

طراحی و پیاده‌سازی یک چارچوب انسان‌درحلقه مبتنی بر بینایی ماشین برای تشخیص محصولات و ارزیابی عملیات چیدمان در صنعت پخش: مطالعه موردی پلتفرم B2Shelf

Design and Implementation of a Human-in-the-Loop Computer Vision Framework for Product Detection and Merchandising Evaluation in the Distribution Industry: A Case Study of the B2Shelf Platform

زینب خورشیدی

سایه‌سمن - تهران، ایران

zkhoshidi@sayesaman.com

چکیده

با پیشرفت روزافزون مدل‌های شبکه‌های عصبی و روش‌های پردازش تصویر، امکان تحلیل خودکار تصاویر فروشگاه‌ها و استخراج اطلاعات قابل‌استفاده برای تحلیل بازار و ارزیابی عملیات اجرایی بیش از گذشته فراهم شده است. ارزیابی عملیات چیدمان در شبکه‌های پخش معمولاً بر مشاهده دستی تصاویر قبل و بعد، شمارش چشمی محصولات و قضاوت کارشناسان متکی است؛ فرایندی که در مقیاس بالا زمان‌بر، مستعد خطا و دشوار برای استانداردسازی است. پژوهش حاضر با رویکرد کاربردی-توسعه‌ای و در قالب مطالعه موردی پلتفرم B2Shelf¹ یک چارچوب انسان‌درحلقه برای تشخیص محصولات قفسه، پیش‌برچسب‌گذاری خودکار و پشتیبانی از ارزیابی عملیات چیدمان ارائه می‌کند. پیکره عملیاتی پروژه شامل تصاویر واقعی از قفسه‌ها و استندهای فروشگاه‌ها بود. نتایج اعتبارسنجی ذخیره‌شده شامل Precision برابر ۰٫۸۳۱۳، Recall برابر ۰٫۸۵۰۳، mAP@50 برابر ۰٫۸۹۵۵ و mAP@50-95 برابر ۰٫۵۹۴۸ بود. مدل آموزش‌دیده در محیط برچسب‌گذاری برای پیش‌برچسب‌گذاری تصاویر جدید به کار گرفته شد و خروجی آن پیش از ورود مجدد به دیتاست توسط نیروی انسانی بررسی و اصلاح می‌شود. برچسب‌های تأییدشده و محصولات جدید دوباره به دیتاست بازمی‌گردند و در چرخه‌های بعدی بازآموزی، به بهبود دامنه و عملکرد مدل کمک می‌کنند. تحلیل کیفی تصاویر قبل و بعد نیز نشان داد که خروجی مدل قابلیت تبدیل تغییرات بصری چینش به داده قابل‌اندازه‌گیری را دارد. نوآوری اصلی پژوهش در ترکیب داده واقعی صنعت پخش، تشخیص شیء و چرخه تکرارشونده پیش‌برچسب‌گذاری، کنترل انسانی و بهبود دیتاست نهفته است.

کلیدواژه‌ها: بینایی ماشین؛ YOLO11s؛ چیدمان؛ انسان درحلقه؛ سهم قفسه؛ صنعت پخش

¹ <https://b2shelf.com>

۱. مقدمه

تحول دیجیتال در صنعت پخش زمانی ارزش عملیاتی ایجاد می‌کند که داده‌های میدانی پراکنده را به شاخص‌های قابل اتکا برای تصمیم‌گیری تبدیل کند. یکی از منابع مهم داده در شرکت‌های پخش و تولیدکنندگان کالاهای تندگردش، تصاویر قفسه‌ها، یخچال‌ها و استندهای فروشگاهی است. این تصاویر می‌توانند وضعیت حضور برند، تعداد اقلام قابل مشاهده، میزان پرشدگی فضای عرضه، کیفیت چینش و تغییرات ناشی از عملیات چیدمان را منعکس کنند؛ اما در بسیاری از فرایندهای عملیاتی، ارزیابی آن‌ها همچنان به مشاهده انسانی وابسته است.

در پلتفرم B2Shelf، مأموریت‌های فروشگاهی در اختیار نیروهای اجرایی قرار می‌گیرد. نیروی اجرا پس از پذیرش مأموریت، از وضعیت اولیه قفسه یا استند تصویر تهیه می‌کند، عملیات چیدمان یا تکمیل کالا را انجام می‌دهد و تصویر پس از اجرا را بارگذاری می‌کند. تیم پشتیبانی سپس دو تصویر را بررسی می‌کند تا میزان تغییر، صحت اجرای مأموریت و مبنای پرداخت را ارزیابی کند. افزایش تعداد مأموریت‌ها موجب می‌شود بررسی دستی به گلوگاه عملیاتی تبدیل شود؛ زیرا شمارش محصولات در تصاویر شلوغ، تشخیص بسته‌بندی‌های مشابه و مقایسه دو تصویر با زاویه و نور متفاوت، زمان و تمرکز بالایی می‌طلبد.

