه YOLOv5 با تغییراتی در CSPlayer (به نام ماژول C2f) استفاده می‌کند. ماژول C2f ویژگی‌های سطح بالا را با اطلاعات متنی ترکیب می‌کند تا دقت تشخیص را بهبود بخشد.



شکل ۵: معماری YOLOv8 (Contributors, 2023)

YOLOv8 از یک مدل بدون لنگر با سر جداشده برای پردازش مستقل وظایف تشخیص وجود یک شیء، طبقه‌بندی و رگرسیون استفاده می‌کند. این طراحی به هر بخش اجازه می‌دهد تا بر روی وظیفه خود تمرکز کند و دقت کلی مدل را بهبود بخشد. در لایه خروجی YOLOv8، از تابع سیگموئید به عنوان تابع فعال‌سازی برای امتیاز شیء و از تابع سافت‌مکس برای احتمالات کلاس‌ها استفاده می‌شود. این توابع هزینه‌های CIoU^۱ و DFL^۲ را برای محدود کردن تابع هزینه جعبه محدود کننده و آنتروپی متقاطع باینری برای تابع هزینه طبقه‌بندی

1 . Complete Intersection over Union
2 . Distribution Focal Loss

به کار می‌برد. این توابع هزینه عملکرد تشخیص اشیاء، به ویژه در هنگام برخورد با اجسام کوچکتر را بهبود بخشیده‌اند (Zheng et al., 2019; Li et al., 2020).

علاوه بر این، YOLOv8 یک مدل بخش‌بندی معنایی به نام YOLOv8-Seg را نیز ارائه می‌دهد. این مدل از ستون فقرات CSPDarknet53 به‌عنوان استخراج‌کننده ویژگی و یک ماژول C2f استفاده می‌کند. ماژول C2f دارای دو سر بخش‌بندی است که یاد می‌گیرند ماسک‌های بخش‌بندی را برای تصویر ورودی پیش‌بینی کنند. این مدل با حفظ سرعت و کارایی بالا، به نتایج پیشرفته‌ای در معیارهای مختلف تشخیص اشیاء و بخش‌بندی دست یافته است. همچنین، دارای چندین ادغام برای برچسب‌گذاری، آموزش و به‌کارگیری است.

YOLOv8 را می‌توان از رابط خط فرمان (CLI) اجرا کرد و یا آن را به‌عنوان بسته PIP^۲ نصب کرد.

طراحی و پیاده‌سازی

در این پژوهش، از زبان برنامه‌نویسی پایتون و چارچوب Tensorflow برای پیاده‌سازی مدل استفاده شد. برای انجام فرایندهای آموزش و آزمایش مدل، محیط Google Colab مبتنی بر ابر و دارای دسترسی به GPU NVIDIA Tesla T4 مورد استفاده قرار گرفت. انتخاب این چارچوب و محیط برای پیاده‌سازی مدل به دلیل قدرت بالا و امکان استفاده از GPU^۳ برای شتاب بخشی به فرایند آموزش بوده است. در این مطالعه، رویکرد یادگیری انتقالی برای آموزش مدل انتخاب شد. به دلیل محدودیت‌های موجود در مجموعه داده تهیه شده، از وزن‌های از پیش آموزش دیده مدل YOLOv7 از مجموعه داده COCO^۴ به‌عنوان مقداردهی اولیه استفاده شد. این رویکرد امکان استفاده از اطلاعات یادگیری شده از مجموعه داده بزرگ COCO و انتقال این دانش به مجموعه داده کوچک‌تر موجود در این پروژه را فراهم می‌کند.

در طول مرحله آموزش مدل، از تکنیک‌های تقویت داده مختلف بر روی تصاویر مجموعه داده استفاده شد. تقویت داده تکنیکی برای افزایش مصنوعی مجموعه آموزشی با ایجاد نسخه‌های اصلاح شده با استفاده از داده‌های موجود است. این تکنیک به طور گسترده در حوزه‌های مختلف، مانند بینایی کامپیوتر و پردازش زبان طبیعی، برای بهبود عملکرد و نتایج مدل‌های یادگیری ماشین با ایجاد مجموعه داده‌های متنوع‌تر استفاده می‌شود. تکنیک‌های خاص تقویت داده به نوع داده‌های مورد تجزیه و تحلیل بستگی دارد. به عنوان مثال، در مورد تصویر، تکنیک‌هایی مانند چرخش، تغییر اندازه، برش و تنظیم روشنایی معمولاً استفاده می‌شود. استفاده از تقویت داده‌ها می‌تواند به کاهش بیش‌برازش و افزایش دقت مدل در مجموعه داده‌های دیده نشده کمک کند (Mumuni., 2022). در این پژوهش از چندین روش برای تقویت تصاویر مانند چرخش، تغییر اندازه، انتقال، چرخش از راست به چپ و ... با هدف تقویت قابلیت تعمیم مدل و جلوگیری از بیش‌برازش در فرایند آموزش استفاده شد. همچنین تعداد تکرارها ۴۰ و اندازه دسته ۳۲ و مقدار اولیه نرخ یادگیری ۰/۰۰۱ انتخاب گردید و بهینه‌ساز آدام نیز جهت به‌روزرسانی وزن‌ها در طول آموزش به کار گرفته شد.

جدول ۲ نماینده فرآیندهای مورد استفاده در مرحله آموزش مدل است.

lr0: نرخ یادگیری اولیه برای کنترل اندازه گام که در هر تکرار در طول آموزش استفاده می‌شود.

تکانه: برای بهینه‌ساز که بر میزان همگرایی در طول آموزش تأثیر می‌گذارد.

iou_t: آستانه IOU، برای تعیین آستانه IOU در طول آموزش استفاده می‌شود.

anchor_t: آستانه لنگر، بر آستانه لنگر در طول آموزش تأثیر می‌گذارد.

hsv_s, hsv_v, hsv_h: پارامترهای تقویت داده HSV برای رنگ، اشباع و مقدار که برای تقویت تصویر استفاده می‌شود.

translate, scale, shear, perspective: پارامترهای تقویت تصویر به ترتیب برای انتقال، مقیاس‌بندی، برش و تبدیل پرسپکتیو.

استفاده می‌شود.

flip_lr: پارامتر تقویت تصویر برای چرخش چپ-راست است.

mixup: برای ترکیب داده‌ها استفاده می‌شود، این روش به طور معمول با ایجاد ترکیب‌های محوری از نمونه‌ها و برچسب‌های آن‌ها،

عمل می‌کند. این روش به عنوان یک ابزار گسترده و ساده از تقویت داده‌ها استفاده می‌شود.

1 . Command Line Interface
2 . Python Installs Package
3 . Graphics Processing Unit
4 . Common Objects in Context

فراپارامترهایی که فرض شده بودند به طور قابل توجهی بر عملکرد تشخیص مدل تأثیر می‌گذارند برای آزمایش انتخاب شدند این فراپارامترها را می‌توان به عنوان فراپارامترهای "وزن و یادگیری" یا "تقویت داده" طبقه‌بندی کرد. مورد اول بر تعداد وزن‌های مدل یا نحوه به‌روزرسانی آن‌ها تأثیر می‌گذارد. مورد دوم تصاویر ورودی را در طول آموزش اصلاح می‌کند (Dehaerne et al., 2023). مدل‌ها با فراپارامترهای مختلف آموزش داده شدند که در نهایت مقادیر زیر به آن‌ها اختصاص داده شد.

