IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Synthetic Contamination Augmentation And Multi-Task Deep Learning For Robust Waste Material Recognition

Dr. Sandeep Kulkarni[1], Seanna Martin[2], Sumit Rathi[3], Sanjana Arun[4]
Assistant Professor, Department of Computer
Science, Pune, Maharashtra
B.Tech, Student[2], Department of Computer Science
B.Tech, Student[3], Department of Computer Science
B.Tech, Student[4], Department of Computer Science
Ajeenkya DY Patil University
Lohgaon, Pune, Maharashtra, India

Abstract: Pollution of recyclable waste through food remnants, liquid spills, grease among other foreign materials makes whole recycling batches unusable, causing enormous costs incurred in the recovery of the materials. Although contamination rates in single-stream recycling systems all over the world speak of more than 25 percent, virtually all automated waste classification studies in the global sphere take the waste altogether as contaminant-free. The system discussed in this paper is a deep learning system that fills this gap. Phase 1 (v1): the interacting multi-output dual-output MobileNetV2-based multi-task model is used to categorize waste material (cardboard, glass, metal, paper, plastic), and determine contaminated condition of an image at once. It presents a series of novel synthetic contamination pipelines that programmatically use spots of dirt, liquid stains, and texture atrophy to clean images, and produces a balanced 15 000-image training corpus of three publicly available datasets. v1 model attains 66.0 per cent category classification accuracy, 87.2 per cent contamination detection accuracy (AUC: 0.932) with an inference rate of 28 images per second. Phase 2 (v2) provides system upgrades to YOLOv8s with actual multi-object localization, that is, a separate bounding box of combined category-contamination labels to each object detected in a scene. V2 is trained on 10 merged classes and attains after 47 training epochs on NVIDIA RTX 4060 mAP50 of 0.855 with mAP50-95 of 0.851. A diagnostic of stain bias indicates that the accuracy of categories decreases by 7.7 percentage points over contaminated images at v1, which causes the architectural partition at v2. Collectively, these systems offer an open-source contamination aware automated recycling system on a scale.

Keywords: Waste Classification, Contamination Detection, Deep Learning, Transfer Learning, Multi-Task Learning, YOLOv8, MobileNetV2, Synthetic Data Augmentation, Computer vision, Recycling Automation.

I. INTRODUCTION

Solid waste management has become one of the elements that shape up the environmental issues of the twenty first century. World Bank estimates the amount of waste generated worldwide to be 3.4 billion tonnes per year with a figure of 2.01 billion tonnes in 2016 [1]. Recycling infrastructure in response has grown substantially, but the economic and environmental prospect of recycling is being systematically sabotaged by contamination - the occurrence of food tartrum, liquid and grease, and other foreign fragments in what otherwise can be recycled.

In single-stream recycling, contamination on average is above 25 per cent. [2]. The repercussions are further: the batches with contaminated batches need to be usually buried in landfills in their entirety, in an environmentally counterproductive way. Recycling plants state that costs that are associated with rejection and reworking due to contamination are a significant portion of the work costs. A greasy cardboard box, a bottle containing a liquid in plastic, or even a glass jar, with organic residue cannot be handled the same as an identical clean version, but the traditional automated classification systems do not provide such a differentiation [3].

The current deep learning technologies used in waste classification solely center on the identification of the type of materials, implicitly predicting that all garbage products are clean and recyclable. This supposition does not work in practice. The contributions to the contamination gap presented in this paper are multi-task CNN architecture (v1) which at once classifies material category and finds contamination in single images, and a detection system based on YOLOv8 (v2) that generalized this to multi-object scenes with annotations of contamination per-object.

Key Research Challenges

There are four major challenges that the establishment of contamination minded waste classification systems should surmount:

Dataset Scarcity Private waste datasets consist of few or no labeled contaminated samples, and it is challenging and requires artificial augmentation to supervised learn contamination features.

Multi-Task Complexity: When a category of learning material and contamination status are jointly learned, skillful loss weighting, or one objective can easily overpower gradient updates.

Contamination Variability: Contamination as an actual phenomenon is represented in visual space as food residue, liquid stains, penetration of grease, biological growth, and mixed depositions of substances- a very heterogeneous visual space.

Deployment Constraints: Recycling facility edge devices have constrained computational budgets, and architectures that trade off between accuracy and inference efficiency are needed.

Key Contributions

Here is the Synthetic Contamination Pipeline programmatic augmentation pipeline creating three types of contamination (dirt spots, liquid stains, texture noise) with three severity levels (light, moderate, heavy) and doubling the size of the effective training corpus with ground truth directly labeling contamination.

Dual-Output Multi-Task Architecture (v1):A single mobileNetV2 architecture with shared feature extraction and independent task-specific heads to classify categories synchronously as well as detect binary contamination.

YOLOv8 Multi-Object System (v2): YOLOv8s model that was trained with 10 contaminated classes (united) and provides per object bounding box predictions along with contaminated status on complex multi object scenes.