تشخیص شیء می‌تواند محل و کلاس اقلام را به‌صورت هم‌زمان برآورد کند و خروجی تصویر را به داده ساختاریافته تبدیل نماید. با این حال، مسئله تشخیص محصول روی قفسه با چالش‌هایی مانند تراکم زیاد هم‌پوشانی، تفاوت کم میان بسته‌بندی‌ها، تغییر طراحی محصول، نور نامتوازن، زاویه دوربین و عدم توازن تعداد نمونه‌های کلاس‌ها مواجه است. مطالعات حوزه خرده‌فروشی نیز نشان داده‌اند که شناسایی محصول در محیط واقعی، به‌ویژه در صحنه‌های متراکم و برای کلاس‌های ریزدانه، از مسائل دشوار بینایی ماشین است. (Goldman et al., 2019; Tonioni et al., 2018; Wei et al., 2020)

پژوهش حاضر بر طراحی یک فرایند کاربردی و تکرارشونده متمرکز است؛ فرایندی که داده واقعی را دریافت می‌کند، با استفاده از مدل محصولات را تشخیص می‌دهد، برچسب‌های پیشنهادی را در اختیار کارشناس قرار می‌دهد و اصلاحات تأییدشده انسانی را به دیتاست باز می‌گرداند تا در چرخه‌های بعدی بازآموزی مورد استفاده قرار گیرد. این رویکرد علاوه بر کاهش وابستگی به برچسب‌گذاری کاملاً دستی، امکان توسعه تدریجی دامنه محصولات و استفاده از خروجی مدل در تحلیل عملیات چیدمان را فراهم می‌کند.

۱-۱. هدف و سؤال‌های پژوهش

هدف اصلی، طراحی و ارزیابی اولیه یک چارچوب انسان‌درحلقه برای خودکارسازی بخشی از فرایند تشخیص و برچسب‌گذاری محصولات قفسه و فراهم کردن داده لازم برای ارزیابی عملیات چیدمان است. سؤال‌های پژوهش عبارت‌اند از: (۱) مدل آموزش‌دیده با چه سطحی از دقت می‌تواند برندهای موجود در تصاویر واقعی را

تشخیص دهد؟ ۲) چگونه می‌توان مدل را در چرخه برچسب‌گذاری و اصلاح انسانی ادغام کرد؟ ۳) بازگشت برچسب‌های اصلاح‌شده و کلاس‌های جدید چگونه به تکمیل دیتاست و بازآموزی مدل کمک می‌کند؟ ۴) خروجی تشخیص چگونه می‌تواند مبنای شاخص‌هایی مانند شمارش محصول، تغییر قبل و بعد و سهم قفسه قرار گیرد؟

۲-۱. مشارکت‌های پژوهش

مشارکت‌های اصلی این پژوهش در پنج محور قابل جمع‌بندی است. این پژوهش با استفاده از تصاویر واقعی مأموریت‌های فروشگاه، یک پیکره دامنه‌محور برای تشخیص محصولات ایجاد کرده و مدل YOLO11s را برای شناسایی چندکلاسه برندها در شرایط واقعی قفسه آموزش داده است. همچنین، با به‌کارگیری رویکرد انسان‌درحلقه، خروجی مدل در فرایند پیش‌برچسب‌گذاری، اصلاح انسانی و توسعه تدریجی کلاس‌ها مورد استفاده قرار گرفته است. در سطح کاربردی نیز خروجی تشخیص برای مقایسه وضعیت قبل و بعد از چیدمان و استخراج شاخص‌هایی مانند تعداد محصولات، میزان تغییر و سهم قفسه به داده‌های قابل تحلیل تبدیل شده است. جدول ۱ این مشارکت‌ها را به‌صورت خلاصه نشان می‌دهد.

جدول ۱. مشارکت‌های اصلی پژوهش

ردیف	مشارکت	شرح
۱	ایجاد پیکره دامنه‌محور	استفاده از تصاویر واقعی مأموریت‌های فروشگاه به‌جای تصاویر استودیویی یا داده عمومی
۲	مدل تشخیص چندکلاسه	آموزش YOLO11s برای تشخیص ۱۲ کلاس برند در شرایط واقعی قفسه
۳	چرخه انسان‌درحلقه	استفاده از خروجی مدل برای پیش‌برچسب‌گذاری، اصلاح کارشناس و توسعه تدریجی کلاس‌ها
۴	ارزیابی عملیات چیدمان	استفاده از خروجی تشخیص برای مقایسه وضعیت قبل و بعد از چینش و تبدیل تغییرات بصری به داده قابل تحلیل
۵	تعریف لایه تحلیلی	صورت‌بندی شاخص‌های شمارش، تغییر قبل/بعد و سهم قفسه بر پایه خروجی جعبه‌های مرزی

۲. پیشینه پژوهش

۱-۲. تشخیص و شناسایی محصولات در محیط خرده‌فروشی

پژوهش‌های شناسایی محصول در فروشگاه از طبقه‌بندی تصاویر منفرد به سمت تحلیل صحنه‌های واقعی قفسه حرکت کرده‌اند. George et al. (2015) مسئله شناسایی ریزدانه محصولات را در تصاویر تلفن همراه بررسی کردند و بر تفاوت اندک میان کلاس‌ها و تغییر مستمر طراحی بسته‌بندی تأکید داشتند (Tonioni et al., 2018). یک خط لوله دومرحله‌ای شامل تشخیص محصول و بازیابی مبتنی بر شباهت پیشنهاد کردند تا افزودن اقلام جدید با تعداد محدودی تصویر مرجع ممکن شود. این آثار نشان می‌دهند که سامانه عملیاتی باید علاوه بر دقت، قابلیت توسعه کلاس و انطباق با تغییر بسته‌بندی را نیز داشته باشد.

