



Intelligent Detection and Separation of Recyclable Urban Waste Using a Deep Learning-Based Computer Vision Model

Ali Jelokhani 

School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran. Email: ali.jelokhani@ut.ac.ir

ARTICLE INFO

Article type:
Research Paper

Article History:

Received 27 February 2026
Revised 07 April 2026
Accepted 28 May 2026
Published Online 23 September 2026

Keywords:

Municipal waste,
Recycling,
Intelligent waste separation,
Computer vision,
Deep learning,
Object detection.

ABSTRACT

Urban waste management is one of the major environmental and economic challenges in modern cities. A considerable part of municipal solid waste consists of recyclable materials such as cardboard, glass, metal, paper, and plastic. Accurate detection and separation of these materials can improve recycling quality and reduce the environmental burden of urban waste. This study proposes a deep learning-based computer vision framework for intelligent detection and separation of recyclable urban waste. Unlike single-label image classification methods, the proposed framework performs multi-instance object detection and generates a bounding box, class label, and confidence score for each detected item. The architecture includes preprocessing and augmentation, lightweight feature extraction, a combined attention module, multi-level feature fusion, and a final detection stage. The dataset contains 4250 images and 6815 annotated waste instances. The proposed method achieved Precision, Recall, F1-score, mAP@0.5, and mAP@0.5:0.95 values of 0.934, 0.919, 0.926, 0.947, and 0.742, respectively. The results indicate that the proposed framework can support automated waste sorting systems, smart bins, robotic separators, and recycling facilities.

Cite this article: Jelokhani, A. (2026). Intelligent Detection and Separation of Recyclable Urban Waste Using a Deep Learning-Based Computer Vision Model. *Urban Development Policy Making*, 3 (3), 411-431. DOI: <http://doi.org/10.22034/judpm.2026.582800.1106>



© Authors retain the copyright and full publishing rights.
DOI: <http://doi.org/10.22034/judpm.2026.582800.1106>

Introduction

The growth of urban population, consumption, and disposable packaging has increased the amount of municipal waste. In many cities, recyclable waste separation is still performed manually or semi-automatically, which can be slow, costly, and prone to human error. Since recyclable materials may appear together in real waste sorting environments, simple image classification is not sufficient for practical separation. Object detection is more suitable because it can determine both the class and location of each waste item. This study focuses on five recyclable categories: cardboard, glass, metal, paper, and plastic, and aims to provide an intelligent framework for detecting multiple waste objects in a single image.

Methodology

This research is applied-developmental in purpose and quantitative-experimental in approach. The study uses the public EE297_Project dataset published in Roboflow Universe under the CC BY 4.0 license. The dataset includes 4250 images and 6815 bounding-box annotations from five recyclable waste classes. The data were divided into training, validation, and test sets with ratios of 70%, 15%, and 15%, respectively. Data augmentation was applied only to the training set to improve

generalization and prevent data leakage. After augmentation, the effective number of training images increased from 2975 to 14875, and the effective number of training instances increased from 4773 to 23865.

The proposed architecture consists of preprocessing, lightweight feature extraction, a combined attention module, multi-level feature fusion, and a detection head. In the feature extraction stage, depthwise and pointwise convolutions are used to reduce computational complexity. The combined attention module emphasizes informative feature channels and important spatial regions. Multi-level feature fusion combines local and semantic features from different layers, allowing the model to detect objects with different sizes and appearances. The final detection stage produces a bounding box, class label, and confidence score for each detected waste object.

The model was trained using 640×640 input images, a batch size of 16, 120 epochs, and the AdamW optimizer with an initial learning rate of 0.0001. Performance was evaluated using Precision, Recall, F1-score, mAP@0.5, and mAP@0.5:0.95. To assess reliability, each model was evaluated in five independent runs with different random seeds, and paired t-tests were used to examine the statistical significance of performance differences.

Results

The proposed method achieved a Precision of 0.934, Recall of 0.919, F1-score of 0.926, mAP@0.5 of 0.947, and mAP@0.5:0.95 of 0.742 on the test set. These results show that the model performs well in both waste recognition and localization. Class-wise evaluation showed stable performance across all five classes, with stronger results for plastic and metal and relatively lower performance for paper. The difference between mAP@0.5 and mAP@0.5:0.95 indicates that part of the challenge is related to precise localization of bounding boxes.

The proposed method was compared with Faster R-CNN, SSD, RetinaNet, YOLOv8n, MRS-YOLO, and a lightweight recyclable domestic waste detection method. The proposed model achieved higher overall performance across the main metrics. Statistical analysis showed that the improvements were significant at the 0.05 level. The model was also evaluated under difficult imaging conditions, including brightness changes, contrast reduction, noise, blur, cluttered background, and partial occlusion. In all tested conditions, mAP@0.5 remained above 0.91, although blur and cluttered backgrounds caused the largest performance drops. Qualitative results also showed that the proposed method can detect multiple waste objects in a single image and produce accurate bounding boxes in complex scenes.

Conclusion

This study presented a deep learning-based computer vision framework for intelligent detection and separation of recyclable urban waste. The proposed method can identify multiple waste objects in a single image and generate object-level outputs suitable for practical sorting systems. The results show that lightweight feature extraction, combined attention, and multi-level feature fusion improve recyclable waste detection, especially in multi-object scenes and difficult imaging conditions. Although the results are promising, future research should use more diverse real-world data, optimize the model for edge devices, and integrate the detection framework with robotic sorting mechanisms.



تشخیص و جداسازی هوشمند پسماندهای شهری قابل بازیافت با استفاده از مدل بینایی ماشین مبتنی بر یادگیری عمیق

علی جلوخانی

دانشکده کامپیوتر، دانشکده فنی دانشگاه تهران. رایانامه: ali.jelokhani@ut.ac.ir

چکیده

اطلاعات مقاله

نوع مقاله:

پژوهشی

تاریخ‌های مقاله:

تاریخ دریافت: ۱۴۰۴/۱۲/۰۸

تاریخ بازنگری: ۱۴۰۵/۰۱/۱۸

تاریخ پذیرش: ۱۴۰۵/۰۳/۰۷

تاریخ انتشار: ۱۴۰۵/۰۷/۰۱

کلیدواژه:

پسماند شهری،

بازیافت،

جداسازی هوشمند پسماند،

بینایی ماشین،

یادگیری عمیق،

تشخیص شیء.

با افزایش حجم پسماندهای شهری و اهمیت بازیافت در کاهش آلودگی محیط زیست، جداسازی دقیق مواد قابل بازیافت به یکی از مراحل مهم مدیریت هوشمند پسماند تبدیل شده است. روش‌های سنتی تفکیک پسماند معمولاً به نیروی انسانی وابسته بوده و در مقیاس‌های بزرگ با محدودیت‌هایی مانند سرعت پایین، خطای انسانی، هزینه عملیاتی بالا و خطرات بهداشتی همراه هستند. در این پژوهش، یک چارچوب مبتنی بر بینایی ماشین و یادگیری عمیق برای تشخیص و جداسازی هوشمند پسماندهای شهری قابل بازیافت ارائه می‌شود. هدف روش پیشنهادی، تشخیص و مکان‌یابی پسماندها در پنج گروه اصلی شامل مقوا، شیشه، فلز، کاغذ و پلاستیک است. به خلاف روش‌های طبقه‌بندی ساده که تنها یک برچسب برای کل تصویر تولید می‌کنند، چارچوب پیشنهادی قادر است در تصاویر تک‌شیء و چندشیء، چند نوع پسماند را به صورت هم‌زمان شناسایی کرده و برای هر نمونه، جعبه محدودکننده، کلاس و امتیاز اطمینان تولید کند. در معماری پیشنهادی، از استخراج ویژگی سبک، ماژول توجه ترکیبی و تلفیق ویژگی‌های چندسطحی استفاده شده است تا سامانه بتواند ویژگی‌های ظاهری، بافتی و ساختاری مواد مختلف را بهتر تشخیص دهد. مجموعه داده مورد استفاده شامل ۴۲۵۰ تصویر و ۶۸۱۵ نمونه برچسب‌خورده در قالب جعبه محدودکننده است. نتایج آزمایش‌ها نشان داد روش پیشنهادی با دستیابی به Precision برابر با ۰/۹۳۴، Recall برابر با ۰/۹۱۹، F1-score برابر با ۰/۹۲۶، مقدار mAP@0.5 برابر با ۰/۹۴۷ و مقدار mAP@0.5:0.95 برابر با ۰/۷۴۲ عملکرد مناسبی در تشخیص و مکان‌یابی پسماندهای قابل بازیافت دارد. همچنین، نتایج به تفکیک کلاس‌ها نشان داد مدل در تشخیص پلاستیک و فلز عملکرد بالاتری داشته و بیشترین چالش مربوط به شباهت ظاهری میان کاغذ و مقوا و همچنین، شیشه و پلاستیک شفاف است. بر اساس نتایج، روش پیشنهادی می‌تواند در سامانه‌هایی مانند خطوط تفکیک خودکار، سطوح‌های هوشمند و ربات‌های جداساز مورد استفاده قرار گیرد.

استناد: جلوخانی، علی (۱۴۰۵). تشخیص و جداسازی هوشمند پسماندهای شهری قابل بازیافت با استفاده از مدل بینایی ماشین مبتنی بر یادگیری عمیق. *سیاست‌گذاری پیشرفت شهری*، ۳(۳) ۴۱۱-۴۳۱.

DOI: <http://doi.org/10.22034/judpm.2026.582800.1106>

© نویسندگان

DOI: <http://doi.org/10.22034/judpm.2026.582800.1106>



۱. مقدمه

افزایش جمعیت شهرنشین، رشد مصرف‌گرایی و گسترش استفاده از بسته‌بندی‌های یک‌بارمصرف باعث شده است مدیریت پسماندهای شهری به یکی از چالش‌های مهم زیست‌محیطی و اقتصادی در شهرهای امروزی تبدیل شود. بخش قابل توجهی از این پسماندها شامل موادی مانند مقوا، شیشه، فلز، کاغذ و پلاستیک است که در صورت جداسازی صحیح، می‌توانند دوباره وارد چرخه تولید شوند. با این حال، در بسیاری از سامانه‌های مدیریت پسماند، جداسازی مواد قابل بازیافت هنوز به صورت دستی یا نیمه‌خودکار انجام می‌شود و همین موضوع باعث کاهش سرعت پردازش، افزایش هزینه عملیاتی و افت کیفیت مواد بازیافتی می‌شود [۱ و ۲]. پژوهش‌های جدید نیز نشان می‌دهند دسته‌بندی دقیق پسماند یکی از مراحل کلیدی در بهبود بازیافت، کاهش دفن زباله و حرکت به سمت شهرهای هوشمند و پایدار است [۳].

کیفیت بازیافت تا حد زیادی به دقت مرحله تفکیک بستگی دارد؛ زیرا خطای جداسازی می‌تواند ارزش اقتصادی مواد بازیافتی و کارایی مراحل پردازش بعدی را کاهش دهد. به همین دلیل، استفاده از روش‌های هوشمند برای شناسایی و تفکیک پسماندهای قابل بازیافت در سال‌های اخیر مورد توجه قرار گرفته است [۴ و ۵].

اهمیت این مسئله در کلان‌شهری مانند تهران پررنگ‌تر است؛ زیرا حجم بالای تولید پسماند، تراکم جمعیت و هزینه‌های جمع‌آوری و پردازش، مدیریت پسماند را به یکی از چالش‌های اصلی شهری تبدیل کرده است. از آنجا که بخشی از پسماند خشک تهران شامل موادی مانند کاغذ، مقوا، پلاستیک، شیشه و فلز است، بهبود فرایند تشخیص و تفکیک این مواد می‌تواند به افزایش کیفیت بازیافت و کاهش خطای انسانی کمک کند.