جدول ۲: فراپارامترهای استفاده شده در مرحله آموزش مدل

مقدار	فراپارمتر	مقدار	فراپارمتر
۰/۲	translate (Image augmentation - translation)	۰/۰۰۱	lr0 (Initial learning rate)
۰/۹	scale (Image augmentation - scale)	۰/۹۳۷	Momentum (Momentum for the optimizer)
.	shear (Image augmentation - shear)	۰/۲	iou_t (IOU threshold)
.	perspective (Image augmentation - perspective)	۴/۰	anchor_t (Anchor threshold)
.	flipud (Image augmentation - flip upside-down)	۰/۰۱۵	hsv_h (HSV augmentation - hue)
۱	fliplr (Image augmentation - flip left-right)	۰/۰۷۷	hsv_s (HSV augmentation - saturation)
.	mosaic (Mosaic augmentation)	۰/۴	hsv_v (HSV augmentation - value)
۰/۱۵	mixup (Mixup augmentation)	۰/۱۵	degrees (Image augmentation - rotation)

معیارهای ارزیابی

الگوریتم‌ها در این مقاله با استفاده از رابطه‌های (۱) - (۶) ارزیابی شده‌اند. معیار دقت (P) تعداد تشخیص‌های صحیح، معیار یادآوری (R) تعداد غذاهایی است که به اشتباه تشخیص داده نشده‌اند و معیار F1 متوسط، تعادل بین دقت و یادآوری را نشان می‌دهد. AP ناحیه زیر منحنی دقت-یادآوری است که نشان‌دهنده تعداد نقاطی است که دقت و یادآوری برابر با هم هستند. معیار mAP میانگین مقادیر AP برای هر دسته از غذاها را نشان می‌دهد. از این معیارها برای ارزیابی عملکرد الگوریتم‌ها استفاده شده است. سرعت، تعداد فریم بر ثانیه (fps) معیار دیگری است که برای ارزیابی عملکرد الگوریتم در این مقاله استفاده شده است. TP مثبت واقعی برای تشخیص صحیح، FN منفی کاذب برای تشخیص‌های اشتباه، FP مثبت کاذب برای تشخیص‌های نادرست و n نشان دهنده تعداد کلاس‌ها است. با استفاده از این مقادیر معیارهای ارزیابی محاسبه می‌شوند (Zhao et al., 2023):

$$P = \frac{TP}{TP + FP} \quad \text{رابطه ۱}$$

$$R = \frac{TP}{TP + FN} \quad \text{رابطه ۲}$$

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad \text{رابطه ۳}$$

$$AP = \int_0^1 P(R) dR \quad \text{رابطه ۴}$$

$$mAP = \frac{\sum AP}{n} \quad \text{رابطه ۵}$$

$$speed = \frac{frames}{time} \quad \text{رابطه ۶}$$

لازم به ذکر است که در این تحقیق، روش ارزیابی الگوریتم‌ها با روش‌های دیگر به دقت مورد مقایسه قرار گرفته است. همچنین، تمامی مراحل ارزیابی با استفاده از یک مجموعه داده خاص و در شرایط یکسان اجرا شده است.

نتایج و بحث

عملکرد مدل تشخیص

در این پژوهش، از روش تشخیص تصویر به نام YOLOv7 برای تشخیص ۲۲ نوع ماده غذایی در تصاویر استفاده شد. این روش با استفاده از شبکه‌های عصبی کانولوشنال عمیق، قابلیت یادگیری نمایش سلسله‌مراتبی تصاویر را داراست. جهت آموزش مدل، بخشی از مجموعه داده غذا در ۱۰۰ دوره آموزش داده شد. نتایج نشان داد که عملکرد سیستم پس از ۳۰ دوره آموزش تقریباً ثابت بوده و تغییر چندانی نکرده است؛ بنابراین، برای آموزش داده‌ها، ۴۰ دوره در نظر گرفته شد.

با توجه به شکل ۶، مشاهده می‌شود که YOLOv7 عملکرد مطلوبی داشته و مقدار دقت متوسط میانگین، ۷۷ درصد را برای تشخیص ۲۲ ماده غذایی کسب کرده است. همچنین، دقت متوسط میانگین و یادآوری سیستم به ترتیب ۷۵ و ۶۶/۹ درصد بوده است.

نمودار اول از سمت چپ در شکل ۶، نمودار جعبه مربوط به معیار خطای جعبه محدود کننده است. این نمودار نشان می‌دهد که چگونه خطای بین جعبه‌های پیش‌بینی شده و جعبه‌های حقیقت زمین را محاسبه می‌کند. هدف از این تابع خطا کاهش اختلاف بین جعبه‌های پیش‌بینی شده و حقیقت زمین است. هر چه خطا کمتر باشد، نشان می‌دهد که شبکه بهتر یاد گرفته و توانایی پیش‌بینی جعبه‌های دقیق‌تر را دارد.

این نمودار رفتار معیار خطای جعبه‌ای را در طول مراحل آموزش نشان می‌دهد. مقادیر خطا روی محور عمودی و شماره‌های دوره‌ها روی محور افقی قرار دارند. در وظایف تشخیص اشیاء، که شامل محل‌یابی و دسته‌بندی است، روش اصلی برای محل‌یابی چندین شیء در یک تصویر استفاده از جعبه‌های محدودکننده است. با توجه به نمودار خطای جعبه شکل ۶ با افزایش تعداد دوره‌ها، روند کاهش مشاهده می‌شود که نشان‌دهنده کاهش خطای بین جعبه‌های پیش‌بینی است.

نمودار دوم در شکل ۶ نشان می‌دهد که چگونه مدل یاد می‌گیرد که اطمینان از وجود یک شیء در یک جعبه محدود کننده را ارزیابی کند. Objectness به اطمینان مدل از حضور یک شیء در یک جعبه محدود کننده اشاره دارد. از طرف دیگر، امتیاز کلاس نشان می‌دهد که احتمال شرطی یک کلاس خاص، با فرض وجود یک شیء در آن جعبه چقدر است. امتیاز اطمینان کلی برای هر کلاس با ضرب اطمینان شیء و امتیاز کلاس به دست می‌آید. در این سناریو، مطلوب است که اطمینان شیء با افزایش تعداد دوره‌ها به سمت صفر کاهش یابد. این نشان می‌دهد که مدل در تشخیص دقیق حضور اشیاء پیشرفت می‌کند. با توجه به شکل ۶ مشاهده می‌شود اطمینان شیء با افزایش تعداد دوره‌ها به سمت صفر کاهش یافته است.

پلات سوم در شکل ۶ نشان می‌دهد که چگونه دقت مدل در تشخیص مواد غذایی از طریق شبکه عصبی عمیق با افزایش تعداد دوره‌ها تغییر می‌کند. دقت یک معیار است که نسبت پیش‌بینی‌های مثبتی را که به درستی طبقه‌بندی شده‌اند، اندازه‌گیری می‌کند. دقت به وجود پیش‌بینی‌های مثبت اشتباه، که به مواردی اشاره می‌کند که به اشتباه به عنوان مثبت شناسایی شده‌اند، حساس است. هدف از دقت این است که مدل بتواند تشخیص‌های مثبت را با دقت بالا انجام دهد و تعداد خطاها را کم کند. هر چه دقت بیشتر باشد، به این معنی است که مدل در تشخیص مواد غذایی کمتر اشتباه می‌کند. مقدار دقت بین ۰ تا ۱ متغیر است، که ۱ به معنی دقت کامل و ۰ به معنی دقت صفر است.

نمودار نشان می‌دهد که دقت مدل با افزایش تعداد دوره‌ها افزایش می‌یابد. این به این معنی است که مدل با یادگیری بیشتر، قادر به تشخیص مواد غذایی با دقت بالاتر می‌شود. مقدار دقت در دوره آخر ۷۷ است که نشان می‌دهد که مدل در ۷۷ درصد موارد، مواد غذایی را به درستی تشخیص داده است.

نمودار یادآوری در شکل ۶ مقدار یادآوری را برای هر مرحله نشان می‌دهد. یادآوری یک معیار است که نسبت تعداد پیش‌بینی‌های مثبتی را که به درستی طبقه‌بندی شده‌اند، به تعداد کل موارد مثبت واقعی اندازه‌گیری می‌کند. این معیار در نظر می‌گیرد که چند مورد منفی اشتباهی وجود دارد که باید به عنوان مثبت شناسایی می‌شدند اما به اشتباه به عنوان منفی طبقه‌بندی شده‌اند. به عبارت دیگر، یادآوری نشان می‌دهد که مدل چقدر قادر است تمام موارد مثبت را در تصاویر شناسایی کند. هر چه مقدار یادآوری بالاتر باشد، نشان می‌دهد که مدل کمتر خطا می‌کند و دقیق‌تر عمل می‌کند.

