Journal of Advanced College of Engineering and Management

Municipal Plastic Detection and Classification in Real Time using YOLOv9 and Custom CNN

[1]Prajwal Chaudhary, [2]Kamal Shrestha, [3]Pabin Khanal, [4]Jesis Upadhayaya, [5]Anup Shrestha

[1,2,3,4,5]*Department of Electronics and Computer Engineering, Thapathali campus, IOE*

Email: [1]prajwalchy25@gmail.com, [2]kamalshrestha17104@gmail.com, [3]kpabin022@gmail.com, [4]upadhyayjesis@gmail.com, [5]anup@tcioe.edu.np

DOI: 10.3126/jacem.v11i1.84527

Abstract

Plastic waste management is a critical global issue, with over 380 million tons of plastic produced annually, much of which pollutes the environment. Manual sorting of municipal plastic waste is labor-intensive and costly, creating the need for automated solutions. This paper presents a real-time system that detects and classifies plastic waste using deep learning. The system integrates YOLOv9 for detecting plastic items and a custom Convolutional Neural Network (CNN) for classifying them into four categories: Polyethylene Terephthalate (PET), Polypropylene (PP), High-Density Polyethylene (HDPE), and Polystyrene (PS). Our experiments show that YOLOv9, evaluated as an object detector, achieved a precision of 95.77%, a recall of 97.04%, and mAP@50–95 of 90.92%, while the CNN, evaluated as a classifier, achieved a precision of 98.41%, a recall of 98.42%, and an F1-score of 98.40%, demonstrating strong classification performance. These results indicate that the developed system provides an effective, scalable, and cost-efficient approach for automating plastic waste sorting, supporting improved recycling and sustainable waste management.

Keywords*—CNN, Plastic waste management, Recycling, Sustainable waste management, YOLOv9.*

1. INTRODUCTION

Plastics are a complex material with properties that are not fully understood by most people. Waste management, particularly municipal plastic waste, has become a major environmental challenge. Municipal plastic waste management has emerged as a significant environmental challenge in recent years. In 2024, the world is expected to generate 220 million tons of plastic waste, with 70 million tons ending up in nature. Municipal plastic waste, which includes household and commercial plastic products, contributes significantly to urban pollution. As plastics do not biodegrade, they accumulate in landfills and oceans, breaking down into harmful micro-plastics. The need for efficient and scalable recycling solutions has become critical.

One of the main reasons for such high levels of waste today is the ineffective and often careless way that materials are handled. This is where the 3R Initiative steps in, promoting the ideals of reducing, reusing, and recycling on a global scale. The goal is to build a society that gets the most out of its resources while generating as little waste as possible. Waste reduction starts with making intelligent choices, i.e., buying only what's needed and foregoing those that will end up in the landfill in a short while. Reusing gives products a longer life by using them more than once instead of throwing them away. Recycling transforms already used material into a new product, either during the manufacturing process or after consumers are through with them. Recycling at the consumer level entails processing the materials in such a way that they can be reused, generally with some change in their form. For recycling to take place efficiently, household waste has to be sorted out properly, for e.g., plastics, metal, organic waste, paper, and glass have to be segregated. Yet not everything within these groups can be recycled, and so the more advanced technology needs to come into play in order to sift them out and identify what can actually be reused. This is especially important with plastics, where slight differences in type will affect whether or not they will be recyclable.

This work tries to address this issue by developing a deep learning-based automated system for plastic waste management using YOLOv9, a powerful object detection model, while a custom-designed Convolutional Neural Network (CNN) model is used for classifying different types of plastics based on their visual features. YOLOv9 is chosen for its real-time processing capabilities and high accuracy in detecting objects within images, whereas a custom-designed Convolutional Neural Network (CNN) is used for the categorization of different types of plastic based on visual features. The system utilizes deep learning and image processing techniques to classify plastic waste such as PET, HDPE, PP, and PS.

2. RELATED WORK

Rahman et al. designed an automated waste management system that integrates deep learning with IoT technology. In this system, deep learning models classify waste, while IoT handles the practical disposal processes. Their study also compared various CNN models to identify the most efficient one, emphasizing the importance of optimizing for both accuracy and computational efficiency [1].

Hassan et al. (2020) investigated the use of advanced data augmentation techniques to improve dataset diversity. By employing methods like synthetic image generation and adaptive augmentation, they enhanced model generalizability, especially for underrepresented plastic types such as PP and PS. These techniques help mitigate the common issue of dataset imbalance in waste classification tasks [2].

Deep learning models, such as YOLO, ResNet, and SSD, have revolutionized waste detection by automating traditionally manual tasks. Models trained on datasets like TrashNet and TACO have demonstrated strong performance, with YOLOv5 and YOLOv8 excelling in real-time detection and handling challenges like occlusion and varying illumination. However, gaps persist, particularly the need for more comprehensive datasets and refined feature extraction techniques. Advances in segmentation and feature extraction are helping bridge these gaps, paving the way for practical waste management applications [3].