در سال‌های اخیر، بینایی ماشین و یادگیری عمیق به عنوان راهکاری مؤثر برای خودکارسازی تشخیص پسماند مورد توجه قرار گرفته‌اند. با وجود این، بسیاری از روش‌های موجود یا بر طبقه‌بندی تک‌برچسبی تصویر تمرکز دارند، یا مدل‌های عمومی تشخیص شیء را بدون توجه کافی به شرایط واقعی پسماندهای قابل بازیافت به کار گرفته‌اند. در محیط واقعی، یک تصویر ممکن است شامل چند نوع پسماند باشد و کیفیت تصویر نیز تحت تأثیر عواملی مانند تغییر روشنایی، هم‌پوشانی اشیاء، پس‌زمینه شلوغ و تغییر شکل مواد قرار گیرد.

در این پژوهش، یک چارچوب مبتنی بر یادگیری عمیق برای تشخیص چندنمونه‌ای پسماندهای شهری قابل بازیافت ارائه می‌شود. روش پیشنهادی برای پنج کلاس مقوا، شیشه، فلز، کاغذ و پلاستیک طراحی شده و برای هر پسماند موجود در تصویر، جعبه محدودکننده، کلاس و امتیاز اطمینان تولید می‌کند. هدف پژوهش، ارائه یک مدل قابل استفاده در سامانه‌های عملی تفکیک پسماند است؛ به گونه‌ای که علاوه بر دقت تشخیص، پایداری در شرایط تصویری دشوار نیز مورد توجه قرار گیرد.

با توجه به بررسی‌های انجام‌شده، بسیاری از پژوهش‌های موجود در حوزه تشخیص پسماند یا بر دسته‌بندی تک‌برچسبی تصویر تمرکز دارند، یا از مدل‌های عمومی تشخیص شیء بدون توجه کافی به ویژگی‌های خاص پسماندهای قابل بازیافت استفاده کرده‌اند. در حالی که در محیط واقعی تفکیک پسماند، یک تصویر ممکن است شامل چند ماده مختلف باشد و برخی کلاس‌ها مانند کاغذ و مقوا یا شیشه و پلاستیک شفاف از نظر ظاهری شباهت زیادی داشته باشند. افزون بر این، شرایط تصویربرداری در محیط‌های واقعی معمولاً کنترل‌شده نیست و عواملی مانند تغییر روشنایی، تاری، نویز، کاهش کنتراست، پس‌زمینه شلوغ و هم‌پوشانی اشیاء می‌توانند عملکرد سامانه‌های تشخیص را تحت تأثیر قرار دهند. بر این اساس، نوآوری‌ها و مشارکت‌های اصلی این پژوهش به صورت زیر قابل بیان است:

۱. ارائه یک چارچوب تشخیص چندنمونه‌ای برای شناسایی هم‌زمان چند نوع پسماند قابل بازیافت در یک تصویر
۲. طراحی معماری سبک و مسئله‌محور بر پایه استخراج ویژگی سبک، توجه ترکیبی و تلفیق ویژگی‌های چندسطحی
۳. ارزیابی پایداری مدل در شرایط دشوار تصویری مانند تغییر روشنایی، نویز، تاری، پس‌زمینه شلوغ و هم‌پوشانی اشیاء
۴. تحلیل کلاس به کلاس، کیفی و آماری نتایج برای بررسی عملکرد مدل فراتر از معیارهای کلی
۵. تولید خروجی قابل استفاده برای سامانه‌های عملی جداسازی، شامل جعبه محدودکننده، نوع ماده و امتیاز اطمینان

پژوهش حاضر از نظر هدف، کاربردی - توسعه‌ای و از نظر رویکرد، کمی و تجربی است. داده‌های پژوهش از دو مسیر گردآوری شدند: نخست، منابع اسنادی و پژوهش‌های پیشین برای تبیین مسئله و بررسی وضعیت مدیریت پسماند؛ دوم، مجموعه داده تصویری عمومی EE297_Project برای آموزش و ارزیابی مدل تشخیص پسماند. جامعه مورد مطالعه شامل تصاویر پسماندهای قابل بازیافت در پنج کلاس مقوا، شیشه، فلز، کاغذ و پلاستیک است و نمونه مورد استفاده شامل ۴۲۵۰ تصویر و ۶۸۱۵ جعبه برچسب‌خورده است. داده‌ها به سه بخش آموزش، اعتبارسنجی و آزمون تقسیم شدند و ارزیابی مدل با معیارهای Precision، Recall، F1-score و $mAP@0.5$ و $mAP@0.5:0.95$ انجام گرفت. همچنین برای افزایش اعتبار نتایج، مقایسه با مدل‌های پایه، تحلیل کلاس به کلاس، تحلیل کیفی، بررسی شرایط دشوار تصویری و آزمون آماری معناداری نتایج انجام شد.

۲. پیشینه پژوهش

مدیریت پسماند شهری در سال‌های اخیر از یک فعالیت خدماتی ساده به بخشی از برنامه‌های توسعه پایدار، اقتصاد چرخشی و شهرهای هوشمند تبدیل شده است. پژوهش‌های مرتبط با سطل‌های هوشمند و سامانه‌های داده‌محور نشان می‌دهند ترکیب حسگرها، اینترنت اشیا و یادگیری عمیق می‌تواند فرایند مدیریت پسماند را از حالت سنتی به سمت سامانه‌های هوشمند و قابل پایش هدایت کند [۶]. همچنین، مطالعات مرتبط با دسته‌بندی پایدار پسماند نشان داده‌اند بازیافت مؤثر منابع، به‌ویژه در محیط‌های شهری، به روش‌های خودکار و دقیق برای شناسایی و تفکیک مواد نیاز دارد [۱ و ۳].

بخش مهمی از پژوهش‌های اولیه در این حوزه بر دسته‌بندی تصویر متمرکز بوده‌اند. در این رویکرد، کل تصویر به یک کلاس مانند کاغذ، مقوا، شیشه، فلز یا پلاستیک نسبت داده می‌شود. معرفی مجموعه داده‌هایی مانند ترشنت، امکان ارزیابی روش‌های یادگیری ماشین و شبکه‌های عمیق را برای دسته‌بندی مواد قابل بازیافت فراهم کرد [۷]. در ادامه، مدل‌های مبتنی بر شبکه‌های کانولوشنی و یادگیری عمیق برای افزایش دقت دسته‌بندی پسماند به کار گرفته شدند [۸]. پژوهش‌هایی مانند مطالعات ساید و همکاران [۱]، سایم و همکاران [۲] و لیو و همکاران [۹] نیز نشان داده‌اند یادگیری عمیق می‌تواند در تشخیص و دسته‌بندی پسماندهای قابل بازیافت عملکرد قابل قبولی داشته باشد. با این حال، ماهیت تک‌برچسبی بسیاری از این روش‌ها باعث می‌شود در شرایطی که چند نوع پسماند هم‌زمان در تصویر وجود دارد، کارایی آن‌ها محدود شود.

در کنار دسته‌بندی تصویر، تشخیص شیء به دلیل توانایی در تعیین هم‌زمان کلاس و موقعیت پسماند، برای کاربردهای عملی تفکیک اهمیت بیشتری پیدا کرده است. مدل‌های عمومی تشخیص شیء مانند فستر آر - سی‌ان‌ان، اس‌اس‌دی، یولو و رتینانت، پایه بسیاری از پژوهش‌های جدید در این حوزه را تشکیل داده‌اند [۱۰-۱۳]. در کاربردهای مرتبط با پسماند نیز زایلان و همکاران [۱۴] و رن و همکاران [۱۵] نشان داده‌اند روش‌های تشخیص شیء می‌توانند برای صحنه‌های دارای پس‌زمینه پیچیده و چند شیء مناسب‌تر از طبقه‌بندی ساده باشند. همچنین وو و همکاران [۱۶] و یانگ و همکاران [۱۷] بر اهمیت طراحی مدل‌های سبک‌تر برای کاربردهای نزدیک به بالادرنگ، سطل‌های هوشمند، خطوط تفکیک و سامانه‌های رباتیک تأکید کرده‌اند. این رویکرد با هدف پژوهش حاضر هم‌راستا است؛ زیرا مدل پیشنهادی باید ضمن حفظ دقت، از نظر هزینه محاسباتی نیز برای کاربرد عملی قابل استفاده باشد.

برخی مطالعات برای افزایش دقت تشخیص جنس ماده، از داده‌های حسی یا تصویربرداری‌های تخصصی استفاده کرده‌اند. برای نمونه، روتس و همکاران [۱۸] از ترکیب حسگرها و یادگیری عمیق برای جداسازی پسماندهای بسته‌بندی پلاستیکی استفاده کردند. چن و همکاران [۱۹] از تصاویر پرتو ایکس برای تشخیص و دسته‌بندی پسماند بهره گرفتند و پژوهش‌های مبتنی بر تصویربرداری فراطیفی نیز نشان داده‌اند داده‌های طیفی می‌توانند اطلاعات بیشتری درباره جنس مواد فراهم کنند [۲۰]. با این حال، این روش‌ها معمولاً به تجهیزات گران‌تر یا شرایط تصویربرداری خاص نیاز دارند. از سوی دیگر، روش‌های بخش‌بندی تصویری مانند نونی - ویست می‌توانند مرز دقیق‌تری از پسماند ارائه دهند، اما به داده‌های برچسب‌خورده دقیق‌تر و هزینه محاسباتی بالاتر وابسته‌اند [۵]. بنابراین، در بسیاری از کاربردهای شهری، تشخیص شیء بر پایه تصاویر معمولی می‌تواند تعادلی مناسب میان دقت، هزینه و قابلیت استقرار ایجاد کند.

از نظر داده و شرایط تصویربرداری، یکی از چالش‌های مهم این حوزه فاصله میان داده‌های آزمایشگاهی و محیط واقعی است. بسیاری از مجموعه داده‌ها تصاویر نسبتاً ساده و کنترل شده دارند، در حالی که پسماند در محیط واقعی ممکن است تغییر شکل یافته، نیمه‌پنهان، دارای آلودگی، هم‌پوشانی یا پس‌زمینه پیچیده باشد. لین و همکاران [۴] و سینگل و همکاران [۲۱] بر اهمیت استفاده از داده‌های واقعی‌تر و ارزیابی مدل‌ها در شرایط نزدیک به محیط عملی تأکید کرده‌اند. از سوی دیگر، مجموعه داده‌هایی مانند تاکو [۲۲]، ترش‌کن [۲۳] و ریل‌ویست [۲۱] تلاش کرده‌اند تنوع محیطی و واقع‌گرایی بیشتری را در داده‌های پسماند فراهم کنند. با وجود این، چالش‌هایی مانند عدم تعادل کلاس‌ها، کمبود نمونه برای برخی مواد و تفاوت زیاد میان شرایط تصویربرداری همچنان باقی است.

پژوهش‌های جدیدتر نشان می‌دهند کاربرد یادگیری عمیق در مدیریت پسماند تنها به پسماندهای شهری عمومی محدود نیست. برای نمونه، روش‌های یادگیری عمیق در جداسازی ضایعات فلزی هم‌جنس اما متفاوت [۲۴] و همچنین، تشخیص پسماندهای ساختمانی یا تزئینی [۲۵] نیز مورد استفاده قرار گرفته‌اند. این مطالعات از نظر دامنه کاربرد با پسماندهای شهری قابل بازیافت تفاوت دارند، اما نشان می‌دهند روند کلی پژوهش‌ها به سمت سامانه‌های دقیق‌تر، کاربردی‌تر و قابل استفاده در محیط واقعی حرکت کرده است.