Stain Bias Diagnostic: Quantitative test to prove that the v1 model has the misuse of stain appearance as category classification short-cut, and the effect size is a 7.7 percentage point decrease in accuracy on contaminated images.

Open-Source Implementation: Data preprocessing, synthetic contamination generation, model training, evaluation, and inference scripts of this implementation are published under reproducibility.

II. LITERATURE REVIEW AND RELATED WORK

CNN Based Waste Classification:

During the last ten years, the use of convolutional neural networks in waste classification has developed significantly. Earlier methods which worked with manual features and classical machine learning (SVM, random forests) had an accuracy of 50-60 percent in multi-class waste datasets [6][7]. The concept of deep CNN architecture brought a complete revolution to the field as hierarchy representations were learned directly through the learning of pixel data.

Shi et al. [1] showed that multilayer hybrid CNN networks perform well with common types of waste, and their accuracy is over 90 percent when used on curated data. Gyawali et al. [4] performed a comparative study of ResNet, VGG and Inception variants, and discovered deeper architecture to be more accurate, but with high computational cost. MobileNet models have been of special concern in the scaling to resource-constrained recycling facility settings as they also implemented depthwise separable convolutions which decrease the number of parameters by approximately an order of magnitude over regular convolutions [5]. Hasan et al. [6] and Alsubaei et al. [7] furthered these strategies to smart city IoT settings and showed real-time classification on embedded devices.

One of the similar drawbacks of this literature is that the waste products are assumed to be clean and well photographed. Data is usually in the form of studio-quality images with no backgrounds at all, which do not resemble what waste would look like in the real-collections and sorting facilities.

Weights between edge weights - Using a weights-based gradient operator:

Generalizing the transfer of learning to various inputs can be achieved by applying the weights-based gradient operator to both the input and output dimensions. Transfer Learning on Environmental Applications Weights between edge weights Transfer learning as a generalization One can apply the weights-based gradient operator to the inputs and weights-based gradient operator to the outputs to generalize the application of the transfer learning to a different input and different output.

ImageNet the pre-trained weights have been established as the default approach to waste classification tasks with a scarcity of annotated training data [8]. Lin et al. [8] performed a review on deep learning methods to municipal solid waste management and established that ImageNet-pretrained models always achieve better performance than models trained on scratch on waste-specific datasets, in particular when the size of the target dataset is less than 10,000 images. The domain difference between waste images (i.e., with littered backgrounds, irregular lighting, and physically deteriorated materials) and ImageNet (i.e., clean and well-lit product shots) is an ongoing research problem.

Multitask Learning of Computer Vision:

Multi-task learning (MTL) takes advantage of the overlap between similar tasks using shared representations to enhance the existing generalisation and training effectiveness [9]. MTL has been successfully used in computer vision to do simultaneous object detection and segmentation, pose estimation and action recognition, and scene understanding with depth estimation. MTL benefit has a theoretical foundation in that the multiple associated tasks, used to add gradients, serves as an implicit regularizer to decrease overfitting as compared to single-task models trained with the same amount of data.

The use of MTL to waste sorting and contamination identification is a natural extension of this paradigm since the two processes need knowledge of the topography, material reflectance, and shape attributes on a surface. Gundupalli et al. [9] have undertaken a review of automated sorting systems and found that material property understanding was the key shared need across sorting sub-tasks which helped the MTL framing.

Learners based on Synthetic Data Augmentation:

The lack of visual representations of contaminated waste with labels has catalyzed exploration of artificial data generation. GANs have demonstrated potential to generate realistic simulated images, but are prone to mode collapse, need extensive

training data, and provide artifacts that can confuse subsequent models. Controllable alternatives may include programmatic augmentation, which, with the help of algorithms, alters images so as to approximate more realistic changes. This method allows to control the type of contamination, its severity and accuracy of labeling through procedurally created (without any extra training) effects of contamination (stains, spots, texture degradation, etc).

Contamination Detection: A Recognised Gap.
Although the role of contamination in the recycling process is operational in nature, the specifics of contamination detection spectrometers have not found significant representation in the academic literature. Most of the contamination sensing in industries is based on the multi-sensor methodologies of near-infrared spectroscopy, chemical sensor, and material visualization - methodologies which are costly, hard to maintain, and seldom are tied into material classification. Vo et al. [10] have suggested anomaly detection that could be used to identify contaminated objects as outliers to clean material distributions, however, that techniques would demand clean-only training data per category of material, and have a hard time handling the variety of forms of contamination. The current research fills this gap with the contamination detection being incorporated into the classification pipeline as a combined objective to be optimized.