Goldman et al. (2019) با معرفی مجموعه SKU-110K نشان دادند که صحنه‌های خرده‌فروشی مترکم به دلیل نزدیکی و شباهت اشیاء، یک مسئله مستقل و دشوار در تشخیص شیء محسوب می‌شوند.

در ادامه، مجموعه‌هایی مانند RP2K با بیش از دو هزار محصول و Unitail با وظایف تشخیص، خواندن متن و تطبیق، ابعاد مختلف تحلیل تصویر در خرده‌فروشی را پوشش دادند (Chen et al., 2022; Peng et al., 2020). با وجود ارزش این مجموعه‌ها برای مقایسه الگوریتم‌ها، مدل‌های سازمانی معمولاً به داده دامنه‌محور نیاز دارند؛ زیرا نوع محصولات، نحوه چیدمان، کیفیت دوربین و اهداف مدیریتی هر شرکت متفاوت است.

مرور Wei et al. (2020) چالش‌های اصلی شناسایی محصول را شامل تعداد زیاد SKUها، کمبود داده برچسب‌خورده، تغییر دامنه میان تصاویر مرجع و فروشگاه و انسداد معرفی می‌کند.

این چالش‌ها در تصاویر B2Shelf نیز مشاهده می‌شوند: برخی بسته‌ها کوچک یا نیمه‌پوشیده‌اند، چندین محصول از یک خانواده طراحی مشابه دارند و تراکم تصویر پس از چیدمان به‌طور محسوسی افزایش می‌یابد.

۲-۲. برچسب‌گذاری نیمه‌خودکار و انسان‌درحلقه

برچسب‌گذاری کادرهای شامل محصولات یکی از پرهزینه‌ترین مراحل توسعه سامانه‌های تشخیص شیء است. Adhikari and Huttunen (2021) روشی تکرارشونده پیشنهاد کردند که در آن مدل روی یک بخش کوچک آموزش می‌بیند، برای دسته بعدی جعبه پیشنهاد می‌کند و کاربر صرفاً خطاها را اصلاح می‌کند. این روش در داده‌های آزمایشی آن‌ها توانست بخش قابل‌توجهی از کار دستی را کاهش دهد. Marchesoni-Acland (2023) and Facciolo نیز در چارچوب IAdet، حاشیه‌نویسی کمکی، آموزش پس‌زمینه و انتخاب داده را در یک حلقه واحد قرار دادند. ادبیات انسان‌درحلقه تأکید میکند که نقش انسان نه حذف کامل، بلکه انتقال از «تولید همه برچسب‌ها» به «بازبینی و اصلاح موارد نامطمئن» است (Budd et al., 2021).

چارچوب B2Shelf از همین منطق استفاده می‌کند، اما آن را در چرخه واقعی تولید داده در صنعت پخش به کار می‌گیرد. مدل برای محصولات شناخته‌شده جعبه و کلاس اولیه تولید می‌کند؛ پشتیبان خطاهای مکان‌یابی یا طبقه‌بندی را اصلاح می‌کند و محصولات جدید را به‌صورت دستی برچسب می‌زند. داده اصلاح‌شده دوباره به دیتاست باز می‌گردد و در دوره‌های بعدی بازآموزی مورد استفاده قرار می‌گیرد. بنابراین، نقش نیروی انسانی از ترسیم همه برچسب‌ها از ابتدا به کنترل کیفیت، اصلاح موارد نادرست و تکمیل کلاس‌های جدید تغییر می‌کند.

۳-۲. شکاف پژوهشی

بخش عمده مطالعات موجود بر افزایش دقت الگوریتم یا معرفی مجموعه‌داده تمرکز دارد؛ در حالی که برای صنعت پخش، نحوه شکل‌گیری تدریجی دیتاست، استفاده از مدل برای پیش‌برچسب‌گذاری، کنترل انسانی و بازگشت اصلاحات به چرخه آموزش نیز اهمیت زیادی دارد. پژوهش حاضر این شکاف را با یک مطالعه موردی مبتنی بر داده واقعی پوشش می‌دهد. معیار موفقیت در اینجا فقط mAP نیست، بلکه توانایی تبدیل تصاویر عملیاتی به داده ساختاریافته و ایجاد یک چرخه قابل تکرار برای تکمیل دیتاست و بهبود مدل است.

۳. روش تحقیق

۳-۱. طرح پژوهش و جامعه مورد مطالعه

پژوهش از نوع کاربردی-توسعه‌ای و روش آن مطالعه موردی فنی است. جامعه مورد مطالعه، تصاویر قفسه و اسناد ثبت‌شده در مأموریت‌های واقعی پلتفرم B2Shelf است. واحد تحلیل در لایه یادگیری ماشین، هر نمونه محصول قابل مشاهده در تصویر است و واحد تحلیل در لایه عملیاتی، زوج تصویر قبل و بعد از چیدمان محسوب می‌شود. مراحل پژوهش شامل گردآوری داده، تعریف کلاس‌ها، برچسب‌گذاری اولیه، تشکیل دیتاست، آموزش مدل، پیش‌برچسب‌گذاری تصاویر جدید، بازبینی انسانی، بازگشت اصلاحات به دیتاست، بازآموزی دوره‌ای و ارزیابی عملکرد مدل است.