در سطح معماری، بسیاری از پژوهش‌های پسماند از شبکه‌های شناخته شده بینایی ماشین به عنوان پایه استفاده کرده‌اند. معماری‌هایی مانند رس‌نت، موبایل‌نت و ایفیشنت به ترتیب برای آموزش شبکه‌های عمیق، اجرای سبک روی تجهیزات محدود و ایجاد تعادل میان دقت و هزینه محاسباتی معرفی شده‌اند [۲۶-۲۸]. اگرچه این معماری‌ها به طور اختصاصی برای پسماند طراحی نشده‌اند، اما ایده‌های آن‌ها در بسیاری از مدل‌های دسته‌بندی و تشخیص پسماند مورد استفاده قرار گرفته است. همین موضوع نشان می‌دهد مسئله اصلی در پژوهش‌های جدید، تنها انتخاب یک شبکه عمومی نیست، بلکه سازگار کردن معماری با ویژگی‌های مسئله، شرایط داده و نیاز کاربردی است.

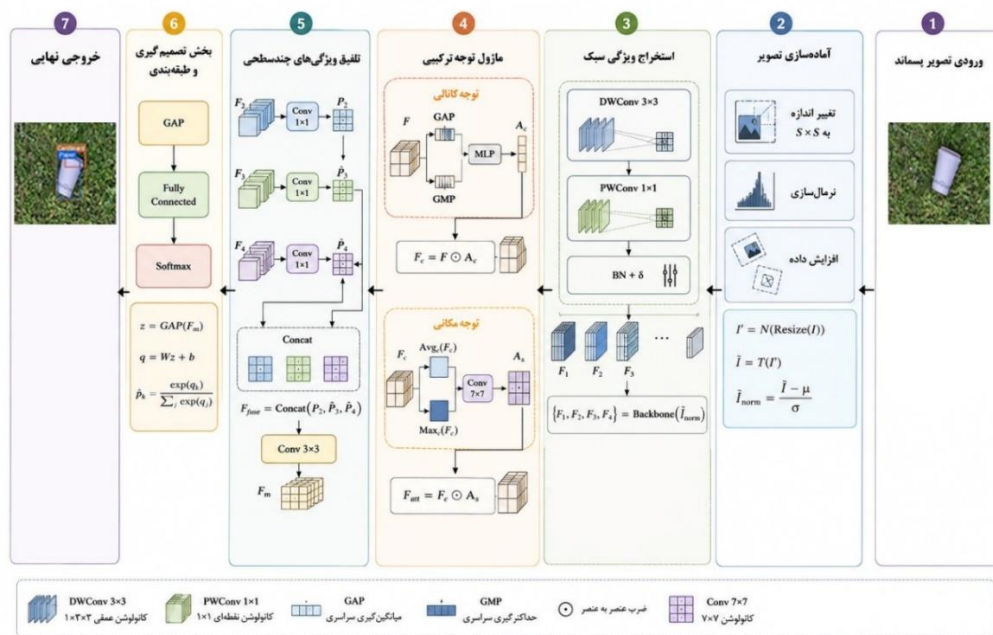
بررسی مطالعات پیشین نشان می‌دهد با وجود پیشرفت‌های قابل توجه، چند محدودیت همچنان باقی است. نخست، بخشی از پژوهش‌ها بر طبقه‌بندی تک‌برچسبی تصویر متمرکزند و برای تصاویر چندشبه کافی نیستند. دوم، بسیاری از مدل‌های تشخیص شیء بدون تحلیل کافی از شرایط واقعی پسماند، مانند تغییر شکل مواد، پس‌زمینه شلوغ، هم‌پوشانی و شرایط نامناسب تصویربرداری استفاده شده‌اند. سوم، ارزیابی‌ها در بسیاری از موارد به معیارهای کلی محدود شده و تحلیل کلاس به کلاس، تحلیل خطا، شرایط دشوار تصویری و معناداری آماری کمتر مورد توجه قرار گرفته است. بر این اساس، پژوهش حاضر با تمرکز بر تشخیص چندنمونه‌ای، معماری سبک و مسئله‌محور، ارزیابی در شرایط دشوار و تحلیل آماری نتایج تلاش می‌کند بخشی از این شکاف‌ها را پوشش دهد.

۳. روش پیشنهادی

در این پژوهش، یک چارچوب مبتنی بر بینایی ماشین و یادگیری عمیق برای تشخیص و جداسازی هوشمند پسماندهای شهری قابل بازیافت ارائه می‌شود. هدف روش پیشنهادی، دریافت تصویر پسماند و تشخیص چندنمونه‌ای مواد قابل بازیافت در پنج گروه اصلی شامل مقوا، شیشه، فلز، کاغذ و پلاستیک است. در این چارچوب، برای هر پسماند موجود در تصویر، جعبه محدودکننده، کلاس و امتیاز اطمینان تولید می‌شود.

۳-۱. طراحی مسئله محور معماری پیشنهادی

معماری پیشنهادی از پنج بخش اصلی تشکیل شده است: آماده‌سازی و افزایش داده، استخراج ویژگی سبک، ماژول توجه ترکیبی، تلفیق ویژگی‌های چندسطحی، و بخش تصمیم‌گیری نهایی. ابتداء، تصویر ورودی به اندازه ثابت تبدیل و نرمال‌سازی می‌شود. سپس تصویر وارد استخراج‌گر ویژگی می‌شود تا نمایش‌های اولیه و عمیق از آن استخراج شود. در ادامه، ماژول توجه ترکیبی، بخش‌های مهم تصویر را تقویت می‌کند و اثر پس‌زمینه یا نواحی غیرمرتبط را کاهش می‌دهد. سپس، ویژگی‌های چندسطحی با یکدیگر ادغام می‌شوند تا هم جزئیات محلی مانند بافت و لبه، و هم اطلاعات کلی مانند شکل و ساختار شیء در تصمیم‌گیری نهایی استفاده شوند. در نهایت، بخش طبقه‌بندی نوع پسماند را همراه با امتیاز اطمینان پیش‌بینی می‌کند.



شکل ۱. معماری کلی روش پیشنهادی برای تشخیص و جداسازی هوشمند پسماندهای قابل بازیافت شامل مراحل آماده‌سازی تصویر، استخراج ویژگی سبک، توجه ترکیبی، تلفیق ویژگی‌های چندسطحی و تشخیص نهایی نوع پسماند

ورودی سامانه یک تصویر رنگی از پسماند است که به صورت رابطه ۱ تعریف می‌شود:

$$I \in R^{H \times W \times 3} \quad (1)$$

در این رابطه، H و W به ترتیب ارتفاع و عرض تصویر هستند و عدد ۳ نشان‌دهنده سه کانال رنگی تصویر است. پس از آماده‌سازی، تصویر به اندازه ثابت $S \times S$ تبدیل شده و مقادیر پیکسلی آن نرمال می‌شوند (رابطه ۲):

$$I' = N(\text{Resize}(I)) \quad (2)$$

در این رابطه، $\text{Resize}(\cdot)$ عملیات تغییر اندازه تصویر و $N(\cdot)$ عملیات نرمال‌سازی را نشان می‌دهد. هدف از این مرحله، کاهش اثر تفاوت اندازه تصاویر، شدت روشنایی و شرایط مختلف تصویربرداری است.

۳-۲. آماده‌سازی و افزایش داده

در محیط واقعی، پسماندها معمولاً در وضعیت‌های کاملاً منظم و ثابت قرار ندارند. ممکن است یک بطری پلاستیکی فشرده شده باشد، تکه‌ای کاغذ چروک یا پاره شده باشد، شیشه نور را بازتاب دهد، یا مقوا بخشی از سطح خود را از دست داده باشد. بنابراین، سامانه نباید فقط روی تصاویر تمیز و منظم عملکرد مناسبی داشته باشد، بلکه باید در برابر تغییرات رایج در محیط واقعی نیز پایدار باشد.

برای افزایش توان تعمیم روش پیشنهادی، روی داده‌های آموزشی عملیات افزایش داده انجام می‌شود. این عملیات شامل چرخش محدود، تغییر روشنایی، تغییر کنتراست، تغییر مقیاس، برش تصادفی، جابه‌جایی مکانی و وارون‌سازی افقی است. تصویر حاصل از افزایش داده به صورت رابطه ۳ تعریف می‌شود:

$$\tilde{I} = T(I') \quad (3)$$

در رابطه ۳، $T(\cdot)$ مجموعه‌ای از تبدیلات تصویری است که به صورت تصادفی روی تصویر اعمال می‌شود. استفاده از این تبدیلات باعث می‌شود روش پیشنهادی نسبت به تغییر زاویه دید، نور محیط، فاصله دوربین و تغییر شکل پسماند مقاوم‌تر شود. همچنین، برای کاهش اثر اختلاف شدت روشنایی، مقادیر پیکسلی تصویر در بازه مشخص قرار می‌گیرند (رابطه ۴):

$$\tilde{I}_{norm} = \frac{\tilde{I} - \mu}{\sigma} \quad (4)$$

در رابطه ۴، μ و σ به ترتیب میانگین و انحراف معیار مقادیر پیکسلی هستند. این کار باعث می شود فرایند آموزش پایدارتر شود و روش پیشنهادی کمتر به شدت روشنایی خام تصویر وابسته باشد.

۳-۳. استخراج ویژگی سبک برای تشخیص پسماندهای متنوع

در این پژوهش، به جای استفاده از استخراج گرهای بسیار سنگین، از یک استخراج گر ویژگی سبک مبتنی بر کانولوشن های عمقی و نقطه ای استفاده می شود. دلیل این انتخاب آن است که سامانه جداسازی پسماند در کاربرد عملی باید بتواند با سرعت مناسب عمل کند. اگر معماری بیش از حد سنگین باشد، استفاده از آن در سطوح های هوشمند، خطوط بازیافت یا سامانه های رباتیک دشوار می شود.

اگر F_{in} نقشه ویژگی ورودی باشد، خروجی کانولوشن عمقی به صورت رابطه ۵ تعریف می شود:

$$F_d = DWConv_{3 \times 3}(F_{in}) \quad (5)$$

سپس با استفاده از کانولوشن نقطه ای، ترکیب کانالی انجام می شود (رابطه ۶):

$$F_p = PWConv_{1 \times 1}(F_d) \quad (6)$$

در این روابط، $DWConv$ کانولوشن عمقی و $PWConv$ کانولوشن نقطه ای را نشان می دهد. خروجی نهایی این بلوک به صورت رابطه ۷ نوشته می شود:

$$F_{out} = \delta(BN(F_p)) \quad (7)$$

در این رابطه، BN نرمال سازی دسته ای و δ تابع فعال سازی است.

خروجی مراحل مختلف استخراج گر ویژگی به صورت مجموعه ای از نقشه های ویژگی تعریف می شود (رابطه ۸):

$$\{F_1, F_2, F_3, F_4\} = Backbone(\tilde{I}_{norm}) \quad (8)$$

استفاده از چند سطح ویژگی باعث می شود مدل هم از جزئیات محلی و هم از نمایش های معنایی عمیق تر استفاده کند. این موضوع برای تشخیص اشیای کوچک، نمونه های تغییر شکل یافته و تصاویر چندشیء اهمیت دارد.

۳-۴. توجه ترکیبی برای تفکیک کلاس های مشابه

برای کاهش اثر نواحی غیرمرتبط و تقویت ویژگی های مؤثر در تشخیص، از ماژول توجه ترکیبی استفاده شد. این ماژول شامل توجه کانالی و توجه مکانی است و به ترتیب اهمیت کانال های ویژگی و نواحی مکانی تصویر را تنظیم می کند.