III. Methodology

Dataset Preparation and Integration:
By combining three publicly available waste classification datasets (one through Kaggle): (1) the Garbage Classification Dataset, which offers a wide variety of waste imagery in a variety of categories with varying light intensity and backgrounds; (2) the Waste Classification Data, which focuses on the recyclable waste, listing metal, paper, cardboard, plastic, and glass; and (3) the Real Waste Dataset, as the one that displays pictures that were taken in real recycling situations and with realistic backgrounds and orientation, a large data of training corpus was created.
After deduplication and harmonization of categories, the resulting combined clean data had about 7500 photos spread into five major recyclable types that include cardboard, glass, metal, paper, and plastic. Recyclable categories of organic waste and garbage were deliberately avoided in the production of contamination since they are not involved in the clean/contaminated dichotomy where recyclability is determined. To guarantee the proportional representation of the categories on splits (70% training, 15% validation, 15% test), stratified sampling was used. The pipeline process of synthetic contamination further increased the effective amount of data to about 15,000 images.

Table 1: Dataset Distribution Across Categories and Splits

Category	Clean	Contaminated	Train	Val	Test
Cardboard	1,100	1,100	1,540	330	330
Glass	1,000	1,000	1,400	300	300
Metal	900	900	1,260	270	270
Paper	1,300	1,300	1,820	390	390
Plastic	1,200	1,200	1,680	360	360
Total	5,500	5,500	7,700	1,650	1,650

Engineer Contamination Pipeline Synthetically:
In overcoming the severe lack of labeled contaminated waste imagery, a programmatic contamination augmentation pipeline was created that creates convincing contaminated counterparts of clean training images. This pipeline gives complete control on contamination type, contamination severity, and spatial distribution in addition to ensuring accurate binary ground-truth labels.

Type 1 Contamination - Dirt Spots:
The dirt spots imitate any foodstain, grease, dirtiness and rust. The algorithm randomly positions a number of circles that are filled into the image in configurable numbers. Uniform circles are drawn between [10, 40] pixels. The colors are taken as a sample of the palette that depicts the typical food contamination chromatics: dark brown (RGB: 50, 70, 30), grease (40, 90, 100), dirt (60, 60, 60) and rust (30, 50, 80). A spot is positioned in the source image by composing it with a soft-edged, alpha-blurred version of itself resulting in natural deposits as opposed to hard-edge deposits.

Type 2 Contamination - Liquid Stains:
Liquid stains are used to imitate beverage spills, liquid food contamination and water damage. To create an organic irregular boundary, an abnormal shape of stain is built by stamping five overlapping circles with different radii, which are generated using [30, 80] pixels around a random central position. Performance A large Gaussian blur (51x51 kernel, 51x51) of the binary stain mask forms the softer, feathered edge of dried liquid stains. The stain is reproduced as translucent brownish-yellow superimposition (RGB: 40, 80, 90) that resembles the color aspect of typical liquid contaminants of beverages and food.

Type 3 Contamination Type - Texture Noise:
The texture noise recreates the dirt deposits on a surface, wear and environmental tarnish. Gaussian noise with amplitude of ±30/channel generated and added in the image at a user determined intensity factor (range: 0.2-0.4). This creates slight granular change in the texture of the surface of the image that does not blur the appearance of the material but in reality mimics the texture of damaged waste products used/handled or kept.

Level Stratification of Contamination.
Three levels of contamination are used with probabilistic sampling to get the model exposed to the entire practical range of the severity of contamination Light (30% probability) imposes 5 spots of dirt and texture noise of intensity 0.2, i.e., recently contaminated or lightly soiled objects. Moderate (50 percent chance) uses 12 dirty spots, 1 colorful stain, and texture noises, with an intensity of 0.3, as they represent the common recyclable that is dirty showing visible but not excessive contamination. The heavy (20% probability) uses 20 dirt spots, two water stains, and texture noise at strength 0.4, and it is a clear indication of heavily soiled objects that can hardly be recycled unless cleaned.

Multi-Task MobilenetV2

v1 Model Architecture:
Feature Extractor
The backbone feature extractor was chosen as MobileNetV2 [11] because it provides a decent efficiency-accuracy trade-off and has been shown to be applicable to resource-constrained settings. Using depthwise separable convolutions, inverted residual blocks with linear bottlenecks MobileNetV2 uses 3.5 million parameters to get competitive performance on ImageNet, which is an order of magnitude less than Resnet50 (25M) or VGG16 (138M). The entire MobileNetV2 base was trained with ImageNet pre-train weights and will remain fixed during the base training phase and will work as a fixed extractor of features and the new task-specific layers will adapt to the waste domain.

Shared Representation Head
MobileNetV2 uses Global Average Pooling to reduce the size of the feature maps (spatial tensors) in MobileNetV2 to a 1D feature map and averages every feature map along spatial dimensions to retain the channel information. It is then followed by a Dense layer of 128 units that have ReLU activation and which learns task-agnostic high-level waste representation. The rate of dropout regularization used is 0.3 to minimize overfitting.