A state-of-the-art review published by IEEE addressed the complexities in plastic waste detection. It stressed the need for algorithms that combine real-world data with advanced techniques, such as transfer learning, to improve model performance and scalability. The study argued for the integration of adaptive systems that can better respond to dynamic environmental conditions [4].

Recent work in plastic waste management has shifted toward using deep learning and machine vision to overcome the limitations of traditional methods like near-infrared spectroscopy (NIRS), which struggles with plastics like PET and PET-G. For example, YOLOv8 has achieved over 91.7% accuracy by leveraging visual features such as texture and shape. These systems often combine automated sorting mechanisms, such as pneumatic sorting, to enhance efficiency, though challenges like lighting variability and dataset limitations remain. Nonetheless, advancements in deep learning and dataset quality provide scalable solutions for improving plastic recycling rates [5].

Ramos, Lopes, and Mendonça reviewed the use of machine learning in waste classification, emphasizing the effectiveness of CNN-based models, particularly YOLO and SSD. They suggested that the development of benchmark datasets would be crucial for advancing the field and facilitating comparisons across different approaches [6].

Guezouli et al. (2022) developed a system utilizing CNNs to detect and classify plastic waste, aiming to enhance recycling efficiency. It is trained on a vast dataset, achieving an impressive accuracy rate of 97%. This shows that there is a potential for neural networks to aid in plastic segmentation and recycling. However, the architecture they have used is rather simple and leaves room for exploration of better architectures. Our approach is to create a more lightweight model in order to assist real-time classification of plastics [7].

"A Deep Learning Approach to Manage and Reduce Plastic Waste in the Oceans" by El Zaar et al. (2022) explains the utilization of deep learning techniques to identify plastic waste from images in order to curb ocean pollution. The authors have applied two transfer learning techniques using CNNs: using pre-trained CNNs as a feature extractor with an SVM classifier and fine-tuning pre-trained CNNs. These approaches are validated on challenging object discovery datasets and texture classifying datasets. The paper illustrates that deep learning may be utilized in the task of managing plastic waste, but without particular measures of performance and straightforward information about datasets [8].

3. MATERIALS AND METHODS

A. System Overview

Our system comprises a camera and two models for plastic detection and classification, i.e., YOLOv9 and a custom CNN model. The proposed system takes real-time images as input from the camera. The input from the camera is passed on to the YOLOv9 frame by frame. The output is received, which contains bounding boxes for individual plastics detected in the image. The exact bounding boxes(ROIs) are cropped and resized to meet the input requirements of our CNN model. It classifies the images into individual plastic types. The outputs from both models are then combined to get the final output. The figure below illustrates our system block diagram.

In our system, YOLOv9 and the CNN work together but serve different purposes. YOLOv9 is used to detect and locate plastic objects in real time, while the CNN classifies the detected objects into their specific categories. These two models are connected in a single pipeline. To check the strength of our CNN, we compared it with well-known CNN models (ResNet50, MobileNetV2, and EfficientB0), which is shown in Table3.

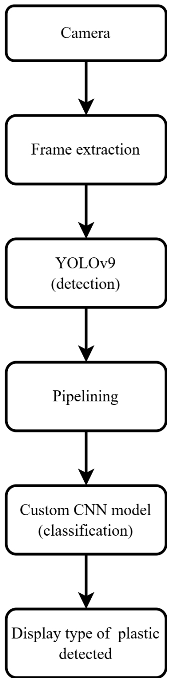


Fig.1. System block diagram

B. Datasets

The dataset used for plastic detection and classification is primarily sourced from WaDaBa (Waste Database), Kaggle open source datasets, and some from Roboflow. The WaDaBa datasets, developed by researchers J. Bobulski and J. Piatkowski [10], serve as the foundation for a dataset of plastic waste detection and classification, which includes high-quality images for machine learning tasks. This dataset contains images of plastics from four different classes, namely HDPE, PET, PP, and PS. There are a total of 3,960 images: 2,200 images of PET plastics, 600 images of HDPE plastics, 640 images of PP plastics, and 520 images of PS plastics. From Kaggle, we collected approximately 3,200 images of PET and HDPE plastic bottles, along with various bottle caps, captured in a real-world setting on a conveyor belt. The images were taken in an industrial environment, where the bottles and caps are moving along the belt. Additionally, images of plastics of the same classes with varying backgrounds were collected from the internet and added to the training dataset to improve recognition accuracy. This dataset consisted of about 4,018 images annotated into four different classes (Aluminium can, HDPE plastic, PET plastic, and UHT box. Out of these, we removed annotations (Aluminium can and UHT box) from the dataset and annotated them as the Null class, declaring that there are no relevant images in them. And then we merged this dataset into our project database.