برای یک نقشه ویژگی F ، ابتدا میانگین گیری سراسری و بیشینه گیری سراسری انجام می شود (روابط ۹ و ۱۰):

$$F_{avg} = GAP(F) \quad (9)$$

$$F_{max} = GMP(F) \quad (10)$$

سپس، این دو بردار از یک لایه مشترک عبور می کند و وزن توجه کانالی تولید می شود (رابطه ۱۱):

$$A_c = \sigma(MLP(F_{avg}) + MLP(F_{max})) \quad (11)$$

در رابطه ۱۱، GAP میانگین گیری سراسری، GMP بیشینه گیری سراسری، MLP شبکه چندلایه کوچک و σ تابع سیگموئید است.

توجه کانالی روی ویژگی ورودی اعمال می شود (رابطه ۱۲):

$$F_c = F \odot A_c \quad (12)$$

در مرحله بعد، توجه مکانی محاسبه می شود. برای این کار، میانگین و بیشینه ویژگی ها در راستای کانال ها محاسبه شده و

سپس با یک کانولوشن ترکیب می شوند (رابطه ۱۳):

$$A_s = \sigma(Conv_{7 \times 7}([\text{Avg}_c(F_c); \text{Max}_c(F_c)])) \quad (13)$$

در نهایت، خروجی ماژول توجه ترکیبی به صورت رابطه ۱۴ به دست می آید:

$$F_{att} = F_c \odot A_s \quad (۱۴)$$

در این رابطه، $\odot$ ضرب عنصر به عنصر را نشان می‌دهد. استفاده از این ماژول باعث می‌شود سامانه به جای وابستگی کامل به ظاهر کلی تصویر، روی نشانه‌های مهم‌تر هر ماده تمرکز کند. این موضوع به‌ویژه برای تفکیک کلاس‌های مشابه مانند کاغذ و مقوا یا شیشه و پلاستیک اهمیت دارد.

۳-۵. تلفیق چندسطحی برای تشخیص اشیای کوچک و چندشئی

از آنجا که اندازه و مقیاس پسماندها در تصاویر متفاوت است، استفاده از تنها یک سطح ویژگی برای تشخیص کافی نیست. در این بخش، ویژگی‌های سطوح مختلف شبکه با یکدیگر تلفیق می‌شوند تا مدل بتواند هم اشیای کوچک و هم نمونه‌های بزرگ‌تر را شناسایی کند.

در روش پیشنهادی، ویژگی‌های استخراج‌شده از سطوح مختلف شبکه ابتدا به ابعاد کانالی یکسان نگاشت می‌شوند (رابطه ۱۵):

$$P_i = \text{Conv}_{1 \times 1}(F_i), \quad i \in \{2, 3, 4\} \quad (۱۵)$$

سپس برای هم‌اندازه‌سازی، ویژگی‌های عمیق‌تر با روش افزایش اندازه به مقیاس مناسب تبدیل می‌شوند (رابطه ۱۶):

$$\hat{P}_i = \text{Upsample}(P_i) \quad (۱۶)$$

در ادامه، ویژگی‌های چندسطحی با یکدیگر ترکیب می‌شوند (رابطه ۱۷):

$$F_{fuse} = \text{Concat}(P_2, \hat{P}_3, \hat{P}_4) \quad (۱۷)$$

پس از الحاق ویژگی‌ها، یک کانولوشن 3×3 برای کاهش نویز و ترکیب بهتر اطلاعات اعمال می‌شود (رابطه ۱۸):

$$F_m = \text{Conv}_{3 \times 3}(F_{fuse}) \quad (۱۸)$$

این بخش باعث می‌شود هم اطلاعات محلی مانند بافت، لبه و جزئیات سطحی، و هم اطلاعات عمیق‌تر مانند شکل کلی و ساختار شیء در تصمیم‌گیری نهایی استفاده شوند. چنین تلفیقی برای مسئله پسماند ضروری است؛ زیرا برخی کلاس‌ها با نشانه‌های سطحی و برخی دیگر با نشانه‌های ساختاری بهتر تشخیص داده می‌شوند.

۳-۶. خروجی عملیاتی برای هدایت پسماند به مسیر جداسازی

پس از استخراج و تلفیق ویژگی‌ها، بخش تشخیص نهایی برای هر پسماند موجود در تصویر، مختصات جعبه محدودکننده، کلاس و امتیاز اطمینان را تولید می‌کند. بنابراین، خروجی مدل به صورت مجموعه‌ای از پیش‌بینی‌ها تعریف می‌شود (رابطه ۱۹):

$$\text{Output} = \{(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i, \hat{c}_i, \hat{s}_i)\}_{i=1}^N \quad (۱۹)$$

در این رابطه، N تعداد پسماندهای تشخیص‌داده‌شده در تصویر است. همچنین، $\hat{x}_i$ و $\hat{y}_i$ مختصات مرکز جعبه محدودکننده، $\hat{w}_i$ و $\hat{h}_i$ عرض و ارتفاع جعبه، $\hat{c}_i$ کلاس پیش‌بینی‌شده، و $\hat{s}_i$ امتیاز اطمینان مربوط به پسماند شماره i هستند.

برای تصمیم‌گیری عملی، تنها پیش‌بینی‌هایی پذیرفته می‌شوند که امتیاز اطمینان آن‌ها از آستانه تعیین‌شده بیشتر باشد (روابط ۲۰ و ۲۱):

$$\text{Decision}_i = \text{Accept} \quad \text{if} \quad \hat{s}_i \geq \tau \quad (۲۰)$$

$$\text{Decision}_i = \text{Review} \quad \text{if} \quad \hat{s}_i < \tau \quad (۲۱)$$

در این رابطه، τ آستانه اطمینان است. این خروجی می‌تواند در مراحل بعدی برای هدایت هر پسماند به مسیر جداسازی متناظر استفاده شود.

۳-۷. تابع زیان

از آنجا که مسئله پژوهش از نوع تشخیص شیء است، تابع زیان مدل از سه بخش اصلی تشکیل می‌شود: زیان مکان‌یابی جعبه، زیان طبقه‌بندی و زیان اطمینان وجود شیء. تابع زیان کلی به صورت رابطه ۲۲ تعریف می‌شود:

$$L_{total} = \lambda_{box} L_{box} + \lambda_{cls} L_{cls} + \lambda_{obj} L_{obj} \quad (۲۲)$$

در این رابطه، L_{box} خطای مکان‌یابی جعبه محدودکننده، L_{cls} خطای پیش‌بینی کلاس پسماند و L_{obj} خطای تشخیص وجود شیء را نشان می‌دهد. ضرایب λ_{cls} و λ_{obj} نیز برای کنترل سهم هر بخش از تابع زیان استفاده می‌شوند. این ساختار باعث می‌شود مدل به صورت هم‌زمان نوع پسماند، موقعیت آن و اطمینان تشخیص را یاد بگیرد.

۴. آزمایش‌ها و نتایج

۴-۱. مجموعه داده و آماده‌سازی تصاویر

برای آموزش و ارزیابی روش پیشنهادی، از مجموعه داده عمومی EE297_Project استفاده شد. این مجموعه داده در پلتفرم Roboflow Universe منتشر شده است. نسخه مورد استفاده در این پژوهش با عنوان ۲۰۲۲-۰۴-۱۲-۱۵۵-final-pm منتشر شده و تحت مجوز CC BY 4.0 ارائه شده است. استفاده از یک مجموعه داده عمومی و دارای مجوز مشخص باعث می‌شود امکان بررسی، دانلود و بازتولید آزمایش‌ها برای سایر پژوهشگران فراهم باشد.

این مجموعه داده برای مسئله تشخیص شیء تهیه شده و شامل تصاویر پسماندهای قابل بازیافت در پنج کلاس اصلی مقوا، شیشه، فلز، کاغذ و پلاستیک است. به خلاف مجموعه داده‌های ساده طبقه‌بندی تصویر که در آن هر تصویر تنها یک برچسب دارد، در این مجموعه داده هر تصویر می‌تواند شامل یک یا چند پسماند باشد و برای هر شیء موجود در تصویر، یک جعبه محدودکننده و برچسب کلاس در نظر گرفته شده است. بنابراین، این داده برای ارزیابی روش‌های تشخیص چندنمونه‌ای پسماند مناسب است.

در این پژوهش، مجموعه داده خام شامل ۴۲۵۰ تصویر و ۶۸۱۵ نمونه برچسب‌خورده در قالب جعبه محدودکننده بود. از میان تصاویر، ۲۵۲۰ تصویر شامل یک نوع پسماند و ۱۷۳۰ تصویر شامل دو یا چند نوع پسماند بودند. این ترکیب باعث شد مدل هم با نمونه‌های ساده تک‌شیء و هم با نمونه‌های پیچیده‌تر چندشیء آموزش ببیند. تعداد ۶۸۱۵ نمونه برچسب‌خورده مربوط به داده‌های خام پیش از اعمال افزایش داده است. افزایش داده تنها روی بخش آموزش اعمال شد و داده‌های اعتبارسنجی و آزمون بدون افزایش داده برای ارزیابی مستقل باقی ماندند.

جدول ۱. توزیع تصاویر و نمونه‌های برچسب‌خورده در مجموعه داده

کلاس	تعداد جعبه‌های برچسب‌خورده
مقوا	۱۲۸۰
شیشه	۱۳۶۵
فلز	۱۱۹۰
کاغذ	۱۴۵۵
پلاستیک	۱۵۲۵
مجموع	۶۸۱۵

برای افزایش توان تعمیم مدل، افزایش داده تنها روی بخش آموزش اعمال شد. این کار با هدف شبیه‌سازی تغییرات رایج در محیط واقعی، مانند تغییر زاویه دید، تغییر شدت نور، تغییر کنتراست، تغییر مقیاس، جابه‌جایی مکانی و وارون‌سازی افقی انجام شد. با اعمال پنج حالت افزایش داده روی تصاویر آموزشی، تعداد مؤثر تصاویر آموزشی از ۲۹۷۵ به ۱۴۸۷۵ و تعداد مؤثر نمونه‌های آموزشی از ۴۷۷۳ به ۲۳۸۶۵ افزایش یافت. با این حال، داده‌های اعتبارسنجی و آزمون بدون افزایش داده باقی ماندند تا ارزیابی نهایی منصفانه و مستقل باشد.

جدول ۲. تعداد نمونه‌های خام و نمونه‌های مؤثر پس از افزایش داده در فرایند آموزش

بخش داده	تعداد تصاویر خام	تعداد جعبه‌های خام	ضریب افزایش داده	تعداد تصاویر مؤثر	تعداد جعبه‌های مؤثر
آموزش	۲۹۷۵	۴۷۷۳	۵ برابر	۱۴۸۷۵	۲۳۸۶۵
اعتبارسنجی	۶۳۸	۱۰۲۲	بدون افزایش داده	۶۳۸	۱۰۲۲
آزمون	۶۳۷	۱۰۲۰	بدون افزایش داده	۶۳۷	۱۰۲۰

مجموعه داده به سه بخش آموزش، اعتبارسنجی و آزمون تقسیم شد. نسبت تقسیم داده ها به ترتیب ۷۰ درصد آموزش، ۱۵ درصد اعتبارسنجی و ۱۵ درصد آزمون در نظر گرفته شد. تقسیم بندی به گونه ای انجام شد که تصاویر تک شیء و چندشیء در هر سه بخش وجود داشته باشند.