اگر مقدار یادآوری پایین باشد، به این معنی است که مدل برخی از مناطق مواد غذایی را از قلم انداخته و این می‌تواند منجر به تشخیص نادرست می‌شود. بنابراین، هدف این است که مقدار یادآوری را بالا ببریم تا مدل بتواند تمام مناطق را با دقت بالا شناسایی کند.



در نمودار یادآوری در شکل ۶ می‌توانیم روند افزایش یادآوری را با افزایش تعداد مراحل مشاهده کنیم. این نشان می‌دهد که مدل به مرور زمان یاد می‌گیرد و قابلیت شناسایی مناطق را بهبود می‌بخشد.

نمودارهای دقت متوسط میانگین معیارهایی هستند که برای ارزیابی دقت مدل در تشخیص و مکان‌یابی مواد غذایی استفاده می‌شوند. برای درک بهتر این معیارها، ابتدا باید مفاهیمی مانند IOU را تعریف شوند.

IOU (اشتراک بر اجتماع) یک معیار است که میزان همپوشانی بین دو جعبه محدود کننده را اندازه‌گیری می‌کند. این معیار به صورت نسبت مساحت اشتراک به مساحت اجتماع دو جعبه محدود کننده محاسبه می‌شود. مقدار IOU بین ۰ تا ۱ متغیر است، که ۱ به معنای همپوشانی کامل و ۰ به معنای عدم همپوشانی است. دو جعبه محدود کننده با هم مقایسه می‌شوند و یک معیار استاندارد از همپوشانی آن‌ها ارائه می‌شود که در ارزیابی دقت پیش‌بینی جعبه محدود کننده در وظایف تشخیص شیء مفید است. در زمان ارزیابی مدل‌های تشخیص شیء، ضروری است که یک معیار برای تعیین موفقیت تشخیص بر اساس همپوشانی جعبه محدود کننده با داده‌های واقعی تعریف شود. برای این منظور از IOU استفاده می‌شود.

معیار دقت متوسط میانگین ۰/۵ دقت را زمانی که آستانه IOU برابر ۰/۵ است محاسبه می‌کند. به عبارت دیگر، اگر همپوشانی بین جعبه محدود کننده پیش‌بینی شده و جعبه محدود کننده واقعی بیش از ۵۰ درصد باشد، تشخیص موفق در نظر گرفته می‌شود. هر چه آستانه IOU بیش‌تر شود، دقت لازم برای تشخیص جعبه محدود کننده سخت‌تر می‌شود و باعث کاهش مقدار دقت متوسط میانگین می‌شود. این نشان می‌دهد که تشخیص دقیق جعبه محدود کننده با یک آستانه همپوشانی بالاتر دشوارتر است.

دقت متوسط میانگین به عنوان میانگینی از دقت میانگین (AP) محاسبه می‌شود. هر مقدار AP نشان‌دهنده تعادل دقت و یادآوری برای یک کلاس خاص در وظیفه تشخیص شیء است. در نمودارهای دقت متوسط میانگین در شکل ۶ روند افزایشی دقت متوسط میانگین با افزایش تعداد دوره مشاهده می‌شود.

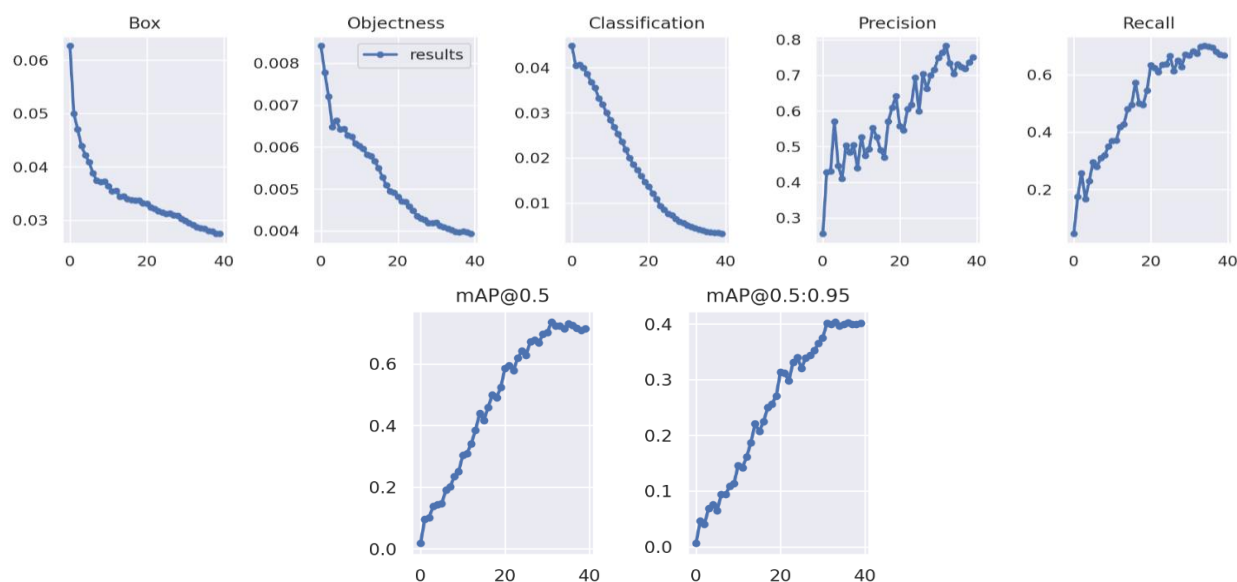
شکل ۷ نیز نشان‌دهنده ماتریس اغتشاش برای YOLOv7 است. ماتریس اغتشاش یک ابزار ارزیابی است که در آن نتایج مدل برچسب‌های پیش‌بینی شده و برچسب‌های واقعی داده‌ها با هم مقایسه می‌شوند. این ماتریس برای بررسی صحت و دقت الگوریتم‌ها استفاده می‌شود. در این ماتریس، اعداد در ردیف‌ها برچسب‌های پیش‌بینی شده توسط مدل را نشان می‌دهند و اعداد در ستون‌ها برچسب‌های واقعی داده‌ها را نمایش می‌دهند. همچنین، در شکل ۸ نمونه‌هایی از نتایج تشخیص بر روی تصاویر مجموعه داده آزمایش به همراه کادرهای مرزی پیش‌بینی شده توسط مدل قابل مشاهده هستند.

با این حال، YOLOv7 در شناسایی و تشخیص کلاس‌های مختلف عملکرد نامتوازی داشته است. بررسی جدول ۳ نشان‌دهنده کمترین دقت برای کلاس کباب و سبزی‌ها (۴۹ و ۵۰ درصد) و بالاترین دقت برای کلاس هویج‌پلو و کوسبزی (۱۰۰ درصد) است. برای کلاس‌های عدس‌پلو، خیار، دوغ، سیب‌زمینی سرخ‌کرده، قیمة، قورمه سبزی، سبزی، هویج‌پلو، جوجه‌کباب، کباب، کچاپ، کتلت، کوکو، لواش، لوبیاپلو، زیتون، خیارشور، پیترزا، برنج، سالاد، گوجه‌فرنگی و نوشابه دقت به ترتیب برابر با ۹۳، ۸۲، ۵۳، ۵۷، ۷۰، ۶۹/۵، ۵۰، ۱۰۰، ۷۱، ۴۹، ۶۲، ۷۰، ۱۰۰، ۸۴، ۵۹، ۷۹، ۸۹، ۹۲، ۸۰، ۵۶، ۸۰ و ۷۸ درصد به دست آمد. دقت پایین در تشخیص برای برخی از اقلام غذایی، به ویژه در کلاس‌های کباب و سبزی می‌تواند به عوامل متعددی نسبت داده شود. یکی از دلایل اصلی پیچیدگی اقلام غذایی است، به ویژه آن‌هایی که شکل، رنگ یا بافت مشابهی دارند تشخیص این اقلام برای مدل چالش برانگیز است. دلیل دیگر، ویژگی‌های خاص حوزه مواد غذایی است، که اغلب دارای ویژگی‌های بصری منحصر به فردی هستند که ممکن است توسط مدل‌های تشخیص شیء به خوبی ثبت نشود.