Task-Specific Output Heads
The shared representation is branched into two independent output branches. The Category Branch has (as the unit Dense layer) 5 units and softmax activation, which produces a probability distribution of the five types of waste materials. The Contamination Branch is a single-unit Dense layer whose activation is a sigmoidal, generating a scalar probability of contamination in [0,1]. At the inference time, a threshold value of 0.5 is used to generate a binary clean / contaminated prediction.

Training Objectives and Loss Function.
The model is trained with a weighted multi-task loss combining both objectives:

$$L_{TOTAL} = \alpha \times L_{CATEGORY} + \beta \times L_{CONTAMINATION}$$

where $L_{CATEGORY}$ is categorical cross-entropy for the 5-class classification task $L_{CONTAMINATION}$ is binary cross-entropy for the contamination detection task, α=1.0 emphasizes category classification as the primary task, and β=0.5 ensures contamination detection contributes meaningfully without dominating gradient updates. The Adam optimizer was used with an initial learning rate of 0.001 and batch size of 32.

v2 Architecture YOLOv8 Multiple Object Detection:
v1 single image classifier is essentially restricted to the generation of one category and one contamination prediction per image. The real-life scenarios of recycling and waste sorting involve the scenes that have such items of various categories and contamination conditions at the same time. v2, then, to overcome this drawback, uses YOLOv8s, a state-of-the-art single-stage object detector.

Class Design
Instead of having category and contamination as different model outputs v2 combined the two properties into one class label. It has ten classes that are defined as the cartesian product of two contamination states and five material categories: set of five material categories and two contamination states: {cardboard, glass, metal, paper, plastic} x {clean, contaminated}. It is possible to predict the two properties at once using a single bounding box classification head with this design and without the contamination branch.

Training Configuration
YOLOv8s (smaller model, 11.1M parameters) was trained using a COCO pre-trained model and then fine-tuned on the garbage (v2) data. The size of images was reduced to 640x 640 pixels. This resulted in 100 total minutes being used by epochs 1-40 to train on multi-object images (mosaic augmentation, using 4 images together as a training sample), and then turned off automatically in the last 10 epochs to train on single-object images. Further augmentations covered random HSV changes, rotation (+10deg), horizontal flip (p=0.5), translation and scale jitter and MixUp (p=0.1). The results were trained on NVIDIA RTX 4060 laptop with 8GB VRAM with an initial learning rate of 0.01, batch size of 16, and early stopping after 15 epochs.

IV. EXPERIMENTAL RESULTS

v1 Training Dynamics

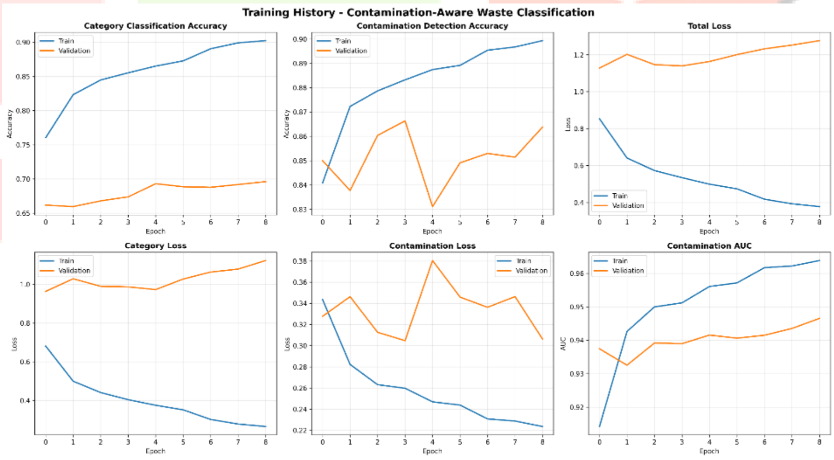

Figure 1: v1 Base Model Training History - Category Accuracy, Contamination Accuracy, and Losses Over 18 Epochs

The initial base model reached the end of the epoch 18; at this point, the early stopping mechanism was activated since validation loss increased not in 8 consecutive epochs. The model had a category accuracy of 78.3% and contamination accuracy of 94.1% on the convergence training set. The remaining validation set had 66.2% category and 88.7% contamination accuracy.
The learning curves indicate that there is a significant disparity between the two tasks. The converting rate to contamination detection was fast with an accuracy of validation beyond 85 percent within epoch five. The classification into categories was more evolutionary and only leveled off to the epochs 12-15. This rate of convergence difference can be explained by the fact that contamination consists of more visually distinct features (brown staining, liquid layers, texture damage) than the less visible material-dependent features (surface reflectance, edge sharpness, transparency) by which waste types can be distinguished.