C. Dataset Preprocessing

The images needed to be annotated before feeding the YOLO model for training. We used Roboflow for extensive annotation of 3,960 raw images from the WaDaBa dataset. Annotating involves the creation of bounding boxes to surround the object in the image. Then, the object class is provided, which in our case is plastic. We performed augmentation on the dataset to further improve the diversity and balance of the dataset. We used Roboflow to apply augmentations on these images, including vertical and horizontal flip, saturation changes between -25% and +25%, brightness changes between -15% and 15%, changes in the exposure in the range of +14% and -14% and random noise was introduced in up to 1.52% of the pixels. ImageDataGenerator was used for geometric and intensity-based augmentations, including rotations (±45°), width/height shifts (20%), shear, zoom (0.8–1.2), flips, and brightness adjustments (0.7–1.3), for our CNN model that used grayscale images. To prevent artifacts, we set fill_mode='nearest'. These enhancements increase robustness to changes in lighting, orientation, and object positioning, decrease overfitting, and improve model generalization. We split our dataset into three categories: train, valid, and test, with a percentage share of 70%, 20% and 10% during annotation and merging. YOLO is mainly used here for plastic detection because of its real-time object detection and localization capabilities, which were the basic requirement of this project. The architecture consists of three primary parts: the backbone, neck, and head. The backbone, typically a Convolutional Neural Network (CNN), is used to extract hierarchical feature maps from the input image, capturing low, mid, and high-level features. These feature maps are then passed to the neck, which aggregates them using structures like the Feature Pyramid Network (FPN). This combined representation is forwarded to the head, which handles both object classification and bounding box prediction. The head can be either a single-stage dense predictor like YOLO or SSD, or a two-stage sparse detector such as those from the R-CNN family.

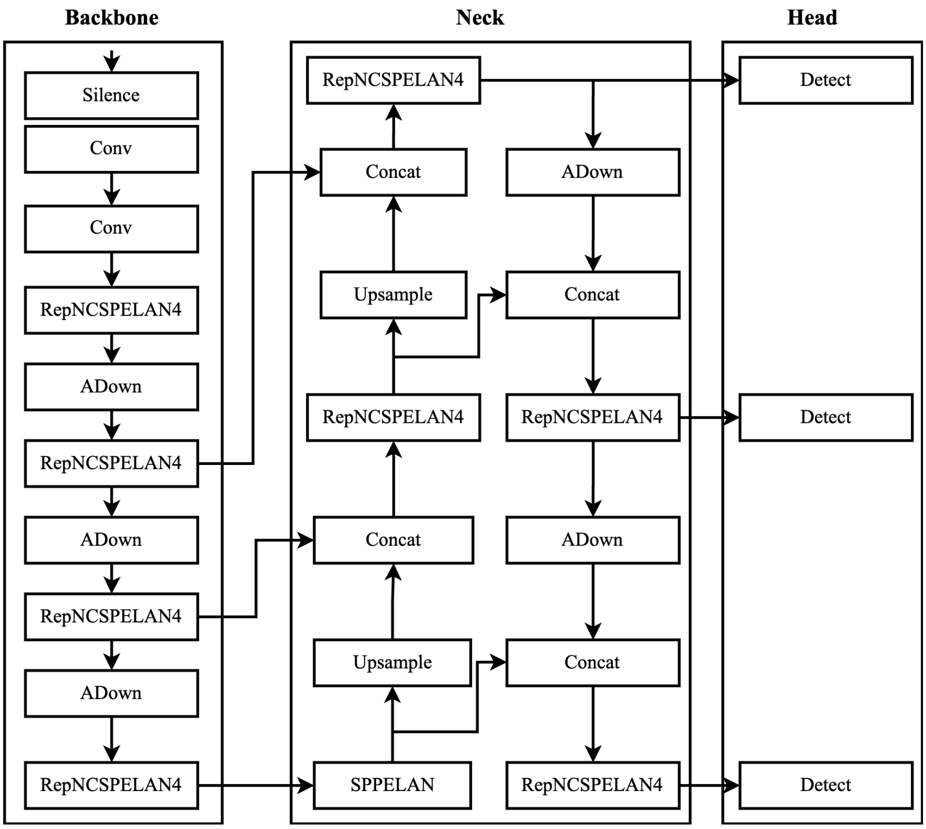


Fig.2. YOLOv9 Architecture [13]

D. Custom CNN for Classification

The CNN has been customized for plastic classification and label generation. It classifies the plastic image and generates different labels, i.e., PET, HDPE, PP, and PS, according to the result of classification. Our custom CNN is inspired by Residual Neural Networks and other modern CNN architectures. It has been divided into three parts, i.e., stem, backbone, and head. The stem is responsible for extracting low-level features. It makes use of a single convolutional layer followed by a batch normalization function and swish activation. The output is max-pooled to reduce the spatial dimension. The backbone makes use of residual and squeeze-and-excitation blocks to learn deeper features. Residual network helps with reducing the vanishing gradient problem and retaining valuable information. The Squeeze-and-Excitation block improves feature extraction by learning channel-wise attention. The head processes the extracted features and provides classification output. It makes use of global average pooling for spatial invariance as the features are averaged to a single value. Fully connected layers then learn to mix and emphasize the most relevant features for classification. Finally, the logits are converted to probabilities for classification.