همچنین یادآوری برای این کلاس‌ها به ترتیب برابر با ۹۳، ۸۰، ۵۲، ۴۸، ۶۰، ۷۵، ۹۰، ۵۴، ۴۵، ۵۴، ۹۳، ۷۷، ۸۶، ۸۷، ۹۴، ۵۵، ۲۹، ۸۹، ۶۵، ۴۲، ۸۸ و ۷۸ درصد بوده و دقت متوسط میانگین نیز برابر با ۶۴ درصد به دست آمد که بالاترین مقدار برای کلاس‌های لوبیاپلو (۹۴ درصد) و عدس‌پلو (۹۳ درصد) است. دقت متوسط میانگین پایین برخی از کلاس‌ها را می‌توان به چالش تشخیص دقیق جعبه‌های محدود کننده با آستانه همپوشانی بالاتر نسبت داد که منجر به کاهش میانگین دقت می‌شود. این به دلیل افزایش دشواری در دستیابی به دقت مورد نیاز برای تشخیص جعبه‌های محدود کننده با افزایش آستانه IOU است. IOU همپوشانی بین دو جعبه محدود کننده پیش‌بینی شده و حقیقت زمین را اندازه‌گیری می‌کند و یک آستانه بالا مستلزم مکان‌یابی دقیق‌تر است که باید جعبه‌های محدود کننده را با دقت بیشتری شناسایی کند که بر دقت متوسط مدل تأثیر می‌گذارد.

نتایج نشان می‌دهد که مدل‌های یادگیری عمیق مانند YOLOv7 به دلیل معماری مؤثر، توانایی پردازش دقیق و سریع تصاویر را

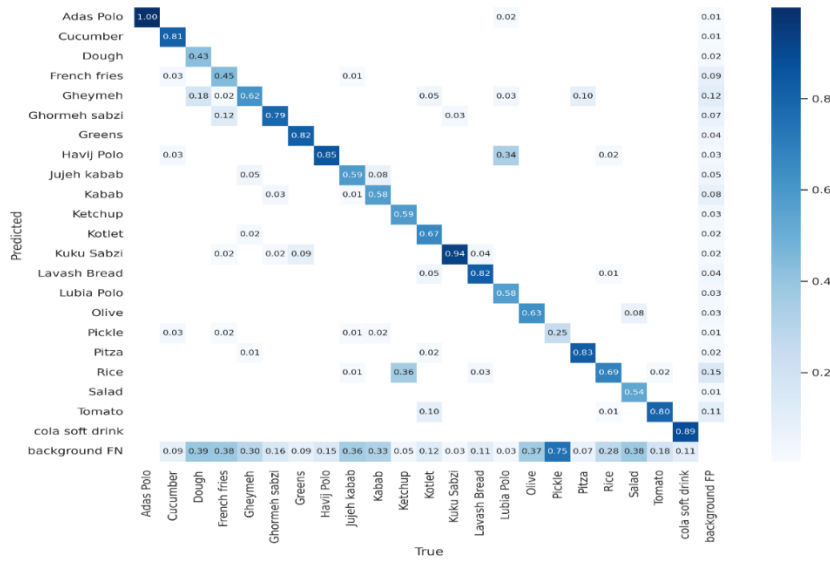
دارد و برای برنامه‌های بلادرنگ ایده‌آل هستند. با این حال، نامتوازی در دقت تشخیص کلاس‌های مختلف می‌تواند به عوامل مختلفی بستگی داشته باشد. از جمله این عوامل می‌توان به اختلاف در وضوح و کیفیت تصاویر و وجود چالش‌های ناشی از پس‌زمینه و اشیاء نامربوط اشاره کرد. همچنین، عوامل دیگری نیز می‌توانند بر عملکرد مدل تأثیر بگذارند؛ از جمله تنظیم فرآیندها مانند اندازه دسته، اندازه تصویر، بهینه‌سازهای مختلف و نرخ یادگیری.



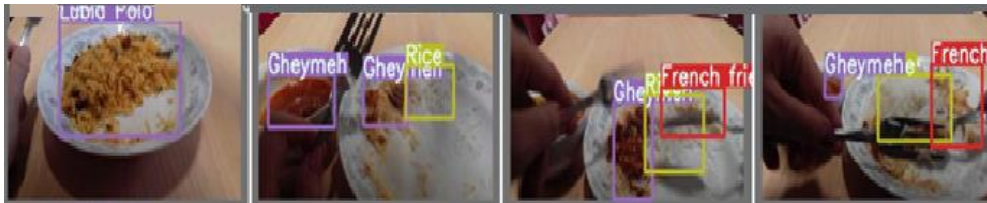
شکل ۶: عملکرد مدل YOLOv7 پیشنهادی، آموزش داده شده در ۴۰ دوره بر روی مجموعه داده غذا

جدول ۳: نتایج مدل YOLOv7 بر روی مجموعه داده غذا بر حسب درصد

کلاس	تصاویر	برچسب‌ها	P	R	mAP@0.5	mAP@0.5-0.95
همه کلاس‌ها	۴۳۹	۱۱۱۷	۷۳	۶۹/۷	۶۴	۳۷
عدس پلو	۴۳۹	۱۵	۹۳	۹۳	۹۴	۷۱
خیار	۴۳۹	۲۹	۸۲	۸۰	۸۰	۴۵
دوغ	۴۳۹	۲۳	۵۳	۵۲	۴۴	۲۵
سیب زمینی سرخ کرده	۴۳۹	۵۶	۵۷	۴۸	۳۴	۱۶
قیمه	۴۳۹	۸۸	۷۰	۶۶	۶۰	۳۱
قورمه سبزی	۴۳۹	۶۴	۶۹/۵	۷۵	۷۱	۳۳
سبزی	۴۳۹	۱۰	۵۰	۹۰	۸۳	۵۸
هویج پلو	۴۳۹	۶۶	۱۰۰	۵۴	۵۵	۳۲
جوجه کباب	۴۳۹	۸۷	۷۱	۴۵	۳۷/۵	۱۶
کباب	۴۳۹	۵۰	۴۹	۵۴	۴۵	۲۰
سس کچاپ	۴۳۹	۱۴	۶۲	۹۳	۸۳	۵۰
کنلت	۴۳۹	۳۵	۷۰	۷۷	۶۳	۳۵
کوکوسبزی	۴۳۹	۳۷	۱۰۰	۸۶	۸۷	۳۸
نان لواش	۴۳۹	۷۷	۸۴	۸۷	۸۲	۵۱
لوبیا پلو	۴۳۹	۳۶	۵۹	۹۴	۹۰	۶۴
زیتون	۴۳۹	۲۰	۷۹	۵۵	۵۰	۱۹
خیارشور	۴۳۹	۲۸	۸۹	۲۹	۲۸	۱۸
پیتزا	۴۳۹	۳۹	۹۲	۸۹	۹۳	۶۴
برنج	۴۳۹	۱۹۴	۸۰	۶۵	۶۰	۳۰
سالاد	۴۳۹	۱۲	۵۶	۴۲	۲۷	۱۵
گوچه فرنگی	۴۳۹	۱۲۸	۸۰	۸۸	۷۰/۵	۳۶
نوشابه	۴۳۹	۹	۷۸	۷۸	۷۴	۴۷