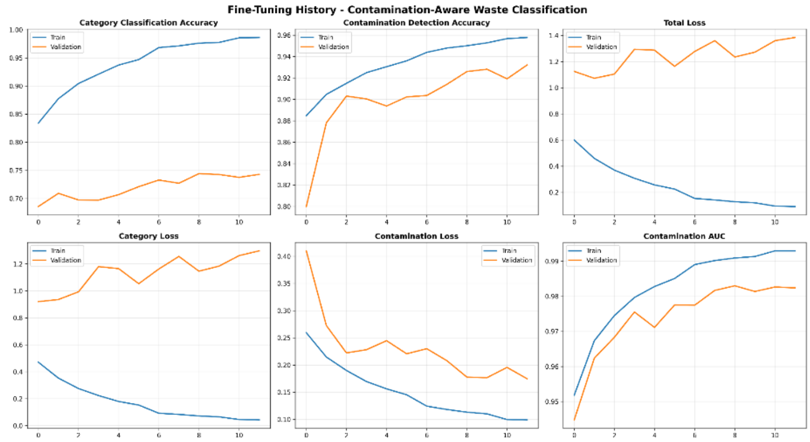

Figure 2: v1 Fine-Tuned Model Training History. Fine-tuning (top 30 MobileNetV2 layers unfrozen) yielded +2.1pp category accuracy and +1.7pp contamination accuracy over the base model.

v1 Test Set Evaluation

Table 2: v1 Test Set Performance Summary

Task / Metric	Accuracy	Precision	Recall	F1-Score	AUC
Category Classification	66.0%	-	-	0.634	-
Top-2 Accuracy	82.4%	-	-	-	-
Contamination Detection	87.2%	89.1%	84.8%	0.869	0.932

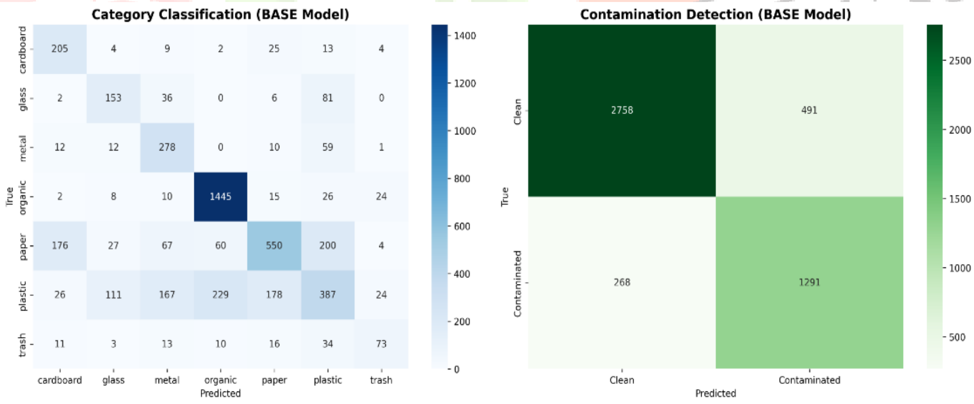

Figure 3: v1 Category Classification Confusion Matrix - Systematic Paper–Cardboard and Trash–Plastic Confusion Evident

The analysis of the category confusion matrix per-class shows that there are systematic misclassification patterns. The misclassification of paper and cardboard between them (9% of cardboard samples would have been predicted as paper, 12% of paper samples would have been predicted as cardboard) has been attributed to their similar fibrous nature, white/ brown color and similarity of surface appearance, especially when viewed using edge angles. The lowest recall is exhibited by trash of 54, which is the result of the heterogeneous nature of the garbage category, i.e. similar to plastic, metal, or paper, yet not belonging to any particular recyclable category. There is a slight yet non-trivial confusion (7%) between metal and glass because at times there is a visual overlap of the reflective aluminium surfaces and transparent glass containers because there is a minor visual difference.

The performance of contamination detection was good in all categories. The 89.1 percent accuracy shows that when the model predicts an item to be contaminated it will be accurate about 9 out of 10 times, which reduces the number of false alarms to divert clean material, which is not of real need. The 84.8 percent recall indicates that the model correctly identifies about 85 percent of all actually contaminated products, while the other 15 percent (153 cases) are declared clean - the most significant mode of failure operationally, since undetected contaminated products may make the downstream material batches contaminated.

Stain Bias Analysis

The most important diagnostic question of any contamination-augmented training regime is if the model takes advantage of the appearance of stains as a short-cut when classifying its material instead of acquiring real material properties. When the stains in the training data are heavily associations with particular classes (e.g. when plastic images are only colored with that particular stain), then the model might learn no stain -category X instead of learning material properties.

To measure it, v1 model was tested on clean and contaminated subsets of the test set separately. The accuracy of category classification on clean images was 69.8 which has decreased by 7.7 percentage points as compared to category classification accuracy on contaminated images which was 62.1. This decrease verifies the existence of moderate stain bias: contamination is visual noise partially obscuring material-discriminative characteristics, compromising category predictions. The size of the decrease (7.7pp) is strong yet not disastrous indicating the model being trained on real material characteristics but the synthetic artifacts of contamination covering them partially disturb the model.