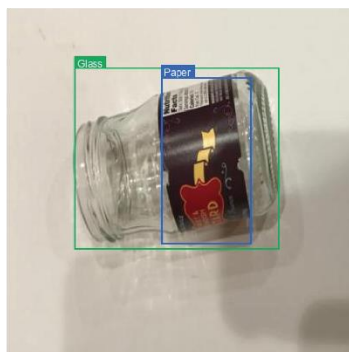
جدول ۳. تقسیم بندی مجموعه داده به بخش های آموزش، اعتبارسنجی و آزمون

بخش داده	تعداد تصاویر	تعداد جعبه های برچسب خورده
آموزش	۲۹۷۵	۴۷۷۳
اعتبارسنجی	۶۳۸	۱۰۲۲
آزمون	۶۳۷	۱۰۲۰
مجموع	۴۲۵۰	۶۸۱۵

برای افزایش اعتمادپذیری داده ها، برچسب ها پیش از آموزش از نظر کلاس نامعتبر، مختصات خارج از محدوده، جعبه های ناقص یا بسیار کوچک و نمونه های مبهم بررسی شدند. موارد دارای خطای ساختاری اصلاح و نمونه های دارای ابهام شدید حذف شدند. خلاصه فرایند کنترل کیفیت در جدول ۴ آمده است.

جدول ۴. خلاصه کنترل کیفیت و اصلاح برچسب ها

اقدام انجام شده	تعداد موارد شناسایی شده	مورد بررسی
حذف یا اصلاح فایل	۲۷	فایل برچسب خالی یا ناقص
اصلاح شناسه کلاس	۱۱	کلاس نامعتبر در فایل برچسب
اصلاح مختصات جعبه	۳۴	جعبه خارج از محدوده تصویر
حذف یا اصلاح جعبه	۴۲	جعبه بسیار کوچک یا غیرواقعی
بازتنظیم جعبه محدودکننده	۸۶	عدم انطباق جعبه با شیء
حذف نمونه از مجموعه نهایی	۳۱	ابهام شدید در نوع پسماند



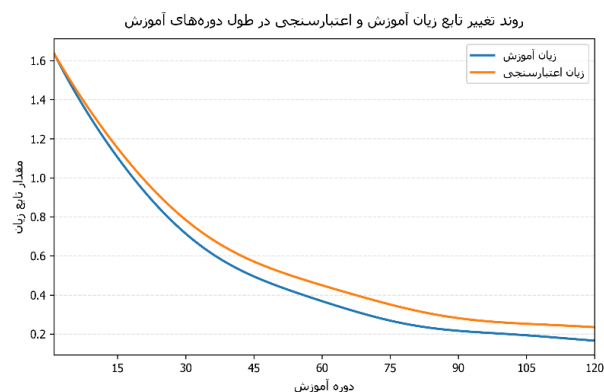
شکل ۲. نمونه هایی از تصاویر برچسب خورده همراه با جعبه های محدودکننده برای چند پسماند در یک تصویر

۲-۴. تنظیمات پیاده‌سازی و آموزش

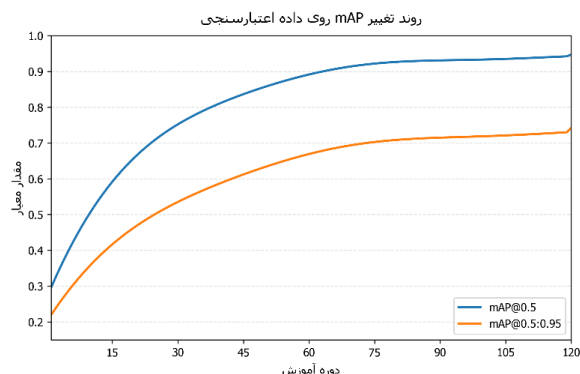
پیاده‌سازی روش پیشنهادی در محیط پایتون و با استفاده از چارچوب یادگیری عمیق انجام شد. تصاویر ورودی با اندازه 640×640 وارد شبکه شدند. اندازه دسته برابر با ۱۶ و تعداد دوره‌های آموزش برابر با ۱۲۰ در نظر گرفته شد. برای بهینه‌سازی وزن‌ها، از بهینه‌ساز AdamW با نرخ یادگیری اولیه 0.001 استفاده شد. بهترین وزن‌های مدل بر اساس مقدار $mAP@0.5$ روی داده اعتبارسنجی ذخیره شدند و ارزیابی نهایی روی مجموعه آزمون انجام گرفت.

برای کاهش اثر تصادفی بودن مقاداردهی اولیه وزن‌ها و ترتیب نمونه‌های آموزشی، هر مدل در پنج اجرای مستقل با بذره‌های تصادفی متفاوت آموزش و ارزیابی شد. در هر اجرا، تقسیم‌بندی آموزش، اعتبارسنجی و آزمون ثابت نگه داشته شد و تنها مقاداردهی اولیه و ترتیب دسته‌های آموزشی تغییر کرد. نتایج نهایی علاوه بر مقدار میانگین، با انحراف معیار گزارش شدند. همچنین، برای بررسی معناداری آماری تفاوت عملکرد روش پیشنهادی با مدل‌های مقایسه‌ای، آزمون t زوجی روی مقادیر $mAP@0.5$ و $mAP@0.5:0.95$ در پنج اجرای مستقل انجام شد. سطح معناداری برابر با 0.05 در نظر گرفته شد.

طی آموزش، مقدار تابع زیان کاهش تدریجی داشت و پس از حدود دوره ۹۰ به حالت پایدار نزدیک شد. همچنین، مقدار mAP روی داده اعتبارسنجی در دوره‌های ابتدایی رشد سریع‌تری داشت و سپس با شیب کمتر بهبود یافت. این روند نشان می‌دهد مدل توانسته است هم ویژگی‌های ظاهری مواد و هم موقعیت مکانی آن‌ها را در تصویر یاد بگیرد.



شکل ۴. روند تغییر تابع زیان آموزش و اعتبارسنجی طی دوره‌های آموزش



شکل ۳. روند تغییر $mAP@0.5$ و $mAP@0.5:0.95$ روی داده اعتبارسنجی

۳-۴. معیارهای ارزیابی

با توجه به ماهیت تشخیص شیء، عملکرد مدل با معیارهای Precision، Recall، F1-score، $mAP@0.5$ و $mAP@0.5:0.95$ ارزیابی شد. معیار $mAP@0.5$ عملکرد مدل را در آستانه هم‌پوشانی 0.5 نشان می‌دهد، در حالی که $mAP@0.5:0.95$ ارزیابی سخت‌گیرانه‌تری از کیفیت مکان‌یابی جعبه‌ها ارائه می‌کند. علاوه بر نتایج کلی، عملکرد مدل به صورت کلاس به کلاس، کیفی، آماری و در شرایط تصویری دشوار بررسی شد.

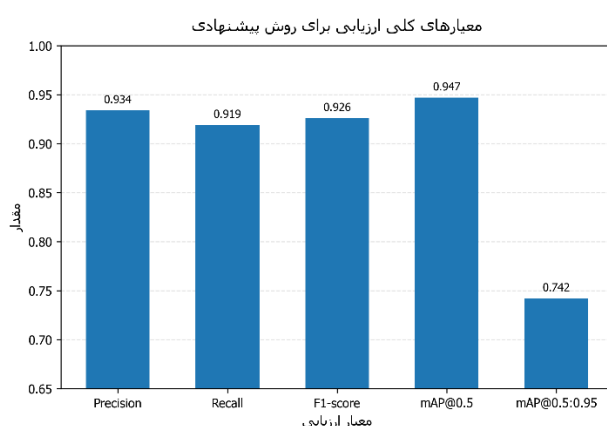
۴-۴. نتایج کلی روش پیشنهادی

نتایج کلی روش پیشنهادی روی مجموعه آزمون در جدول ۵ ارائه شده است. همان‌طور که مشاهده می‌شود، روش پیشنهادی توانسته است مقدار $mAP@0.5$ برابر با 0.947 و مقدار $mAP@0.5:0.95$ برابر با 0.742 به دست آورد. این نتایج نشان می‌دهد مدل علاوه بر تشخیص درست نوع پسماند، در مکان‌یابی آن‌ها نیز عملکرد مناسبی دارد.

جدول ۵. نتایج ارزیابی مدل پیشنهادی روی مجموعه آزمون

مقدار	معیار ارزیابی
۰.۹۳۴	Precision
۰.۹۱۹	Recall
۰.۹۲۶	F1-score
۰.۹۴۷	mAP@0.5
۰.۷۴۲	mAP@0.5:0.95

اختلاف محدود میان Precision و Recall نشان می‌دهد مدل در کاهش تشخیص‌های اشتباه و همچنین، در شناسایی نمونه‌های واقعی تعادل مناسبی برقرار کرده است. این موضوع برای سامانه‌های تفکیک پسماند اهمیت دارد؛ زیرا هم تشخیص اشتباه و هم از دست رفتن یک پسماند قابل بازیافت می‌تواند کیفیت جداسازی را کاهش دهد.



شکل ۵. نمودار میله‌ای معیارهای Precision، Recall، F1-score، mAP@0.5 و mAP@0.5:0.95 برای روش پیشنهادی

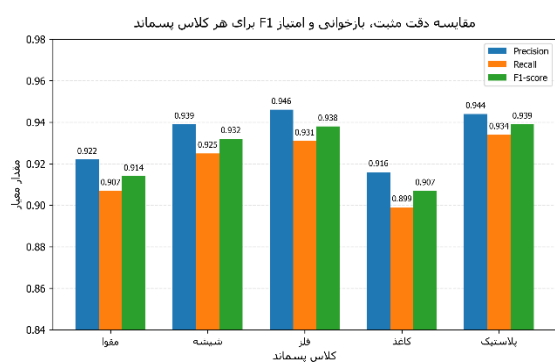
۴-۵. عملکرد روش پیشنهادی به تفکیک کلاس‌ها

برای تحلیل دقیق‌تر، عملکرد روش پیشنهادی برای هر یک از پنج کلاس پسماند در جدول ۶ ارائه شده است. این جدول نشان می‌دهد مدل در تشخیص هر نوع پسماند چه عملکردی داشته و کدام کلاس‌ها چالش بیشتری ایجاد کرده‌اند.

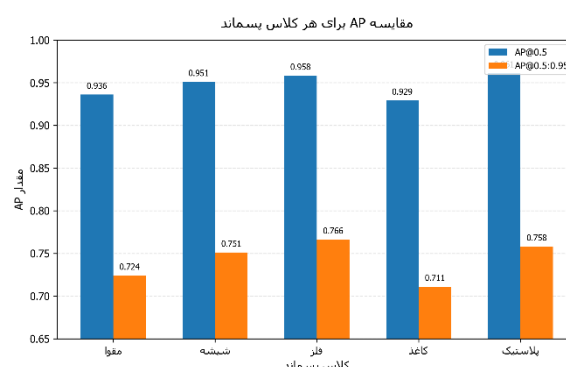
جدول ۶. عملکرد روش پیشنهادی به تفکیک کلاس‌های پسماند

کلاس	Precision	Recall	F1-score	AP@0.5	AP@0.5:0.95
مقوا	۰.۹۲۲	۰.۹۰۷	۰.۹۱۴	۰.۹۳۶	۰.۷۲۴
شیشه	۰.۹۳۹	۰.۹۲۵	۰.۹۳۲	۰.۹۵۱	۰.۷۵۱
فلز	۰.۹۴۶	۰.۹۳۱	۰.۹۳۸	۰.۹۵۸	۰.۷۶۶
کاغذ	۰.۹۱۶	۰.۸۹۹	۰.۹۰۷	۰.۹۲۹	۰.۷۱۱
پلاستیک	۰.۹۴۴	۰.۹۳۴	۰.۹۳۹	۰.۹۶۱	۰.۷۵۸
میانگین	۰.۹۳۴	۰.۹۱۹	۰.۹۲۶	۰.۹۴۷	۰.۷۴۲

نتایج جدول ۶ نشان می‌دهد اختلاف عملکرد میان کلاس‌ها محدود اما قابل تفسیر است. بیشترین $AP@0.5$ مربوط به پلاستیک و فلز و کمترین F1-score مربوط به کاغذ است. همچنین کاهش محسوس‌تر $AP@0.5:0.95$ نسبت به $AP@0.5$ نشان می‌دهد بخشی از خطاها به دقت مکان‌یابی جعبه‌ها مربوط است، نه فقط تشخیص کلاس. بنابراین، تحلیل کلاس به کلاس نشان می‌دهد ارزیابی مدل تنها با معیار کلی mAP کافی نیست.