۳-۲. فرایند عملیاتی B2Shelf

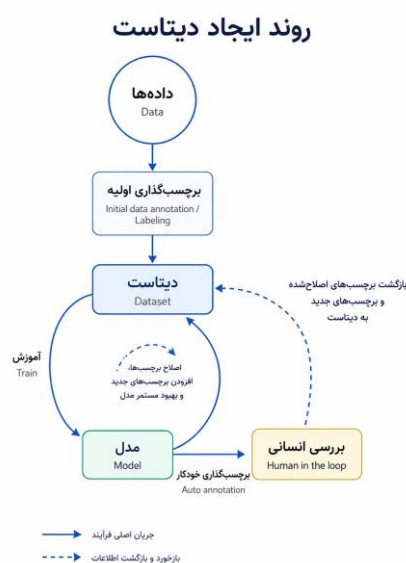
فرایند از ایجاد یا نمایش مأموریت فروشگاهی در اپلیکیشن آغاز می‌شود. نیروی اجرایی مأموریت را می‌پذیرد به نقطه فروش مراجعه می‌کند و تصویر قبل را ثبت می‌نماید. پس از چیدمان، تکمیل یا اصلاح وضعیت نمایش محصول، تصویر بعد بارگذاری می‌شود. در فرایند سنتی، پشتیبان تصاویر را مشاهده و میزان تغییر را برآورد می‌کند. در چارچوب پیشنهادی، مدل بینایی ماشین ابتدا محصولات شناخته‌شده را تشخیص می‌دهد و خروجی ساختاریافته برای بازبینی انسانی و تحلیل بعدی تولید می‌شود.

جدول ۲. جریان داده و یادگیری در چارچوب پیشنهادی

مرحله	مولفه	خروجی
۱	اپلیکیشن B2Shelf	ثبت مأموریت و تصاویر قبل/بعد
۲	مخزن تصاویر	نگهداری داده خام و متادیتای مأموریت
۳	برچسب‌گذاری اولیه	تعریف کادرهای شامل محصولات و کلاس محصولات در بخشی از تصاویر
۴	آموزش YOLO۱۱s	یادگیری مکان و کلاس محصولات از دیتاست برچسب‌خورده
۵	پیش‌برچسب‌گذاری مدل	تولید برچسب اولیه برای تصاویر جدید
۶	بازبینی انسانی	اصلاح خطاها و افزودن محصولات یا کلاس‌های جدید
۷	به‌روزرسانی و بازآموزی	بازگشت برچسب‌های تأییدشده به دیتاست و بهبود دوره‌ای مدل

۳-۳. فرایند ایجاد دیتاست و چرخه انسان‌درحلقه

در این پژوهش، دیتاست به‌صورت یک مجموعه ثابت و یک‌باره در نظر گرفته نشد، بلکه در قالب یک چرخه تکرارشونده میان داده، مدل و کنترل انسانی توسعه یافت. مطابق شکل ۱، فرایند از تصاویر خام قفسه‌ها و اسندهای فروشگاه‌های آغاز می‌شود در مرحله نخست، بخشی از تصاویر به‌صورت دستی برچسب‌گذاری می‌شوند؛ به این معنا که محل هر محصول با کادرهای شامل محصولات مشخص و کلاس مربوط به آن ثبت می‌شود. مجموعه تصاویر و برچسب‌های تأییدشده، دیتاست اولیه را تشکیل می‌دهد و مبنای آموزش نخستین نسخه مدل قرار می‌گیرد.



شکل ۱. فرایند ایجاد دیتاست و بهبود تکرارشونده مدل با رویکرد انسان‌درحلقه

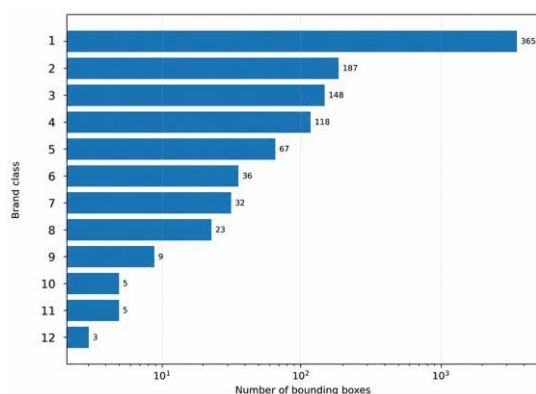
پس از آموزش، مدل برای تصاویر جدید نقش پیش‌برچسب‌گذار را ایفا می‌کند و برای محصولات شناخته‌شده جعبه و کلاس تولید می‌نماید. این خروجی بدون کنترل مستقیم وارد دیتاست نمی‌شود. در مرحله بررسی انسانی، پشتیبان یا کارشناس برچسب‌های پیشنهادی را کنترل می‌کند. موارد نادرست را حذف یا اصلاح می‌کند، مرز جعبه‌ها را در صورت نیاز تغییر می‌دهد و برای محصولاتی که مدل هنوز نمی‌شناسد برچسب جدید ایجاد می‌کند. به این ترتیب، انسان در حلقه نقش کنترل کیفیت و تکمیل دانش مدل را بر عهده دارد. برچسب‌های اصلاح‌شده و کلاس‌های جدید پس از تأیید به دیتاست بازگردانده می‌شوند. با افزایش حجم داده‌های تأییدشده، نسخه به‌روزشده دیتاست در دور بعدی برای بازآموزی مدل استفاده می‌شود. تکرار این چرخه باعث می‌شود پوشش محصولات افزایش یابد و مدل به‌تدریج با نمونه‌های متنوع‌تر، شرایط تصویری جدید و کلاس‌های تازه سازگار شود. مزیت این رویکرد آن است که وابستگی به برچسب‌گذاری کاملاً دستی کاهش می‌یابد، در حالی که کنترل انسانی مانع از انباشته‌شدن خطاها در دیتاست می‌شود. بنابراین چرخه پیشنهادی هم‌زمان یک سازوکار تولید داده، کنترل کیفیت و بهبود دوره‌ای مدل محسوب می‌شود.