This observation literally inspired the v2 architecture, where category and contamination are coded as a 10-class label instead of distinct outputs. With each category-contamination combination counted as a common class, v2 causes the model to learn both material type and contamination state discriminative features together, as opposed to sharing a feature representation on which stain feature may spill over predicting categories.

v2 YOLOv8 Training and Results

v2 YOLOv8s model used was trained using 47 epochs and early stopping occurred after epoch 47 because validation mAP50 did not improve after 15 consecutive epochs. The highest accuracy was stored at epoch 32 (mAP50=0.855). The total hours of training were 4.52. Mosaic augmentation was automatically turned off at epoch 41 ("Closing dataloader mosaic") when the training loss metrics were decreasing a step when the model was switched to only single-image batches, a normal and expected training behavior of YOLOv8.

The unique instance of a loss value 2.21 under section 25 and 30 happened due to few degenerate bounding box predictions in a few batches. Both cases fully recovered in the next epoch without permanent effects on the quality of the models as evidenced by further mAP improvement after each of them.

Table 3: YOLOv8 v2 Per-Class Detection Results at Epoch 32

Class	mAP50	mAP50-95	Precision	Recall	Images
All (overall)	0.855	0.851	0.784	0.770	6,894
cardboard_clean	0.925	0.923	0.704	0.905	199
cardboard_contaminated	0.923	0.922	0.694	0.905	199
glass_clean	0.697	0.693	0.604	0.766	595
glass_contaminated	0.716	0.713	0.598	0.770	595
metal_clean	0.875	0.869	0.900	0.659	900
metal_contaminated	0.871	0.867	0.886	0.658	900
paper_clean	0.922	0.918	0.858	0.838	869
paper_contaminated	0.907	0.903	0.856	0.801	869
plastic_clean	0.856	0.852	0.867	0.701	884
plastic_contaminated	0.852	0.850	0.874	0.700	884

Glass is always the poorest category (mAP50: 0.697 0.716) in both clean and contaminated versions of the material. This is credited to the fact that glass containers are clear or semi transparent, and therefore the visual features are highly variable with background, level, and viewing angle resulting in extracting features that cannot be consistently obtained. All other categories of materials are above mAP50=0.85 with the best scores given by the cardboard and paper parts (0.907-0.925).

Comparative Summary: v1 vs v2

Table 4: v1 vs v2 System Comparison

Feature	v1 - MobileNetV2 (Multi-Task)	v2 -YOLOv8s (Detection)
Architecture	Transfer learning + dual head	Single-stage object detector
Objects per image	1 (image-level classification)	Multiple (bounding boxes)
Category performance	66.0% accuracy	mAP50: 0.855
Contamination output	87.2% binary accuracy	Unified 10-class label
Stain bias	7.7pp accuracy drop (confirmed)	Mitigated by unified classes
Training time	45 min / 18 epochs	4.5 hrs / 47 epochs
Inference speed	28 images/sec (35ms)	3.0ms per image
Framework	TensorFlow 2.x / Keras	PyTorch / Ultralytics
GPU (training)	RTX 3080 (reference)	RTX 4060 Laptop 8GB

V. Discussion

Synthetic Contamination Training Strategy:

The great contamination detection performance of the v1 model (87.2%, AUC: 0.932) with the exclusively synthetic training data confirms the programs of the contamination pipeline as the viable solution to the issue of the lacking labeled data. The main benefit of the pipeline compared to GAN-based methods is that the contamination labels are always accurate - all augmented images have a known contamination type, grade, and location, which cannot be guaranteed with the generative models.

The stain bias of 7.7 percentage points when using categories to classify is a significant result which is only relevant to this particular system. It indicates that any model that has been trained using contamination augmentation on only a single group of samples is prone to learn the appearance of stain as a category-discriminative attribute. It is a very subtle way of data bias that could not be apparent with aggregate accuracy measures. The use of stain bias diagnostics will have to be incorporated into the standard evaluation of such researches made by investigators with the help of similar methods.

Glass as a Continuing Dilemma:

Glass had the lowest score in terms of material of both v1 (f1: 0.71) and v2 (mAP50: 0.697-0.716). This is also the consistent weakness of glass as respects its basic visual characteristics: the background objects are more visible in the field of view, the effect of specular lighting on the scene changes radically with light sources and camera position, and partially filled containers have entirely different appearances than empty ones. These factors render glass extremely vulnerable to the domain gap between training images (which can be of studio quality) and deployment environments (changing light, mixed backgrounds, etc.). Specific glass classification interventions might be: (1) data augmentation with a variety of backgrounds and lighting conditions particular to glass; (2) data collection techniques which emphasize the glass containers at different levels of fullness and (3) the building architecture to emphasize edges and gradients of transparency instead of surface texture.