شکل ۷. نمودار مقایسه Precision، Recall و F1-score برای کلاس های مقوا، شیشه، فلز، کاغذ و پلاستیک

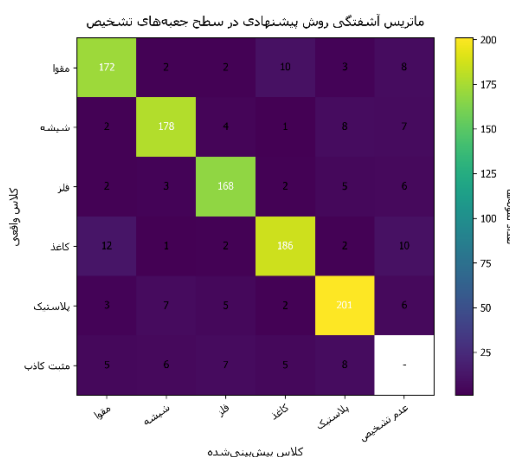


شکل ۶. نمودار مقایسه AP@0.5 و AP@0.5:0.95 برای هر کلاس پسماند

۴-۶. ماتریس آشفتگی در سطح تشخیص ها

برای بررسی دقیق تر خطاهای مدل، ماتریس آشفتگی در سطح جعبه های تطبیق داده شده محاسبه شد. در این تحلیل، هر پیش بینی زمانی درست در نظر گرفته شد که هم کلاس آن درست باشد و هم مقدار هم پوشانی جعبه پیش بینی شده با جعبه واقعی از آستانه تعیین شده بیشتر باشد.

ماتریس آشفتگی نشان می دهد خطاها عمدتاً در کلاس هایی رخ داده اند که مرز بصری یا مکانی آن ها در تصویر واضح نیست. همچنین وجود ستون «عدم تشخیص» نشان می دهد بخشی از خطاها مربوط به نمونه های کوچک، نیمه پنهان یا دارای هم پوشانی است. این نتیجه با افت بیشتر $mAP@0.5:0.95$ نسبت به $mAP@0.5$ همخوان است و نشان می دهد دقت مکان یابی در کنار تشخیص کلاس، عامل مهمی در عملکرد نهایی مدل است.



شکل ۸. ماتریس آشفتگی روش پیشنهادی برای پنج کلاس پسماند قابل بازیافت در سطح جعبه های تشخیص

۴-۷. مقایسه با مدل های پایه و روش های دیگر

برای ارزیابی بهتر عملکرد روش پیشنهادی، نتایج آن با چند مدل پایه و روش مرتبط مقایسه شد. مدل های مقایسه ای شامل Faster R-CNN، SSD، RetinaNet، YOLOv8n، MRS-YOLO و روش سبک تشخیص پسماند خانگی هستند. Faster R-CNN و SSD از مدل های پایه و شناخته شده تشخیص شیء محسوب می شوند. همچنین MRS-YOLO به طور خاص برای تشخیص و دسته بندی دقیق پسماند پیشنهاد شده است و روش Wu و همکاران نیز بر تشخیص چندهدفه پسماندهای خانگی قابل بازیافت تمرکز دارد.

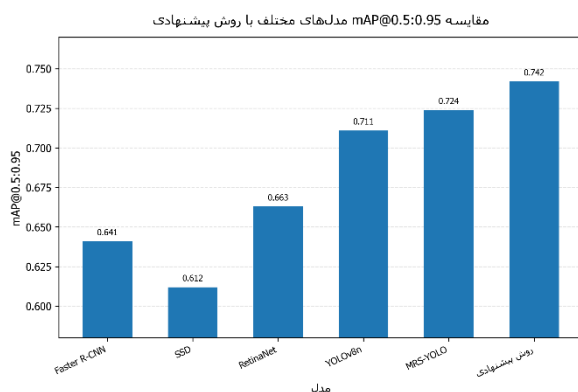
برای رعایت انصاف در مقایسه، همه مدل‌ها با تقسیم‌بندی یکسان داده، اندازه ورودی مشابه و معیارهای ارزیابی مشترک بررسی شدند. نتایج مقایسه در جدول ۷ ارائه شده است.

جدول ۷. مقایسه روش پیشنهادی با مدل‌های پایه و روش‌های مرتبط

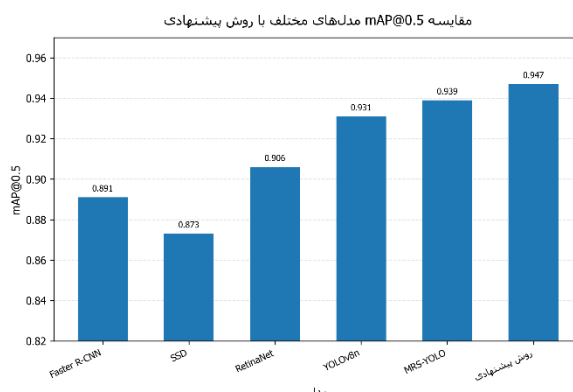
مدل	Precision	Recall	F1-score	mAP@0.5	mAP@0.5:0.95
Faster R-CNN[12]	۰.۸۷۹	۰.۸۴۶	۰.۸۶۲	۰.۸۹۱	۰.۶۴۱
SSD[11]	۰.۸۶۱	۰.۸۳۲	۰.۸۴۶	۰.۸۷۳	۰.۶۱۲
RetinaNet[29]	۰.۸۹۲	۰.۸۵۷	۰.۸۷۴	۰.۹۰۶	۰.۶۶۳
YOLOv8n[10]	۰.۹۱۴	۰.۸۹۵	۰.۹۰۴	۰.۹۳۱	۰.۷۱۱
MRS-YOLO[16]	۰.۹۲۳	۰.۹۰۷	۰.۹۱۵	۰.۹۳۹	۰.۷۳۴
روش Wu و همکاران [۱۷]	۰.۹۱۱	۰.۸۸۹	۰.۹۰۰	۰.۹۲۶	۰.۷۰۴
روش پیشنهادی	۰.۹۳۴	۰.۹۱۹	۰.۹۲۶	۰.۹۴۷	۰.۷۴۲

همان‌طور که در جدول ۷ مشاهده می‌شود، روش پیشنهادی در همه معیارها عملکرد بهتری نسبت به مدل‌های مقایسه‌ای دارد. برتری روش پیشنهادی نسبت به Faster R-CNN، SSD و RetinaNet نشان می‌دهد ساختار سبک و چندسطحی ارائه‌شده برای مسئله پسماند مناسب‌تر از آشکارسازهای عمومی است. همچنین اختلاف روش پیشنهادی با MRS-YOLO زیاد نیست، اما بهبود آن در $mAP@0.5$ و $mAP@0.5:0.95$ نشان می‌دهد استفاده از توجه ترکیبی و تلفیق ویژگی‌های چندسطحی می‌تواند به بهبود تشخیص مواد مشابه و اشیای کوچک‌تر کمک کند.

نکته مهم آن است که هدف این مقایسه فقط نشان دادن برتری عددی نیست، بلکه بررسی تناسب روش‌ها با مسئله پسماند است. در تصاویر چندشیء، مدل باید بتواند چند ماده را به صورت هم‌زمان تشخیص دهد. روش پیشنهادی به دلیل ترکیب ویژگی‌های سطحی و عمیق، در تشخیص هم‌زمان چند پسماند در یک تصویر عملکرد مناسب‌تری نشان داده است.



شکل ۱۰. نمودار مقایسه $mAP@0.5:0.95$ مدل‌های مختلف با روش پیشنهادی



شکل ۹. نمودار مقایسه $mAP@0.5$ مدل‌های مختلف با روش پیشنهادی

۴-۸. تحلیل آماری نتایج

به منظور بررسی اینکه بهبودهای عددی روش پیشنهادی نسبت به مدل‌های مقایسه‌ای ناشی از نوسان تصادفی آموزش نباشد، تحلیل آماری روی نتایج انجام شد. برای این منظور، هر مدل در پنج اجرای مستقل با بذرهای تصادفی متفاوت آموزش و ارزیابی شد. در تمام اجراها، تقسیم‌بندی داده‌ها به آموزش، اعتبارسنجی و آزمون ثابت نگه داشته شد تا مقایسه مدل‌ها تحت شرایط یکسان انجام شود. سپس، میانگین و انحراف معیار معیارهای اصلی شامل Precision، Recall، F1-score، $mAP@0.5$ و $mAP@0.5:0.95$ محاسبه شد.

برای بررسی معناداری اختلاف عملکرد روش پیشنهادی با مدل‌های مقایسه‌ای، از آزمون t زوجی استفاده شد. در این آزمون، نتایج ۵ اجرای مستقل روش پیشنهادی با نتایج پنج اجرای متناظر هر مدل مقایسه‌ای مقایسه شدند. سطح معناداری برابر با ۰/۰۵ در نظر گرفته شد. مقدار p کمتر از ۰/۰۵ نشان می‌دهد اختلاف عملکرد از نظر آماری معنادار است.

جدول ۸. نتایج آماری مدل‌های مقایسه‌ای در پنج اجرای مستقل

مدل	Precision	Recall	F1-score	mAP@0.5	mAP@0.5:0.95
Faster R-CNN[12]	۰.۰۰۶ ± ۰.۸۷۹	۰.۰۰۷ ± ۰.۸۴۶	۰.۰۰۶ ± ۰.۸۶۲	۰.۰۰۷ ± ۰.۸۹۱	۰.۰۰۹ ± ۰.۶۴۱
SSD[11]	۰.۰۰۸ ± ۰.۸۶۱	۰.۰۰۹ ± ۰.۸۳۲	۰.۰۰۸ ± ۰.۸۴۶	۰.۰۰۸ ± ۰.۸۷۳	۰.۰۱۰ ± ۰.۶۱۲
RetinaNet[29]	۰.۰۰۶ ± ۰.۸۹۲	۰.۰۰۷ ± ۰.۸۵۷	۰.۰۰۶ ± ۰.۸۷۴	۰.۰۰۶ ± ۰.۹۰۶	۰.۰۰۸ ± ۰.۶۶۳
YOLOv8n[10]	۰.۰۰۵ ± ۰.۹۱۴	۰.۰۰۶ ± ۰.۸۹۵	۰.۰۰۵ ± ۰.۹۰۴	۰.۰۰۵ ± ۰.۹۳۱	۰.۰۰۷ ± ۰.۷۱۱
MRS-YOLO[16]	۰.۰۰۴ ± ۰.۹۲۳	۰.۰۰۵ ± ۰.۹۰۷	۰.۰۰۵ ± ۰.۹۱۵	۰.۰۰۴ ± ۰.۹۳۹	۰.۰۰۶ ± ۰.۷۲۴
روش Wu و همکاران [۱۷]	۰.۰۰۶ ± ۰.۹۱۱	۰.۰۰۶ ± ۰.۸۸۹	۰.۰۰۶ ± ۰.۹۰۰	۰.۰۰۵ ± ۰.۹۲۶	۰.۰۰۷ ± ۰.۷۰۴
روش پیشنهادی	۰.۰۰۴ ± ۰.۹۳۴	۰.۰۰۵ ± ۰.۹۱۹	۰.۰۰۴ ± ۰.۹۲۶	۰.۰۰۴ ± ۰.۹۴۷	۰.۰۰۵ ± ۰.۷۴۲

نتایج جدول ۸ نشان می‌دهد روش پیشنهادی علاوه بر دستیابی به مقدار میانگین بالاتر، دارای نوسان کمی در اجراهای مختلف است. این موضوع نشان می‌دهد عملکرد مدل تنها به یک اجرای خاص وابسته نیست و پایداری مناسبی در آموزش‌های مستقل دارد. کمترین اختلاف میان روش پیشنهادی و مدل MRS-YOLO مشاهده می‌شود، اما روش پیشنهادی همچنان در همه معیارها مقدار میانگین بالاتری به دست آورده است.