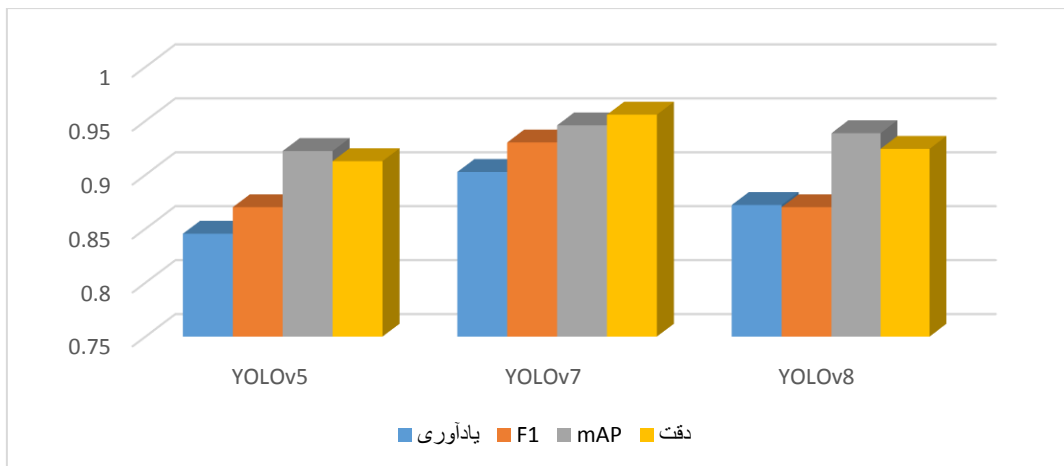
شکل ۷: ماتریس اغتشاش برای YOLOv7



شکل ۸: نمونه‌ای از تصاویر شناسایی شده توسط مدل YOLOv7 بر روی مجموعه داده آزمایش

نتایج مدل‌های بخش‌بندی

مدل‌های بخش‌بندی نمونه‌ای YOLOv5، YOLOv7 و YOLOv8، موفق به تشخیص و بخش‌بندی غذاهای ایرانی شدند. با توجه به شکل ۹ نتایج آزمایش‌ها نشان داد که YOLOv7 با بالاترین دقت ۹۵/۵ درصد، دارای بهترین عملکرد در بخش دقت است. به دنبال آن، YOLOv8 با دقت ۹۲/۴ درصد و سپس YOLOv5 با کمترین دقت ۹۱/۲ درصد قرار داشتند. همچنین، در سایر بخش‌ها نیز در صدر قرار گرفت و دارای دقت متوسط میانگین ۹۴/۵، امتیاز F1 برابر با ۹۳ و یادآوری ۹۰ درصد بود. YOLOv8 نیز با یادآوری نزدیک ۷۰، امتیاز F1 برابر با ۸۷ و دقت متوسط میانگین ۹۳ درصد در جایگاه دوم قرار داشت و در نهایت، YOLOv5 با یادآوری ۸۴/۵، امتیاز F1 برابر با ۸۷ و کمترین دقت متوسط میانگین ۹۲/۲ درصد در جایگاه سوم قرار گرفت.



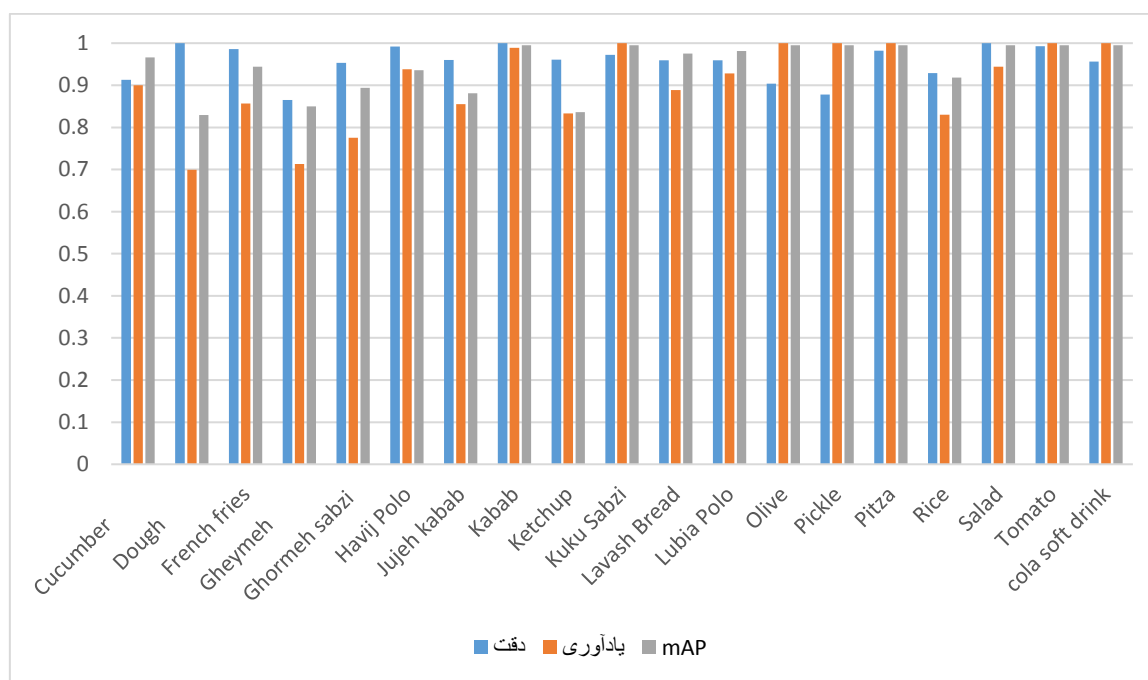
شکل ۹: مقایسه عملکرد مدل‌های بخش‌بندی نمونه‌ای YOLOv5، YOLOv7 و YOLOv8

بر اساس نتایج به دست آمده، مشخص شد که YOLOv7 در ۹۵ درصد مواقع موفق به تشخیص دسته‌های غذایی مختلف بوده است. به طور مشابه، YOLOv8 و YOLOv5 نیز با موفقیت به ترتیب ۹۲ و ۹۱ درصد دسته‌های غذایی را تشخیص داده‌اند. همچنین، با توجه به یادآوری، YOLOv7 در ۹۳ درصد موارد به درستی یک دسته خاص از غذاها را تشخیص داد. YOLOv8 هر دسته‌ای را در ۸۷ درصد موارد تشخیص داد و YOLOv5 نیز میزان موفقیت ۸۷ درصد را به دست آورد. همچنین، در نتایج مربوط به بخش دقت متوسط میانگین نیز YOLOv7 با دقت متوسط میانگین ۹۴/۵ درصد به عنوان بهترین مدل شناخته شد. YOLOv5 با دقت متوسط میانگین ۹۲ درصد کمترین دقت را داشت و YOLOv8 با دقت متوسط ۹۳/۸ درصد در میان دو مدل دیگر قرار گرفت.

شکل ۱۰ عملکرد مدل بخش‌بندی YOLOv7 برای غذاهای ایرانی را به استناد به دقت‌های مختلف نمایش می‌دهد. در این نمودار، مدل بخش‌بندی YOLOv7 برای تشخیص خیار، دوغ، سیب‌زمینی سرخ‌کرده، قیمة، قورمه سبزی، هویج‌پلو، جوجه‌کباب، کباب، سس گوجه‌فرنگی، نان لواش، لوبیاپلو، زیتون، خیارشور، برنج، سالاد، گوجه و نوشابه استفاده شده است. درصد دقت مدل برای هر یک از این موارد به ترتیب به شرح زیر است: ۹۱/۳ برای خیار، ۱۰۰ برای دوغ، ۹۸/۶ برای سیب‌زمینی سرخ‌کرده، ۸۶/۵ برای قیمة، ۹۵/۳ برای قورمه سبزی، ۹۹/۲ برای هویج‌پلو، ۹۶ برای جوجه‌کباب، ۱۰۰ برای کباب، ۹۶/۱ برای سس گوجه‌فرنگی، ۹۷/۲ برای کباب، ۹۵/۹ برای نان لواش، ۹۵/۹ برای لوبیاپلو، ۹۰/۴ برای زیتون، ۸۷/۸ برای خیارشور، ۹۸/۲ برای برنج، ۱۰۰ برای سالاد، ۹۹/۳ برای گوجه و ۹۵/۶ برای نوشابه. این نتایج نشان می‌دهد که مدل توانسته است حداقل ۸۶/۵ و حداکثر ۱۰۰ درصد از موارد را به درستی تشخیص دهد. کمترین دقت به ترتیب مربوط به قیمة (۸۶/۵ درصد) و خیارشور (۸۷/۸ درصد) بود. از سوی دیگر، برای دوغ، کباب و سالاد بالاترین مقدار ۱۰۰ درصد برای دقت به دست آمده است.