Implication of Practical Deployments:

With a v1 inference rate of 28 images per second (35ms/image) on an NVIDIA card, the speed can be integrated into the sorting systems of the conveyor running at 0.5-1.0 m/s, which is actually the typical speed of any small sorting system. With item sizes of 10-20cm, which are common in normal waste sizes, items can be captured in a persistent camera image only once or twice (100-200ms) which is sufficient to allow a complete classification and contamination assessment cycle to be completed. MobileNet V2 MobileNetV2 has a small backbone (3.5M parameters) that makes it possible to run on edge computing devices like boards of NVIDIA Jetson series, and create self-contained sorting units without relying on cloud inference.

v2 YOLOv8 system, which is more computationally intensive, offers the multi-object scene analysis ability that is useful when sorting a large number of items in a single camera frame as they can be in the field of view. The 3.0ms inference time per image Ultralytics is reported to be the unaccelerated forward pass through the model; a real end-to-end latency would be more inclusive of getting the image, preprocessing and actuator signal generation.

VI. Research Gaps and Future Directions

Identified Gaps in the Current Literature

Contamination Integration Gap: The overwhelming majority of waste classification systems treat contamination as either irrelevant or as a post-classification anomaly detection problem. Integrated, jointly optimized contamination-aware systems remain rare.

Dataset Scale and Realism: Available waste datasets (2,000–5,000 images) are insufficient for deep learning at scale. More critically, real-world conditions - cluttered backgrounds, mixed-item scenes, variable lighting, physical degradation - are poorly represented. Datasets exceeding 50,000 images with real-world contamination labels are needed.

Contamination Severity Quantification: Binary clean/contaminated labels do not capture the operational spectrum from lightly soiled (potentially recoverable through cleaning) to permeated/permanently contaminated (irrecoverable). Continuous severity scores or ordinal severity categories would enable more nuanced sorting decisions.

Domain Adaptation: Models trained on one recycling facility or dataset degrade significantly when deployed in a different environment. Robust unsupervised domain adaptation methods specific to waste imagery are needed.

Explainability and Operator Trust: Black-box predictions are insufficient for operational deployment where human operators must understand and trust system decisions. GradCAM visualization, attention maps, and calibrated confidence scores are needed.

Real Multi-Object Annotation: The v2 system currently uses whole-image bounding boxes as a training surrogate. Per-object annotations of cluttered scenes using tools such as Roboflow or Label Studio are required to realize true multi-object detection capability.

Future Research Directions

Several high-value research directions emerge directly from the limitations identified in this work:

Advanced Contamination Characterization: Developing contamination taxonomy with source identification (food, liquid, grease, biological), severity grading, and recoverability prediction. Spatial localization via segmentation masks would provide guidance for automated cleaning systems.

Physics-Based Synthetic Data: Replacing programmatic augmentation with 3D-rendered waste models with physically-based material rendering and fluid dynamics contamination simulation, enabling higher-realism training data with perfect ground truth labels.

Continual and Federated Learning: Online adaptation systems that update models with new waste types and contamination patterns encountered in deployment, combined with federated learning across facilities to aggregate improvements without sharing raw data.

Multi-Modal Sensing Integration: Fusion of visual classification with near-infrared spectroscopy for plastic subtype identification, hyperspectral imaging for material composition analysis, and acoustic sensing for density estimation.

Robotic Sorting Integration: Coupling the classification system with robotic arm grasp planning, trajectory optimization that avoids contaminated surfaces, and bin assignment logic based on both material category and contamination state.

VII. CONCLUSION

This paper has introduced a unified deep learning architecture of simultaneously categorizing and detecting contamination of waste materials, an otherwise important and understudied problem of automated recycling system architecture. It is shown that the operationally valuable information of contamination detection can be efficiently added to the waste classification pipeline at minimal extra computational cost and that the resultant system can obtain operationally useful information that cannot be made available by pure category classifiers.

The v1 MobileNetV2 dual-output model has proven that it is possible to implement multi-task category and contamination learning on 100 purposefully designed synthetic training images with 66.0% category classification and 87.2% contamination detection accuracy with an AUC of 0.932 doing so at a rate of 28 images per second. The magnitude and speed of contamination detection - category classification - contamination detection The magnitude and speed of the contamination detection process confirmed that the synthetic contamination pipeline produced visually differentiable and learnable signals of contamination. The weighted loss formulation (where α=1.0 in the case of category, and β=0.5 in the case of contamination) was useful in balancing the two objectives that does not permit either of the two to overshadow gradient update in training.

One of the more practically important results of this study is the stain bias diagnosis that has indicated a decline of 7.7 percentage points between the clean and contaminated image subset of category accuracy (69.8% vs 62.1). It shows that contamination-enhanced training, without a precise configuration, may add a small and yet significant amount of dataset bias whereby the model partially uses the lack of stain features as a feature to predict the category. This conclusion does not just apply to this system but also to any research on waste classification which uses contamination augmentation as training strategy. Stain bias diagnostic is an evaluation aspect which should be taken into consideration as a standard along with standard accuracy and F1 measures.