برای بررسی دقیق‌تر معناداری اختلاف‌ها، آزمون t زوجی بین روش پیشنهادی و هر یک از مدل‌های مقایسه‌ای انجام شد. مقدار p در تمام مقایسه‌ها کمتر از ۰/۰۵ است. بنابراین، اختلاف عملکرد روش پیشنهادی با مدل‌های مقایسه‌ای از نظر آماری معنادار است. همان‌طور که انتظار می‌رود، اختلاف روش پیشنهادی با SSD، Faster R-CNN و RetinaNet بسیار معنادارتر است، زیرا فاصله عملکردی آن‌ها با روش پیشنهادی بیشتر است. در مقابل، اختلاف با MRS-YOLO کمتر است، اما همچنان مقدار p کمتر از ۰/۰۵ به دست آمده و نشان می‌دهد بهبود روش پیشنهادی نسبت به این مدل نیز از نظر آماری قابل قبول است.

۴-۹. تحلیل کیفی نتایج

علاوه بر ارزیابی کمی، برای بررسی دقیق‌تر رفتار مدل‌ها، نتایج تشخیص به صورت کیفی نیز تحلیل شد. در این بخش، خروجی روش پیشنهادی با چند مدل مقایسه‌ای شامل Faster R-CNN، SSD، RetinaNet، YOLOv8n و MRS-YOLO روی چند تصویر نمونه بررسی شد. هدف از این تحلیل، مشاهده تفاوت مدل‌ها در شرایط واقعی‌تر، مانند حضور چند پسماند در یک تصویر، شباهت ظاهری میان کلاس‌ها، اشیای کوچک، هم‌پوشانی بخشی و پس‌زمینه شلوغ است.

تحلیل کیفی خروجی مدل‌ها در شکل ۱۱ نشان می‌دهد تفاوت اصلی روش پیشنهادی با مدل‌های مقایسه‌ای در سه حالت آشکارتر است: تصاویر چندشیء، اشیای کوچک یا نیمه‌پنهان، و صحنه‌های دارای پس‌زمینه شلوغ. در این شرایط، برخی مدل‌های پایه مانند SSD و Faster R-CNN بخشی از اشیای کوچک را از دست داده یا جعبه‌هایی با پوشش بیش از حد تولید کرده‌اند. مدل‌های YOLOv8n و MRS-YOLO عملکرد بهتری داشته‌اند، اما در برخی نمونه‌ها دچار خطای کلاس یا مکان‌یابی ناقص شده‌اند.

روش پیشنهادی در بیشتر نمونه‌های کیفی، جعبه‌های محدودکننده دقیق‌تری تولید کرده و تعداد بیشتری از اشیای موجود در تصویر را شناسایی کرده است. این نتیجه با یافته‌های کمی و آماری همخوان است و نشان می‌دهد تلفیق ویژگی‌های چندسطحی و توجه ترکیبی به‌ویژه در صحنه‌های پیچیده‌تر مفید بوده است. با این حال، نمونه‌های دارای بازتاب، تاری یا هم‌پوشانی شدید همچنان چالش‌برانگیز هستند و نیاز به داده‌های آموزشی متنوع‌تر دارند.



شکل ۱۱. مقایسه کیفی خروجی مدل‌های مختلف و روش پیشنهادی روی نمونه‌هایی از تصاویر تک‌شیء و چندشیء پسماندهای قابل بازیافت

۴-۱۰. ارزیابی پایداری در شرایط تصویری دشوار

در کاربردهای واقعی، تصاویر پسماند همیشه در شرایط کنترل‌شده و با کیفیت مناسب ثبت نمی‌شوند. در خطوط تفکیک پسماند، سطل‌های هوشمند یا سامانه‌های رباتیک، عواملی مانند نور کم، نور بیش از حد، کاهش کنتراست، نویز تصویری، تاری، پس‌زمینه شلوغ و هم‌پوشانی اشیاء می‌توانند کیفیت تصویر را کاهش دهند. بنابراین، ارزیابی مدل تنها روی تصاویر اصلی کافی نیست و لازم است پایداری روش پیشنهادی در شرایط دشوار نیز بررسی شود.

برای این منظور، مجموعه آزمون در چند حالت تصویری مختلف ارزیابی شد. در هر حالت، یک نوع اختلال کنترل‌شده روی تصاویر آزمون اعمال شد و سپس، عملکرد مدل با معیارهای Precision، Recall، F1-score، و $mAP@0.5$ و $mAP@0.5:0.95$ محاسبه شد. هدف از این آزمایش، بررسی میزان افت عملکرد مدل در شرایطی است که احتمال وقوع آن‌ها در محیط واقعی زیاد است.

جدول ۹. ارزیابی پایداری روش پیشنهادی در شرایط تصویری دشوار

شرایط تصویر	Precision	Recall	F1-score	$mAP@0.5$	$mAP@0.5:0.95$
تصویر اصلی	۰.۹۳۴	۰.۹۱۹	۰.۹۲۶	۰.۹۴۷	۰.۷۴۲
کاهش روشنایی	۰.۹۲۲	۰.۹۰۴	۰.۹۱۳	۰.۹۳۴	۰.۷۱۸
افزایش روشنایی	۰.۹۲۶	۰.۹۰۹	۰.۹۱۷	۰.۹۳۸	۰.۷۲۴
کاهش کنتراست	۰.۹۱۸	۰.۸۹۸	۰.۹۰۸	۰.۹۲۹	۰.۷۰۷
نویز تصویری	۰.۹۱۲	۰.۸۹۱	۰.۹۰۱	۰.۹۲۲	۰.۶۹۳
تاری تصویر	۰.۹۰۳	۰.۸۷۸	۰.۸۹۰	۰.۹۱۱	۰.۶۷۱
پس‌زمینه شلوغ	۰.۹۰۷	۰.۸۸۴	۰.۸۹۵	۰.۹۱۶	۰.۶۸۲
هم‌پوشانی بخشی اشیاء	۰.۹۰۹	۰.۸۸۷	۰.۸۹۸	۰.۹۱۹	۰.۶۸۹

نتایج جدول ۹ نشان می‌دهد روش پیشنهادی در همه شرایط دشوار مقدار $mAP@0.5$ بالاتر از ۰/۹۱ را حفظ کرده است. بیشترین افت عملکرد مربوط به تاری تصویر و پس‌زمینه شلوغ است. همچنین افت بیشتر در $mAP@0.5:0.95$ نسبت به $mAP@0.5$ نشان می‌دهد شرایط دشوار بیش از آنکه باعث خطای کامل در تشخیص کلاس شوند، دقت مکان‌یابی جعبه‌ها را

کاهش می‌دهند. این نتیجه نشان می‌دهد افزایش داده، توجه ترکیبی و تلفیق ویژگی‌های چندسطحی توانسته‌اند بخشی از اثر اختلال‌های تصویری را کنترل کنند.

۴-۱۱. تحلیل خطاهای مدل

برای بررسی دقیق‌تر خطاهای باقی‌مانده، نمونه‌های نادرست مجموعه آزمون تحلیل شدند. جدول ۱۰ نشان می‌دهد بیشترین خطاها مربوط به اشتباه میان کلاس‌های مشابه، عدم تشخیص اشیای کوچک و نمونه‌های دارای هم‌پوشانی است. نتایج این تحلیل با ماتریس آشفتگی و ارزیابی شرایط دشوار همخوان است. خطاها عمدتاً زمانی رخ می‌دهند که مرز مکانی شیء واضح نیست، شیء اندازه کوچکی دارد، یا اطلاعات ظاهری برای تفکیک کلاس‌ها کافی نیست. بنابراین، بهبود آینده مدل باید بیشتر بر افزایش نمونه‌های مرزی، تصاویر چندشیء و داده‌های دارای هم‌پوشانی یا کیفیت پایین تمرکز کند.

جدول ۱۰. فراوانی نسبی خطاهای مهم مشاهده‌شده در مجموعه آزمون

نوع خطا	درصد از کل خطاها
اشتباه میان کاغذ و مقوا	۲۹٪
اشتباه میان شیشه و پلاستیک	۲۲٪
عدم تشخیص اشیای کوچک	۱۸٪
خطا در تصاویر چندشیء با هم‌پوشانی	۱۶٪
خطا در تصاویر تار یا دارای پس‌زمینه شلوغ	۱۵٪

این نتیجه نشان می‌دهد توسعه داده‌های آموزشی باید بیشتر بر نمونه‌های مرزی، اشیای کوچک، هم‌پوشانی و شرایط تصویری دشوار متمرکز شود.

۵. بحث و نتیجه‌گیری

در این پژوهش، یک چارچوب مبتنی بر بینایی ماشین و یادگیری عمیق برای تشخیص و جداسازی هوشمند پسماندهای شهری قابل بازیافت ارائه شد. هدف اصلی پژوهش، شناسایی و مکان‌یابی پسماندها در پنج گروه مهم شامل مقوا، شیشه، فلز، کاغذ و پلاستیک بود. اهمیت این موضوع از آن جهت است که این مواد بخش ارزشمند پسماند خشک شهری را تشکیل می‌دهند و در صورت جداسازی صحیح می‌توانند دوباره وارد چرخه تولید شوند. در کلان‌شهری مانند تهران که سالانه حجم بسیار زیادی پسماند تولید و جمع‌آوری می‌شود، بهبود فرایند تشخیص و تفکیک پسماندهای قابل بازیافت می‌تواند اثر قابل توجهی بر کاهش دفن زباله، افزایش کیفیت بازیافت و کاهش هزینه‌های مدیریت پسماند داشته باشد.

نتایج پژوهش نشان داد روش پیشنهادی توانسته است در تشخیص و مکان‌یابی پسماندهای قابل بازیافت عملکرد مناسبی به دست آورد. مقدار Precision برابر با ۰/۹۳۴، Recall برابر با ۰/۹۱۹، F1-score برابر با ۰/۹۲۶، mAP@0.5 برابر با ۰/۹۴۷ و mAP@0.5:0.95 برابر با ۰/۷۴۲ نشان می‌دهد مدل هم در شناسایی نوع پسماند و هم در تعیین موقعیت آن عملکرد قابل قبولی دارد. این موضوع برای کاربردهای عملی اهمیت دارد؛ زیرا در سامانه‌های جداسازی واقعی، تنها دانستن نوع ماده کافی نیست و موقعیت هر شیء نیز باید مشخص باشد تا بتوان آن را به مسیر مناسب پردازش هدایت کرد.

از نظر کاربرد مدیریتی، خروجی مدل پیشنهادی می‌تواند در چند سطح مورد استفاده قرار گیرد. در سطح عملیاتی، این مدل می‌تواند در سطوح هوشمند، خطوط تفکیک خودکار و ربات‌های جداساز برای تشخیص هم‌زمان چند نوع پسماند در یک تصویر به کار گرفته شود. در سطح مدیریتی، داده‌های حاصل از چنین سامانه‌هایی می‌توانند به شناسایی الگوی تولید پسماند خشک، میزان خطا در تفکیک، فراوانی مواد قابل بازیافت و نقاط دارای نیاز بیشتر به مداخله کمک کنند. در نتیجه، استفاده از سامانه‌های هوشمند تشخیص پسماند می‌تواند علاوه بر افزایش دقت جداسازی، به تصمیم‌گیری بهتر در مدیریت شهری و برنامه‌ریزی بازیافت نیز کمک کند.