برای بهبود دقت تشخیص یک کلاس خاص، می‌توان از روش‌های مختلفی مانند افزایش حجم داده‌های آموزشی، تنظیم دقیق مدل و همچنین از تکنیک‌های پس‌پردازش استفاده کرد. برای مثال، افزودن تصاویر بیشتری به مجموعه داده، تغییر وزن‌های مدل برای تمرکز بیشتر روی کلاس خاص، و استفاده از عملیات مورفولوژیکی و هموارسازی برای حذف نویز و اصلاح پیش‌بینی‌های مدل می‌تواند بهبود دقت را به همراه داشته باشد. یک نکته مهم برای ارتقاء عملکرد مدل، تعادل بین دقت و یادآوری در هر یک از کلاس‌ها است؛ بنابراین، ارزیابی عملکرد مدل بر روی همه کلاس‌ها از اهمیت ویژه‌ای برخوردار است.

همچنین، زمان پیش‌بینی مدل برای هر تصویر آزمایشی حدود ۶۲/۳ میلی‌ثانیه به دست آمد که نشان‌دهنده سرعت خوب مدل در پیش‌بینی‌ها است.



شکل ۱۰: مقایسه عملکرد کلاسی تقسیم‌بندی نمونه‌های YOLOv7

در نهایت، نتایج تشخیص بصری و بخش‌بندی مدل بر روی مجموعه داده آزمایشی در شکل ۱۱ نشان می‌دهد که مدل YOLOv7

موفق به تشخیص و بخش‌بندی ماسک‌های مربوط به مواد غذایی در تصاویر با شرایط مختلف از جمله ظروف متفاوت، روشنایی متغیر و همپوشانی‌های مختلف شده است. همچنین، مدل با دقت بالایی مواد غذایی با اندازه‌ها و اشکال مختلف را شناسایی کرده و امتیاز اطمینان بالایی را به هر کلاس اختصاص داده است. این نتایج نشان‌دهنده توانایی برتر مدل در تشخیص مواد غذایی مختلف با دقت و قابلیت اعتماد بالا است.

به‌طور کلی، نتایج این تحقیق نشان‌دهنده کارایی مدل YOLOv7 در میان مدل‌های مورد بررسی است.



شکل ۱۱: نتایج تشخیص بصری و بخش‌بندی مدل بر روی مجموعه داده آزمایش

در پژوهشی یک برنامه موبایل برای تشخیص غذاها و نمایش اطلاعات تغذیه‌ای آن‌ها با استفاده از شبکه‌های عصبی عمیق و روش‌های تشخیص پیشرفته معرفی شد. آن‌ها از دو مجموعه داده UECFood100 و UECFood256 برای آموزش استفاده کرده‌اند و دقت تشخیص غذا با روش YOLOv2 به ترتیب ۷۶/۳۶ و ۷۵/۰۵ درصد به دست آمد. مدل به یک برنامه موبایل تبدیل شده است که با API Nutritionix اطلاعات تغذیه‌ای هر غذا را نشان می‌دهد (Sun et al., 2019). در پژوهشی دیگر یک مدل برای تشخیص ۲۵ نوع غذای فیلپینی با استفاده از الگوریتم YOLO معرفی شد. پژوهشگران از مجموعه داده‌هایی که با استفاده از Pyimagesearch جمع‌آوری شده بود، برای آموزش مدل استفاده کردند. مدل با دقت ۷۷ درصد توانست غذاها را تشخیص دهد (Yumang et al., 2021). مطابق مشاهدات و نتایج به دست آمده بر روی مجموعه داده تهیه شده از غذاهای ایرانی، دقت تشخیص و بخش‌بندی الگوریتم‌های YOLO استفاده شده در این تحقیق، به ترتیب ۷۷ و ۹۵ درصد به دست آمد که نشان‌دهنده عملکرد موفقیت‌آمیز این روش است. همچنین، در مقایسه با تحقیقات دیگر زمان پیش‌بینی مدل برای هر تصویر نیز بررسی شد و مدل آموزش دیده توانست با سرعت حدود ۶۲/۳ میلی‌ثانیه پیش‌بینی را انجام دهد. نکته حائز اهمیت پژوهش‌های دیگر فرآیند آموزش بهتری و دقت بالاتر، به دلیل دسترسی به مجموعه داده‌های برچسب‌گذاری شده بزرگ برای غذاهای خارجی است. به همین دلیل، برای دستیابی به دقت بالاتر، ضروری است از مجموعه داده‌های بزرگتری استفاده کنیم که تعداد تصاویر در هر کلاس متعادل باشد. در مقایسه با تحقیقات دیگر، تحقیق فعلی نیز نتایج قابل توجهی به ارمغان آورده است و به‌عنوان یک روش قوی و کارا در شناسایی و بخش‌بندی مواد غذایی ایرانی ارزیابی می‌شود.

نتیجه‌گیری

در این تحقیق، به بررسی روش‌های تشخیص و بخش‌بندی خودکار مواد غذایی ایرانی با استفاده از تصاویر رنگی پرداخته شد. برای شناسایی و مکان‌یابی مواد غذایی، از آشکارساز شی YOLOv7 مبتنی بر یادگیری عمیق استفاده شد. سپس با بهره‌گیری از روش‌های بخش‌بندی نمونه‌های YOLOv7، YOLOv5 و YOLOv8 ماسک‌های پیکسلی هر نمونه از ماده غذایی شناسایی شده و استخراج گردید. این روش‌ها امکان تفکیک نمونه‌های متعلق به یک کلاس واحد در یک تصویر را فراهم می‌کنند و بنابراین هر ماده غذایی قابل تشخیص و جداسازی است. نتایج حاصل از تحقیق نشان می‌دهد که بین سه روش بخش‌بندی، YOLOv7 دارای دقت بالاتر و پایدارتر در استخراج ماسک‌های پیکسلی است. همچنین، YOLOv7 عملکرد خوبی در تشخیص مواد غذایی روی مجموعه داده آماده شده نشان داده است. به‌طور کلی، روش YOLOv7 به‌عنوان یک راه‌حل مناسب و سریع برای تشخیص مواد غذایی ایرانی پیشنهاد می‌شود و دقت تشخیص و بخش‌بندی با استفاده از این روش به ترتیب ۷۷ و ۹۵ درصد به دست آمد.

برای بهبود عملکرد مدل‌ها، آماده‌سازی مجموعه داده‌های بزرگ‌تری از تصاویر مواد غذایی ایرانی و آموزش مدل‌ها برای تشخیص و بخش‌بندی کلاس‌های اضافی توصیه می‌شود.

در ادامه به برخی از محدودیت‌های الگوریتم YOLO می‌پردازیم. الگوریتم YOLO برای دستیابی به دقت بالا نیاز به شبکه‌های عمیق و پهن دارد که منجر به افزایش تعداد پارامترها و محاسبات می‌شود. این الگوریتم برای تشخیص اشیاء از شبکه‌های محلی‌سازی

استفاده می‌کند که هر کدام یک جعبه محدود کننده را پیش‌بینی می‌کنند. این روش ممکن است برای اشیاء با اندازه‌ها و نسبت‌های مختلف کارایی کمتری داشته باشد.

الگوریتم YOLO برای بهبود عملکرد خود نیاز به استفاده از روش‌های مختلفی دارد که ممکن است پیچیدگی و هزینه آموزش را افزایش دهند.

منابع

طاهری گراوند، امین؛ نصیری، امین و بنان، اشکان. (۱۴۰۰). توسعه الگوریتم یادگیری عمیق به منظور تشخیص و طبقه بندی هوشمند گونه های ماهی کپور، نشریه مهندسی بیوسیستم/ایران، ۵۲ (۳)، ۳۹۱-۴۰۷.