۳-۴. داده‌ها و فرایند برچسب‌گذاری اولیه

طبق گزارش پروژه، حدود ۱۶۱۳ تصویر واقعی برای توسعه مدل گردآوری شد. تصاویر شامل نمونه‌های قبل و بعد از عملیات، اندازه‌ها و نسبت‌های متفاوت و شرایط متنوع فروشگاه بودند. برچسب‌ها به‌شکل کادرهای شامل محصولات و نام برند در محیط Labeling^۲ ثبت شدند.

جدول ۳. مشخصات داده‌های در دسترس پژوهش

مقدار	شاخص
۱۶۱۳	تعداد تصاویر پیکره عملیاتی
۴۲۹۱	تعداد جعبه‌های مرزی در نمونه
۱۲	کلاس‌های مدل نهایی



شکل ۲. توزیع لگاریتمی جعبه‌های مرزی کلاس‌ها در فایل XML نمونه

محیط برچسب‌گذاری^۲

۳-۵. آموزش مدل YOLO11s

مدل YOLO11s از وزن‌های ازپیش‌آموزش‌دیده مقداردهی اولیه شد. YOLO11 از خانواده مدل‌های یک‌مرحله‌ای Ultralytics^۳ است و برای تشخیص، قطعه‌بندی، طبقه‌بندی و سایر وظایف بینایی ماشین ارائه شده است (Ultralytics, 2024). نسخه S با هدف ایجاد توازن میان دقت و پیچیدگی محاسباتی انتخاب شد.

جدول ۴. تنظیمات اصلی آموزش مدل

پارامتر	مقدار
معماری	YOLO11s
اندازه ورودی	۶۴۰ × ۶۴۰ پیکسل
تعداد دوره آموزش ^۴	۲۶۰
Batch size ^۵	۴
بهینه‌ساز	AdamW

۳-۶. معیارهای ارزیابی

ارزیابی مدل بر مبنای معیارهای رایج تشخیص شیء انجام شد. دقت مثبت یا Precision نسبت تشخیص‌های درست به کل تشخیص‌های مثبت مدل است بازخوانی یا Recall سهم نمونه‌های واقعی کشف‌شده را نشان می‌دهد. در روابط زیر TP تعداد مثبت درست، FP تعداد مثبت کاذب و FN منفی است:

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

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

شاخص mAP میانگین مساحت زیر منحنی Precision-Recall در کلاس‌ها است. mAP@50 از آستانه هم‌پوشانی IoU برابر ۰٫۵ استفاده می‌کند، در حالی که mAP@50-95 میانگین آستانه‌های ۰٫۵ تا ۰٫۹۵ است و نسبت به دقت مکان‌یابی حساس‌تر است. استفاده از این دو شاخص امکان تفکیک «تشخیص درست کلاس» از «مکان‌یابی دقیق جعبه» را فراهم می‌کند. (Lin et al., 2014)

^۳ آموزش، ارزیابی و استقرار مدل‌های بینایی ماشین YOLO

^۴ Epoch

^۵ اندازه دسته

۷-۳. شاخص‌های پیشنهادی برای تحلیل چیدمان

خروجی تشخیص می‌تواند سهم قفسه و حضوربصری برند را برآورد کند. برای برند c سهم تعدادی قفسه از نسبت تعداد محصولات قابل مشاهده آن برند به تعداد کل محصولات تشخیص داده شده به دست می‌آید:

$$\text{ShelfShare}_{(c)} = \frac{N_c}{\sum N_j} \times 100 \quad (3)$$

در صورت کنترل زاویه و محدوده قفسه، سهم سطحی نیز از مجموع مساحت جعبه‌های یک برند نسبت به مساحت جعبه‌های همه برندها قابل محاسبه است:

$$\text{ShelfShare}_{\text{area}(c)} = \frac{\sum \text{Area}(\text{BBox}_c)}{\sum \text{Area}(\text{BBox}_{\text{all}})} \times 100 \quad (4)$$