The fine-tuning step (freezing the top 30 MobileNetV2 layers with a lower learning rate of 5 x 10 -5) provided the slight improvement in category accuracy (+2.1 percentage points) and contamination accuracy (+1.7 percentage points). This small increase hints at the capacity of the model and depth of the learned features not being the primary limiting factor on v1 performance, but instead the size of the dataset and intrinsic visual complexity of waste images a conclusion which directly inspires the dataset augmentation and annotation effort advised in the future directions section.

v2 YOLOv8s system is a qualitative architectural development. v2 removes the representational sharing that allows stain bias in v1 by encoding the category and contamination as a single 10-class label instead of distinct model outputs. The system reached epoch 32 mAP50 of 0.855 and mAP50-95 of 0.851 through all 10 classes (cardboard and paper had the highest per-class scores, mAP50: 0.907- 0.925) and glass as the most difficult material (mAP50: 0.697-0.716). The weakness of the glass, which remains in both v1 and v2, points to one particular area where future actions of data collection and augmentation can be focused. The model was trained with epochs involving mosaic augmentation 1 40 and the model was subjected to simulated multi-object scenes, which also allowed the model to achieve the excellent overall detection performance even with the information available only as whole-image bounding boxes in the dataset.

Even the synthetic contamination pipeline has become a methodological contribution on its own. The three type and three severity system (light- (30-percent) probability, moderate- (50-percent) and heavy- (20-percent)-probability) produced a 15,000-image corpus based on an estimated 7,500 clean source images with accurately known contamination labels. The reproducibility and open-source nature of the pipeline implies that any researcher who has a waste image dataset of clean images can generate a training corpus with contamination added without using real contaminated samples, which has the benefit of reducing the threshold to entry to further development of contamination-aware systems.

The combination of the v1 and v2 systems is a continuum of an entry-level, single-image-based method to build a fully modeled multi-object detection pipeline. The two systems are entirely open-source, having data preprocessing, contamination generation, model training, fine-tuning, evaluation and inference scripts. The reported discrepancy between whole-image bounding box and per-object annotation gives a understood and effective direction on the upcoming row of such work. The v2 architecture using the real per-object annotations provided by scenes of cluttered multi-items can provide the object-level contamination evaluation that automated sorting facilities eventually need.

REFERENCES

[1] C. Shi et al., "A waste classification method based on a multilayer hybrid convolution neural network," Applied Sciences, vol. 11, no. 18, p. 8572, 2021.
[2] A. Jahanbakhshi et al., "Waste management using an automatic sorting system for carrot fruit based on image processing technique and improved deep neural networks," Energy Reports, vol. 7, pp. 5248–5256, 2021.
[3] M. Arun, D. Barik, and S. R. Chandran, "Exploration of material recovery framework from waste - A revolutionary move towards clean environment," Chemical Engineering Journal Advances, vol. 18, p. 100589, 2024.
[4] D. Gyawali et al., "Comparative analysis of multiple deep CNN models for waste classification," arXiv preprint arXiv:2004.02168, 2020.
[5] O. Adedeji and Z. Wang, "Intelligent waste classification system using deep learning convolutional neural network," Procedia Manufacturing, vol. 35, pp. 607–612, 2019.
[6] M. K. Hasan et al., "Smart waste management and classification system for smart cities using deep learning," in Proc. IEEE ICBATS, 2022.
[7] F. S. Alsubaei, F. N. Al-Wesabi, and A. M. Hilal, "Deep learning-based small object detection and classification model for garbage waste management in smart cities," Applied Sciences, vol. 12, no. 5, p. 2281, 2022
[8] K. Lin et al., "Toward smarter management and recovery of municipal solid waste: A critical review on deep learning approaches," Journal of Cleaner Production, vol. 346, p. 130943, 2022.
[9] S. P. Gundupalli, S. Hait, and A. Thakur, "A review on automated sorting of source-separated municipal solid waste for recycling," Waste Management, vol. 60, pp. 56–74, 2017.

[10] A. H. Vo, M. T. Vo, and T. Le, "A novel framework for trash classification using deep transfer learning," IEEE Access, vol. 7, pp. 178631–178639, 2019.
[11] N. K. Gyamfi et al., "Malware detection using convolutional neural network, a deep learning framework: comparative analysis," Journal of Internet Services and Information Security, vol. 12, no. 4, pp. 102–115, 2022.
[12] L. W. Ming et al., "AI as a driver of efficiency in waste management and resource recovery," International Transactions on Artificial Intelligence, vol. 2, no. 2, pp. 128–134, 2024.
[13] J. Li et al., "Automatic detection and classification system of domestic waste via multimodel cascaded convolutional neural network," IEEE Transactions on Industrial Informatics, vol. 18, no. 1, pp. 163–173, 2022.
[14] S. K. Pal et al., "Deep learning in multi-object detection and tracking: state of the art," Applied Intelligence, vol. 51, no. 9, pp. 6400–6429, 2021.
[15] J. Bobulski and M. Kubanek, "Deep learning for plastic waste classification system," Applied Computational Intelligence and Soft Computing, vol. 2021, p. 6626948, 2021.