REFERENCES

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), 2274–2281. <https://doi.org/10.1109/TPAMI.2012.120>
- Afzaal, U., Bhattarai, B., Pandeya, Y. R., & Lee, J. (2021). An Instance Segmentation Model for Strawberry Diseases Based on Mask R-CNN. *Sensors* 2021, Page 6565, 21(19), 6565. <https://doi.org/10.3390/S21196565>
- Ando, Y., Ege, T., Cho, J., & Yanai, K. (2019). DepthCalorieCam: A mobile application for volume-based food calorie estimation using depth cameras. *MADiMa 2019 - Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management, Co-Located with MM 2019*, 76–81. <https://doi.org/10.1145/3347448.3357172>
- Aslan, S., Ciocca, G., Mazzini, D., & Schettini, R. (2020). Benchmarking algorithms for food localization and semantic segmentation. *International Journal of Machine Learning and Cybernetics*, 11(12), 2827–2847. <https://doi.org/10.1007/S13042-020-01153-Z/FIGURES/13>
- Ciocca, G., Napoletano, P., & Schettini, R. (2015). Food recognition and leftover estimation for daily diet monitoring. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9281, 334–341. https://doi.org/10.1007/978-3-319-23222-5_41/COVER
- Dehaerne, E., Dey, B., Halder, S., & De Gendt, S. (2023). Optimizing YOLOv7 for semiconductor defect detection. *In Metrology, Inspection, and Process Control XXXVII*, 12496, 635-642. <https://doi.org/10.1117/12.2657564>
- Dehais, J., Anthimopoulos, M., & Mougiakakou, S. (2015). Dish detection and segmentation for dietary assessment on smartphones. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9281, 433–440. https://doi.org/10.1007/978-3-319-23222-5_53/COVER
- Dutta, A., & Zisserman, A. (2019). The VIA annotation software for images, audio and video. *In MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia* (pp. 2276–2279). Association for Computing Machinery, Inc. <https://doi.org/10.1145/3343031.3350535>
- G. Jocher, A. C. and J. Q. (2020). YOLO by Ultralytics. <https://github.com/ultralytics/ultralytics>, 2023. Accessed: February 30, 2023.
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation (pp. 580–587).
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-October, 2980–2988. <https://doi.org/10.1109/ICCV.2017.322>
- He, K., Zhang, X., Ren, S., & Sun, J. (2014). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8691 LNCS(PART 3), 346–361. https://doi.org/10.1007/978-3-319-10578-9_23
- Islam, A., Chowdhury, T., Hossain, M., Nahid, N., & Rifat, A. I. (2022). An Automatic System for Identifying and Categorizing Tribal Clothing Based on Convolutional Neural Networks. *4th International Conference on Emerging Research in Electronics, Computer Science and Technology, ICERECT 2022*. <https://doi.org/10.1109/ICERECT56837.2022.10060409>
- Jamnekar, R.V., Keole, R.R., Mohod, S.W., Mahore, T.R., Pande, S. (2023). Food Classification Using Deep



- Learning Algorithm. *International Conference on Innovative Computing and Communications*. Lecture Notes in Networks and Systems, vol 492. Springer, Singapore. https://doi.org/10.1007/978-981-19-3679-1_62
- Jiang, L., Qiu, B., Liu, X., Huang, C., & Lin, K. (2020). DeepFood: Food Image Analysis and Dietary Assessment via Deep Model. *IEEE Access*, 8, 47477–47489. <https://doi.org/10.1109/ACCESS.2020.2973625>
- Jiang, S., Min, W., Liu, L., & Luo, Z. (2020). Multi-Scale Multi-View Deep Feature Aggregation for Food Recognition. *IEEE Transactions on Image Processing*, 29, 265–276. <https://doi.org/10.1109/TIP.2019.2929447>
- Kaur, P., Sikka, K., Wang, W., Belongie, S., & Divakaran, A. (2019). FoodX-251: A Dataset for Fine-grained Food Classification. <https://arxiv.org/abs/1907.06167v1>
- Kaur, R., Kumar, R., & Gupta, M. (2023). Deep neural network for food image classification and nutrient identification: A systematic review. *Reviews in Endocrine and Metabolic Disorders*, 1–21. <https://doi.org/10.1007/S11154-023-09795-4/METRICS>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25.
- Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., Tang, J., & Yang, J. (2020). Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *Advances in Neural Information Processing Systems*, 2020-December. <https://arxiv.org/abs/2006.04388v1>
- Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection (pp. 2117–2125).
- Lo, F. P. W., Sun, Y., Qiu, J., & Lo, B. P. L. (2020). Point2Volume: A vision-based dietary assessment approach using view synthesis. *IEEE Transactions on Industrial Informatics*, 16(1), 577–586. <https://doi.org/10.1109/TII.2019.2942831>
- Lu, Y., Stathopoulou, T., Vasiloglou, M. F., Christodoulidis, S., Stanga, Z., & Mougiakakou, S. (2021). An Artificial Intelligence-Based System to Assess Nutrient Intake for Hospitalised Patients. *IEEE Transactions on Multimedia*, 23, 1136–1147. <https://doi.org/10.1109/TMM.2020.2993948>
- Lu, Y., Stathopoulou, T., Vasiloglou, M. F., Pinault, L. F., Kiley, C., Spanakis, E. K., & Mougiakakou, S. (2020). goFOODTM: An artificial intelligence system for dietary assessment. *Sensors (Switzerland)*, 20(15), 1–18. https://doi.org/10.3390/S20154283/SENSORS_20_04283_PDF.PDF
- M. Contributors. (2023). YOLOv8 by MMYOLO. <https://github.com/open-mmlab/mmyolo/tree/main/configs/yolov8>, 2023. Accessed: May 13, 2023.
- Mahmoodi-Eshkaftaki, M., Haghghi, A., & Houshyar, E. (2020). Land suitability evaluation using image processing based on determination of soil texture–structure and soil features. *Soil Use and Management*, 36(3), 482–493. <https://doi.org/10.1111/sum.12572>
- Mao, R., He, J., Shao, Z., Yarlagaadda, S. K., & Zhu, F. (2021). Visual Aware Hierarchy Based Food Recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12665 LNCS, 571–598. https://doi.org/10.1007/978-3-030-68821-9_47/COVER
- Mariappan, A., Bosch, M., Zhu, F., Boushey, C. J., Kerr, D. A., Ebert, D. S., & Delp, E. J. (2009). Personal dietary assessment using mobile devices. <https://doi.org/10.1117/12.813556>, 7246, 294–305. <https://doi.org/10.1117/12.813556>
- Matsuda, Y., & Yanai, K. (2012). Multiple-food recognition considering co-occurrence employing manifold ranking. *Proceedings of the 21st International Conference on Pattern Recognition*, 2017–2020. <https://ieeexplore.ieee.org/document/6460555>
- MMYOLO Contributors. (2022). YOLOv7 by MMYOLO. <https://github.com/open-mmlab/mmyolo/tree/main/configs/yolov7>, 2023. Accessed: May 13, 2023.
- Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, 100258. <https://doi.org/10.1016/j.array.2022.100258>
- Nivedhitha, P., Anurithi, P., Meenashree, S. S., & Pooja Kumari, M. (2022). FOOD NUTRITION AND CALORIES ANALYSIS USING YOLO. *2022 1st International Conference on Computational Science and Technology, ICCST 2022 - Proceedings*, 382–386. <https://doi.org/10.1109/ICCST55948.2022.10040454>
- Okamoto, K., & Yanai, K. (2016). An automatic calorie estimation system of food images on a smartphone. *MADiMa 2016 - Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary*