برای زوج تصاویر متناظر، تغییر تعداد قابل مشاهده برند از رابطه زیر محاسبه می‌شود:

$$\Delta N_c = N_{c, \text{after}} - N_{c, \text{befor}} \quad (5)$$

۴. یافته‌ها و تحلیل

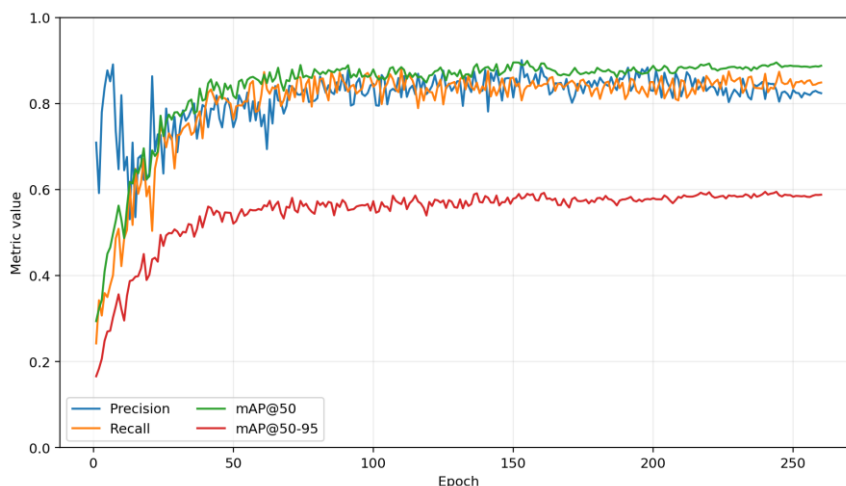
۴-۱. عملکرد مدل در اعتبارسنجی

جدول ۵. نتایج اعتبارسنجی ذخیره شده در نقطه واری نهایی

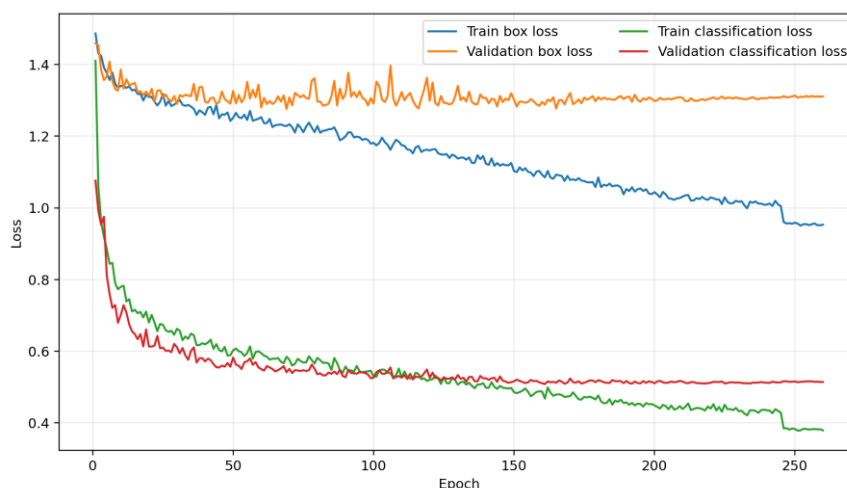
معیار	مقدار	تفسیر
Precision	۰٫۸۳۱۳۳	درصد از تشخیص‌های مثبت مدل درست بوده‌اند ۸۳٫۱۳
Recall	۰٫۸۵۰۲۵	مدل حدود ۸۵٫۰۳ درصد نمونه‌های واقعی را بازیابی کرده است
mAP@50	۰٫۸۹۵۵۴	عملکرد مناسب در آستانه IoU برابر ۰٫۵
mAP@50-95	۰٫۵۹۴۸۲	کاهش طبیعی در آستانه‌های سخت گیرانه تر مکان‌یابی

مقدار mAP@50 نزدیک به ۰٫۹ نشان می‌دهد که مدل در سطح هم‌پوشانی متعارف، توان مناسبی برای تشخیص برندها در تصاویر واقعی دارد. فاصله میان mAP@50 و mAP@50-95 بیانگر آن است که بخشی از خطاها به دقت مکان‌یابی جعبه‌ها مربوط می‌شود. این موضوع برای شمارش معمولاً کمتر از کاربردهای اندازه‌گیری دقیق سطح مسئله ساز است؛ اما برای محاسبه سهم سطحی قفسه باید با دقت بیشتری بررسی شود. بررسی منحنی‌های آموزش نشان داد که معیارها در دوره‌های ابتدایی به سرعت افزایش یافته و سپس به

ناحیه پایدار رسیده‌اند. بیشترین $mAP@50$ ثبت‌شده ۰٫۸۹۹۰۵ در دوره ۱۵۵ و بیشترین $mAP@50-95$ برابر ۰٫۵۹۴۸۲ در دوره ۲۴۴ بود. این الگو نشان می‌دهد ادامه آموزش پس از رسیدن به ناحیه پایدار، بهبود محدودی ایجاد کرده است و انتخاب نقطه واریسی بر مبنای معیار هدف اهمیت دارد.