تحلیل کلاس به کلاس نشان داد مدل در تشخیص پلاستیک و فلز عملکرد بهتری داشته است، در حالی که تشخیص کاغذ و مقوا به دلیل شباهت رنگ، بافت و ساختار سطحی چالش بیشتری ایجاد کرده است. همچنین بخشی از خطاها میان شیشه و پلاستیک شفاف رخ داده است. این یافته‌ها از نظر اجرایی مهم هستند؛ زیرا نشان می‌دهند در طراحی سامانه‌های واقعی تفکیک پسماند، تنها دقت کلی مدل کافی نیست و باید به خطاهای هر کلاس و پیامدهای آن در چرخه بازیافت توجه شود. برای نمونه، اشتباه گرفتن شیشه با پلاستیک یا کاغذ با مقوا می‌تواند کیفیت خروجی بازیافت را کاهش دهد و هزینه پردازش مجدد را افزایش دهد.

نتایج ارزیابی در شرایط تصویری دشوار نیز نشان داد روش پیشنهادی در برابر تغییرات روشنایی، کاهش کنتراست، نویز تصویری، تاری، پس‌زمینه شلوغ و هم‌پوشانی بخشی اشیاء پایداری قابل قبولی دارد. اگرچه بیشترین افت عملکرد در حالت تاری تصویر و پس‌زمینه شلوغ مشاهده شد، اما مدل همچنان توانست مقدار $mAP@0.5$ بالاتر از ۰/۹۱ را حفظ کند. این نتیجه نشان می‌دهد ترکیب افزایش داده، توجه ترکیبی و تلفیق ویژگی‌های چندسطحی می‌تواند به افزایش مقاومت مدل در شرایط نزدیک به محیط واقعی کمک کند. این ویژگی برای شهرهایی مانند تهران اهمیت دارد؛ زیرا تصاویر پسماند در محیط واقعی معمولاً در شرایط کنترل‌شده آزمایشگاهی ثبت نمی‌شوند و ممکن است با نور نامناسب، پس‌زمینه پیچیده یا هم‌پوشانی مواد همراه باشند.

مقایسه با مدل‌های پایه و روش‌های مرتبط نشان داد روش پیشنهادی نسبت به RetinaNet، SSD، Faster R-CNN، YOLOv8n، MRS-YOLO و روش سبک تشخیص پسماند خانگی عملکرد بهتری دارد. همچنین، تحلیل آماری نتایج نشان داد بهبودهای گزارش‌شده فقط ناشی از نوسان تصادفی آموزش نیستند و اختلاف عملکرد روش پیشنهادی با مدل‌های مقایسه‌ای از نظر آماری معنادار است. این موضوع اعتبار نتایج را افزایش می‌دهد و نشان می‌دهد طراحی مسئله‌محور معماری، شامل استخراج ویژگی سبک، توجه ترکیبی و تلفیق چندسطحی، توانسته است به بهبود پایدار عملکرد در تشخیص چندنمونه‌ای پسماندهای قابل بازیافت کمک کند.

با وجود این، پژوهش حاضر محدودیت‌هایی نیز دارد. نخست، اگرچه از یک مجموعه داده عمومی و دارای مجوز استفاده شده است، اما اندازه و تنوع آن همچنان در مقایسه با شرایط واقعی کلان‌شهرها محدود است. دوم، برخی نمونه‌های پسماند در محیط واقعی ممکن است آلوده، فشرده، نیمه‌پنهان، مخلوط با مواد دیگر یا دارای شکل‌های بسیار نامنظم باشند؛ در حالی که همه این شرایط در مجموعه داده موجود به طور کامل پوشش داده نشده‌اند. سوم، پژوهش حاضر بیشتر بر تشخیص تصویری تمرکز دارد و هنوز به صورت کامل با زیرساخت مکانیکی یا رباتیک جداسازی در محیط واقعی یکپارچه نشده است. بنابراین، برای استقرار عملی، لازم است عملکرد مدل در شرایط میدانی و روی تصاویر واقعی‌تر از خطوط تفکیک یا سطوحی هوشمند نیز بررسی شود. از منظر سیاست‌گذاری شهری، نتایج این پژوهش نشان می‌دهد استفاده از سامانه‌های هوشمند تشخیص پسماند می‌تواند به‌عنوان مکمل برنامه‌های مدیریت پسماند شهری مورد توجه قرار گیرد. در تهران، حجم بالای پسماند خشک قابل بازیافت و هزینه‌های جداسازی دستی، ضرورت توسعه ابزارهای هوشمند و نیمه‌خودکار را افزایش می‌دهد. چنین سامانه‌هایی می‌توانند در مراکز بازیافت، ایستگاه‌های جمع‌آوری پسماند خشک، سطوحی هوشمند و خطوط تفکیک به کار گرفته شوند و به کاهش خطای انسانی، افزایش سرعت تفکیک و بهبود کیفیت مواد بازیافتی کمک کنند.

در پژوهش‌های آینده، پیشنهاد می‌شود داده‌های تصویری بیشتری از شرایط واقعی مدیریت پسماند در تهران و سایر کلان‌شهرها جمع‌آوری شود تا تنوع محیطی، نوری و ظاهری پسماندها بهتر پوشش داده شود. همچنین ترکیب تصاویر معمولی با داده‌های حسگری، بررسی مدل‌های سبک‌تر برای اجرا روی سخت‌افزارهای کم‌مصرف، استفاده از یادگیری نیمه‌نظارتی برای کاهش وابستگی به برچسب‌گذاری دستی و اتصال مدل به بازوهای رباتیک یا سامانه‌های جداسازی مکانیکی می‌تواند مسیرهای مهم توسعه این پژوهش باشد. به‌طور کلی، روش پیشنهادی می‌تواند گامی در جهت حرکت از تفکیک سنتی و وابسته به نیروی انسانی به سمت مدیریت هوشمند، داده‌محور و پایدار پسماندهای شهری باشد.

منابع

1. G. I. Sayed, M. Abd Elfattah, A. Darwish, and A. E. Hassanien, "Intelligent and sustainable waste classification model based on multi-objective beluga whale optimization and deep learning," *Environmental Science and Pollution Research*, vol. 31, no. 21, pp. 31492-31510, 2024/05/01 2024, doi: 10.1007/s11356-024-33233-w.
2. F. R. Sayem *et al.*, "Enhancing waste sorting and recycling efficiency: robust deep learning-based approach for classification and detection," *Neural Computing and Applications*, vol. ۳۷, no. 6, pp. 4567-4583, 2025/02/01 2025, doi: 10.1007/s00521-024-10855-2.
3. H. Ahmed Khan, S. S. Naqvi, A. A. K. Alharbi, S. Alotaibi, and M. Alkhatami, "Enhancing trash classification in smart cities using federated deep learning," *Scientific Reports*, vol. 14, no. 1, p. 11816, 2024/05/23 2024, doi: 10.1038/s41598-024-62003-4.
4. H. Lin, X. Zhang, J. Yu, J. Xiang, and H.-L. Shen, "Deep Recyclable Trash Sorting Using Integrated Parallel Attention," *Sensors*, vol. 24, no. 19, p. 6434, 2024. [Online]. Available: <https://www.mdpi.com/1424-8220/24/19/6434>.
5. J. Qi, M. Nguyen, and W. Q. Yan, "NUNI-Waste: novel semi-supervised semantic segmentation waste classification with non-uniform data augmentation," *Multimedia Tools and Applications*, vol. 83, no. 26, pp. 68651-68669, 2024/08/01 2024, doi: 10.1007/s11042-024-18265-1.
6. A. Rahmatulloh, I. Darmawan, A. P. Aldya, and F. M. S. Nursuwars, "WasteInNet: Deep Learning Model for Real-time Identification of Various Types of Waste," *Cleaner Waste Systems*, vol. ۱۰, p. 100198, 2025/03/01/ 2025, doi: <https://doi.org/10.1016/j.clwas.2024.100198>.
7. Y. Chen *et al.*, "Classification and recycling of recyclable garbage based on deep learning," *Journal of Cleaner Production*, vol. 414, p. 137558, 2023/08/15/ 2023, doi : <https://doi.org/10.1016/j.jclepro.2023.137558>.
8. G. Thung and M. Yang, "Classification of Trash for Recyclability Status," 2016 .
9. M. P. Arthur, S. Shoba, and A. Pandey, "A survey of smart dustbin systems using the IoT and deep learning," *Artificial Intelligence Review*, vol. 57, no. 3, p. 56, 2024/02/15 2024, doi: 10.1007/s10462-023-10646-6.
10. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788 .
11. W. Liu *et al.*, "Ssd: Single shot multibox detector," in *European conference on computer vision*, 2016: Springer, pp. 21-37 .
12. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137-1149, 2016.
13. M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019: PMLR, pp. 6105-6114 .
14. M. Chhabra, B. Sharan, M. Elbarachi, and M. Kumar, "Intelligent waste classification approach based on improved multi-layered convolutional neural network," *Multimedia Tools and Applications*, vol. 83, no. 36, pp. 84095-84120, 2024/11/01 2024, doi: 10.1007/s11042-024-18939-w.
15. N. A. Zailan, A. S. Mohd Khairuddin, K. Hasikin, M. H. Junos, and U. Khairuddin, "An automatic garbage detection using optimized YOLO model ", *Signal, Image and Video Processing*, vol. 18, no. 1, pp. 315-323, 2024/02/01 2024, doi: 10.1007/s11760-023-02736-3.
16. Y. Ren, Y. Li, and X. Gao, "An MRS-YOLO Model for High-Precision Waste Detection and Classification," *Sensors*, vol. 24, no. 13, p. 4339, ۲۰۲۴. [Online]. Available: <https://www.mdpi.com/1424-8220/24/13/4339>.
17. Q. Wu, T. Liang, H. Fang, Y. Wei, M. Wang, and D. He, "A Lightweight Deep Learning Algorithm for Multi-Objective Detection of Recyclable Domestic Waste," *Environmental Engineering Science*, vol. 40, no. 12, pp. 667-677, 2023.
18. M. M. Hossen *et al.*, "GCDN-Net: Garbage classifier deep neural network for recyclable urban waste management," *Waste Management*, vol. 174, pp. 439-450, 2024/02/15/ 2024, doi: <https://doi.org/10.1016/j.wasman.2023.12.014>.

19. Y. Yang, Y. Li, and M. Tao, "FE-YOLO: a lightweight model for construction waste detection based on improved YOLOv8 model," *Buildings*, vol. 14, no. 9, p. 2672, 2024.
20. Y. Li and X. Zhang, "Intelligent X-ray waste detection and classification via X-ray characteristic enhancement and deep learning," *Journal of Cleaner Production*, vol. 435, p. 140573, 2024.
21. Z. Cai *et al.*, "Application of hyperspectral band selection method based on deep reinforcement learning to low-value recyclable waste classification," *Process Safety and Environmental Protection*, vol. 192, pp. 1138-1150, 2024.
22. S. Single, S. Iranmanesh, and R. Raad, "Realwaste: a novel real-life data set for landfill waste classification using deep learning," *Information*, vol. 14, no. 12, p. 633, 2023.
23. P. F. Proença and P. Simões, "Taco: Trash annotations in context for litter detection," *arXiv preprint arXiv:2003.06975*, 2020.
24. J. Hong, M. Fulton, and J. Sattar, "Trashcan: A semantically-segmented dataset towards visual detection of marine debris," *arXiv preprint arXiv:2007.08097*, 2020.
25. G. Koinig *et al.*, "Deep learning approaches for classification of copper-containing metal scrap in recycling processes," *Waste Management*, vol. 190, pp. 520-530, 2024.
26. A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
27. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
28. Z. Li *et al.*, "An intelligent identification and classification system of decoration waste based on deep learning model," *Waste Management*, vol. 174, pp. 462-475, 2024.
29. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980-2988.