- Management, Co-Located with ACM Multimedia* 2016, 63–70. <https://doi.org/10.1145/2986035.2986040>
- Pallathadka, H., Jawarneh, M., Sammy, F., Garchar, V., Sanchez, D. T., & Naved, M. (2022). A Review of Using Artificial Intelligence and Machine Learning in Food and Agriculture Industry. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 2215-2218). IEEE. <https://doi.org/10.1109/ICACITE53722.2022.9823427>
- Pouladzadeh, P., Kuhad, P., Peddi, S. V. B., Yassine, A., & Shirmohammadi, S. (2016). Food calorie measurement using deep learning neural network. *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, 2016-July. <https://doi.org/10.1109/I2MTC.2016.7520547>
- Pouladzadeh, P., & Shirmohammadi, S. (2017). Mobile Multi-Food Recognition Using Deep Learning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 13(3s). <https://doi.org/10.1145/3063592>
- Qiu, J., Lo, F. P. W., & Lo, B. (2019). Assessing individual dietary intake in food sharing scenarios with a 360 camera and deep learning. *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks, BSN 2019 - Proceedings*. <https://doi.org/10.1109/BSN.2019.8771095>
- Rajayogi, J. R., Manjunath, G., & Shobha, G. (2019). Indian Food Image Classification with Transfer Learning. *CSITSS 2019 - 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution, Proceedings*. <https://doi.org/10.1109/CSITSS47250.2019.9031051>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016a). You Only Look Once: Unified, Real-Time Object Detection (pp. 779–788).
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016b). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December, 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems*, 28.
- Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- Sezer, A., & Altan, A. (2021). Detection of solder paste defects with an optimization-based deep learning model using image processing techniques. *Soldering and Surface Mount Technology*, 33(5), 291–298. <https://doi.org/10.1108/SSMT-04-2021-0013/FULL/PDF>
- Sezer Bülent Ecevit Üniversitesi, A., Altan Bülent Ecevit Üniversitesi, A., Sezer, A., & Altan, A. (2021). Optimization of deep learning model parameters in classification of solder paste defects. *Ieeexplore.Ieee.Org*. <https://doi.org/10.1109/HORA52670.2021.9461342>
- Shima, R., Yunan, H., Fukuda, O., Okumura, H., Arai, K., & Bu, N. (2018). Object classification with deep convolutional neural network using spatial information. *ICIIBMS 2017 - 2nd International Conference on Intelligent Informatics and Biomedical Sciences*, 2018-January, 135–139. <https://doi.org/10.1109/ICIIBMS.2017.8279704>
- Sun, C., Zhan, W., She, J., & Zhang, Y. (2020). Object Detection from the Video Taken by Drone via Convolutional Neural Networks. *Mathematical Problems in Engineering*, 2020. <https://doi.org/10.1155/2020/4013647>
- Sun, J., Radecka, K., & Zilic, Z. (2019). FoodTracker: A Real-time Food Detection Mobile Application by Deep Convolutional Neural Networks. <https://arxiv.org/abs/1909.05994v2>
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December, 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>
- Taheri-Garavand, A., Nasiri, A., & Banan, A. (2021). Deep Learning Algorithm Development for Intelligent Detection and Classification of Carp Species. *Biosystem Engineering of Iran*, 52(3), 391–407. [In Persian]
- Tahir, G. A., & Loo, C. K. (2021). A Comprehensive Survey of Image-Based Food Recognition and Volume Estimation Methods for Dietary Assessment. *Healthcare* 2021, Page 1676, 9(12), 1676. <https://doi.org/10.3390/HEALTHCARE9121676>
- Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks (pp. 6105–6114). PMLR. <https://proceedings.mlr.press/v97/tan19a.html>
- Tan, M., Pang, R., & Le, Q. V. (2020). EfficientDet: Scalable and Efficient Object Detection (pp. 10781–10790).
- Wang, C. Y., Mark Liao, H. Y., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). CSPNet: A new



- backbone that can enhance learning capability of CNN. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2020-June, 1571–1580. <https://doi.org/10.1109/CVPRW50498.2020.00203>
- Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. <http://arxiv.org/abs/2207.02696>
- Wang, C.-Y., Liao, H.-Y. M., & Yeh, I.-H. (2022). Designing Network Design Strategies Through Gradient Path Analysis. <https://arxiv.org/abs/2211.04800v1>
- Weng, W., & Zhu, X. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *IEEE Access*, 9, 16591–16603. <https://doi.org/10.1109/ACCESS.2021.3053408>
- Xu, R., Lin, H., Lu, K., Cao, L., & Liu, Y. (2021). A forest fire detection system based on ensemble learning. *Forests*, 12(2), 1–17. <https://doi.org/10.3390/F12020217>
- Yumang, A. N., Banguilan, D. E. S., & Veneracion, C. K. S. (2021). Raspberry PI based Food Recognition for Visually Impaired using YOLO Algorithm. *2021 5th International Conference on Communication and Information Systems, ICCIS 2021*, 165–169. <https://doi.org/10.1109/ICCIS53528.2021.9645981>
- Zhao, H., Xu, D., Lawal, O. M., Lu, X., Ren, R., Wang, X., & Zhang, S. (2023). Jujube Fruit Instance Segmentation Based on Yolov8 Method. <https://doi.org/10.2139/SSRN.4482151>
- Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2019). Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, 12993–13000. <https://doi.org/10.1609/aaai.v34i07.6999>
- Zhou, X., Wang, D., & Krähenbühl, P. (2019). Objects as Points. <https://arxiv.org/abs/1904.07850v2>
- Zhu, F., Mariappan, A., Boushey, C. J., Kerr, D., Lutes, K. D., Ebert, D. S., & Delp, E. J. (2008). Technology-assisted dietary assessment. <https://doi.org/10.1117/12.778616>, 6814, 276–285. <https://doi.org/10.1117/12.778616>

Detection of Iranian Foods in Images Using Deep Learning

EXTENDED ABSTRACT

Introduction

Recent attention has been drawn to the application of deep learning models in various domains, with a particular focus on nutritional analysis and food quality evaluation. This study explores the use of YOLO-based models, including YOLOv5, YOLOv7, and YOLOv8, for the automatic detection and segmentation of Iranian cuisine.

Objective:

The primary aim of this study is to assess the effectiveness of several YOLO-based algorithms in detecting specific food classes commonly found in Iranian meals. YOLOv7 is employed for the detection of 22 food classes, while instance segmentation for 19 different food classes is conducted using YOLOv5, YOLOv7, and YOLOv8.

Method:

A meticulously curated dataset of Iranian food images serves as the training data for the models. To ensure the models' robustness and generalization, the dataset comprises images captured under various lighting conditions and from different viewpoints. Transfer learning strategies and hyperparameter optimization techniques are employed to enhance model precision and effectiveness.

Findings:

The YOLOv7 image detection method was employed to detect 22 types of Iranian food items. YOLOv7 utilized deep convolutional neural networks for hierarchical feature learning from images. After training the model with 100 epochs, its performance remained stable, justifying the choice of 40 epochs for training. YOLOv7 achieved satisfactory results, with average precision values of 77% for food detection. The model demonstrated good performance with a mean average precision and recall of 75.0% and 66.9%, respectively. However, YOLOv7 exhibited imbalanced accuracy in classifying different food classes, ranging from 50% for "Greens" and "Ketchup" to 100% for "Havij Polo" and "Kuku Sabzi." Improving accuracy in specific classes can be achieved by augmenting the training dataset and fine-tuning the model. Other factors, like hyperparameter adjustments, can also influence performance.

Evaluation of classification models, i.e., YOLOv5, YOLOv7, and YOLOv8 indicated that YOLOv7 outperformed the others with an accuracy value of 0.955. YOLOv7 showed a mean average precision value of 94.5%, making it the best model. Fine-tuning and post-processing techniques could further improve the accuracy in specific classes. In conclusion, YOLOv7 proved to be a strong and efficient method for detecting and classifying Iranian food items.

Conclusion:

The research highlights the significance of accurate food detection and segmentation in Iranian cuisines, enabling applications in food quality assessment, health monitoring, and dietary analysis. Furthermore, the study emphasizes the impact of different YOLO-based models on performance metrics and their potential to enhance computer vision applications.