شکل ۳. روند Precision، Recall و mAP طی ۲۶۰ دوره آموزش



شکل ۴. روند خطای آموزش و اعتبارسنجی

۲-۴. تحلیل نمونه قبل و بعد از چیدمان

در نمونه خام قبل از عملیات، بخش قابل توجهی از فضای استند خالی، چیدمان پراکنده و تعدادی محصول روی کف استند مشاهده می‌شود. در تصویر بعد، ردیف‌های محصول منظم‌تر و سطح بیشتری از استند پوشش داده شده است این زوج تصویر ارزش عملیاتی مسئله را روشن می‌کند: تغییر بصری برای انسان آشکار است اما تبدیل آن به تعداد، کلاس و شاخص قابل مقایسه نیازمند تشخیص خودکار است.

در نمونه دیگری که خروجی مدل روی تصاویر ترسیم شده است، تصویر قبل تعداد محدودی جعبه تشخیص در فضای نسبتاً خالی دارد؛ در حالی که تصویر بعد شامل تعداد بیشتری جعبه و تنوع بالاتر محصولات است. این مشاهده کیفی نشان می‌دهد که خروجی تشخیص می‌تواند مبنایی برای مقایسه تعداد اقلام قابل مشاهده تغییر پوشش قفسه و ارزیابی اثر عملیات چیدمان باشد. برای گزارش عددی دقیق در سطح مأموریت، لازم است تصاویر قبل و بعد به صورت صحیح جفت شوند و شاخص‌های یکسان روی هر دو تصویر محاسبه گردد.



شکل ۵. مقایسه تصاویر خام قبل (a) و بعد (b) و خروجی مدل روی تصاویر قبل (c) و بعد (d) از چیدمان

۳-۴. ارزش عملیاتی چرخه انسان در حلقه

در مرحله اولیه، کارشناس باید جعبه تمام محصولات را از ابتدا ترسیم و کلاس آن‌ها را انتخاب می‌کند. در چرخه جدید، مدل برای کلاس‌های آموخته‌شده پیش‌برچسب تولید می‌کند و نقش پشتیبان به کنترل، حذف تشخیص‌های نادرست، اصلاح مرزها و افزودن محصولات جدید تغییر می‌یابد. این تغییر ساختاری دو پیامد دارد: نخست، آماده‌سازی داده جدید ساده‌تر می‌شود؛ دوم، اصلاحات انسانی به‌عنوان داده تأییدشده برای نسخه‌های بعدی دیتاست و بازآموزی مدل قابل استفاده است.

۴-۴. کاربردهای عملیاتی و دامنه خروجی مدل

جدول ۶. کاربردهای عملیاتی خروجی مدل در چارچوب پیشنهادی

قابلیت	شرح
تشخیص برند و کادرهای شامل محصولات	شناسایی کلاس محصول و موقعیت آن در تصویر
شمارش هر برند	تجمع تشخیص‌ها برای برآورد تعداد اقلام قابل مشاهده هر برند
پیش‌برچسب‌گذاری تصاویر	تولید برچسب اولیه برای محصولات شناخته‌شده در محیط Labeling
بازبینی و اصلاح انسانی	کنترل خطاهای مدل، اصلاح جعبه‌ها و تأیید برچسب‌ها پیش از بازگشت به دیتاست
افزودن محصولات و کلاس‌های جدید	برچسب‌گذاری دستی محصولات جدید و استفاده از آن‌ها در چرخه‌های بعدی بازآموزی
مقایسه قبل و بعد	نیازمند جفت‌سازی صحیح تصاویر متناظر و محاسبه شاخص‌های یکسان
محاسبه سهم قفسه	امکان محاسبه سهم تعدادی یا سطحی بر اساس خروجی تشخیص، با رعایت شرایط مقایسه

۵. یافته‌های طراحی

یافته‌های این مطالعه نشان می‌دهد که استفاده از YOLO11s در تصاویر واقعی B2Shelf از منظر دقت اعتبارسنجی و قابلیت استفاده در چرخه برچسب‌گذاری قابل دفاع است. صحنه‌های متراکم و بسته‌بندی‌های مشابه، مکان‌یابی دقیق را دشوار می‌کنند و تغییر مستمر محصولات، رویکرد ایستا را ناکافی می‌سازد (Goldman et al., 2019; George et al. 2015). چرخه B2Shelf با نگرداشتن کارشناس در حلقه،

این مسئله را به فرایند تکرارشونده تکمیل دیتاست و بازآموزی مدل تبدیل می‌کند. در مقایسه با روش‌های برچسب‌گذاری تکرارشونده، مزیت مطالعه حاضر استفاده از تصاویر واقعی عملیات چیدمان و بازگشت اصلاحات کارشناسان به دیتاست است. در فرایند استنتاج فعلی، آستانه $Confidence^6$ برابر ۰.۸۰ در نظر گرفته شده است. این آستانه احتمال مثبت کاذب را کاهش می‌دهد.

۶. نتیجه‌گیری و پیشنهادها

این پژوهش یک چارچوب انسان‌درحلقه مبتنی بر YOLO11s برای تشخیص محصولات در قفسه و پشتیبانی از ارزیابی عملیات چیدمان ارائه کرد. مدل با داده‌های واقعی Shelf2B توسعه یافت و در اعتبارسنجی به Precision برابر ۰.۸۳۱۳، Recall برابر ۰.۸۵۰۳، $mAP@50$ برابر ۰.۸۹۵۵ و $mAP@50-95$ برابر ۰.۵۹۴۸ دست یافت. نتایج نشان داد که تصاویر عملیاتی قفسه را می‌توان با استفاده از تشخیص شیء به داده‌های قابل شمارش و تحلیل تبدیل کرد و از مدل برای پیش‌برچسب‌گذاری تصاویر جدید و کاهش وابستگی به برچسب‌گذاری کاملاً دستی استفاده نمود.

نوآوری اصلی، استفاده صرف از یک نسخه YOLO نیست؛ بلکه یکپارچه‌سازی چرخه داده واقعی، برچسب‌گذاری اولیه، پیش‌بینی مدل، اصلاح انسانی و افزودن کلاس‌های جدید در یک فرایند تکرارشونده است. در این چرخه، برچسب‌های اصلاح‌شده به دیتاست بازمی‌گردند و در بازآموزی‌های بعدی به تکمیل دامنه مدل کمک می‌کنند. این ترکیب می‌تواند بار کار دستی را کاهش دهد، استانداردسازی ارزیابی را بهبود بخشد و داده لازم برای محاسبه سهم قفسه و تغییر قبل/بعد را فراهم کند.

در ادامه این مسیر، پیشنهاد می‌شود از مدل‌های جدید تشخیص و تقسیم‌بندی عمومی اشیاء استفاده شود که بتوانند نواحی مربوط به محصولات موجود در تصویر را بدون نیاز به ترسیم دستی اولیه شناسایی کنند. در این حالت، مدل قادر محصولات یا ناحیه پیشنهادی هر محصول را تولید کرده و نیروی انسانی تنها وظیفه تأیید، اصلاح و تعیین کلاس یا برچسب صحیح محصول را بر عهده خواهد داشت. این رویکرد می‌تواند زمان و هزینه برچسب‌گذاری اولیه را بیش از پیش کاهش دهد. همچنین در مراحل بعدی می‌توان از مدل‌های چندوجهی یا مدل‌های دارای قابلیت تشخیص بازکلاس برای پیشنهاد خودکار برچسب محصول استفاده کرد تا وابستگی فرایند به مداخله انسانی کمتر شود و چرخه توسعه دیتاست و بازآموزی مدل با سرعت بیشتری انجام گیرد.

^۶ امتیاز اطمینان

- Adhikari, B., & Huttunen, H. (2021). Iterative bounding box annotation for object detection. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 4040–4046). IEEE. <https://doi.org/10.1109/ICPR48806.2021.9412956>
- Budd, S., Robinson, E. C., & Kainz, B. (2021). A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical Image Analysis*, 71, 102062. <https://doi.org/10.1016/j.media.2021.102062>
- Chen, F., Zhang, H., Li, Z., Dou, J., Mo, S., Chen, H., Zhang, Y., Ahmed, U., Zhu, C., & Savvides, M. (2022). Unitail: Detecting, reading, and matching in retail scene. In *Computer Vision – ECCV 2022* (pp. 705–722). Springer. https://doi.org/10.1007/978-3-031-20071-7_41
- CVAT.ai Corporation. (2023). Computer Vision Annotation Tool (CVAT) (Version 2.4.8) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.8070041>
- George, M., Mircic, D., Sörös, G., Floerkemeier, C., & Mattern, F. (2015). Fine-grained product class recognition for assisted shopping. In 2015 IEEE International Conference on Computer Vision Workshop (ICCVW) (pp. 546–554). IEEE. <https://doi.org/10.1109/ICCVW.2015.79>
- Goldman, E., Herzig, R., Eisenschtat, A., Ratzon, O., Levi, I., Goldberger, J., & Hassner, T. (2019). Precise detection in densely packed scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5222–5231).
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2014). Microsoft COCO: Common objects in context. In *Computer Vision – ECCV 2014* (pp. 740–755). Springer. https://doi.org/10.1007/978-3-319-10602-1_48
- Marchesoni-Acland, F., & Facciolo, G. (2023). IAdet: Simplest human-in-the-loop object detection. *arXiv*. <https://doi.org/10.48550/arXiv.2307.01582>
- Peng, J., Xiao, C., & Li, Y. (2020). RP2K: A large-scale retail product dataset for fine-grained image classification. *arXiv*. <https://doi.org/10.48550/arXiv.2006.12634>
- Sinha, A., Banerjee, S., & Chattopadhyay, P. (2022). An improved deep learning approach for product recognition on racks in retail stores. *arXiv*. <https://doi.org/10.48550/arXiv.2202.13081>
- Tonioni, A., Serra, E., & Di Stefano, L. (2018). A deep learning pipeline for product recognition on store shelves. In 2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS). IEEE. <https://doi.org/10.1109/IPAS.2018.8708890>
- Ultralytics. (2024). Ultralytics YOLO11. <https://docs.ultralytics.com/models/yolo11/>
- Ultralytics. (2024). Object detection. <https://docs.ultralytics.com/tasks/detect/>
- Wei, Y., Tran, S., Xu, S., Kang, B., & Springer, M. (2020). Deep learning for retail product recognition: Challenges and techniques. *Computational Intelligence and Neuroscience*, 2020, 8875910. <https://doi.org/10.1155/2020/8875910>