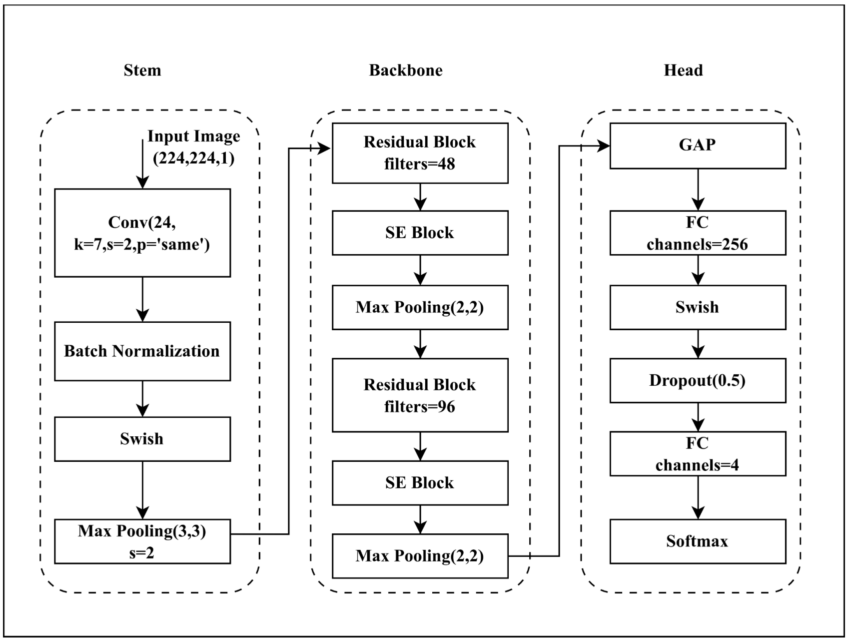
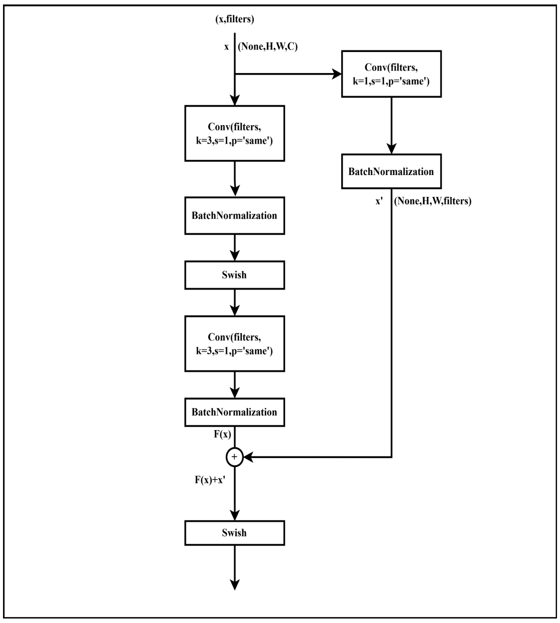


Fig.3. Residual Network-based Custom CNN Architecture

Residual block is used here to solve the vanishing gradient problem. The key idea behind residual networks is that some part of the input is left unoperated for better information retention while passing into further layers. Here, some of the output is processed, and some of it is passed directly to the output. This helps with information retention and better gradient flow. It uses two convolution layers with a kernel size of 3x3 for feature extraction. A 1x1 kernel is used in the skip (shortcut) connection to adjust the number of channels. The processed output is added to the output from the skip connection; thus, adjusting channels is needed to maintain consistency.

The SE block performs three main operations on the feature maps provided to the block: squeezing, excitation, and reshaping. The addition of this block in CNN networks improves the classification performance of the networks without adding any computational overhead. The squeeze block makes use of global average pooling to reduce the feature maps to a single representative value. If the input feature maps are H x W x C, they will get transformed into a vector 1 x 1 x C. The vector obtained after squeezing goes through two subsequent small fully connected layers. One layer reduces the dimension and another expands it. This operation helps highlight the most important features and suppress the ones that are less important. The original feature maps are now multiplied using the obtained vector after the excitation operation. This makes sure that only the most important features are highlighted and the less important ones are ignored.



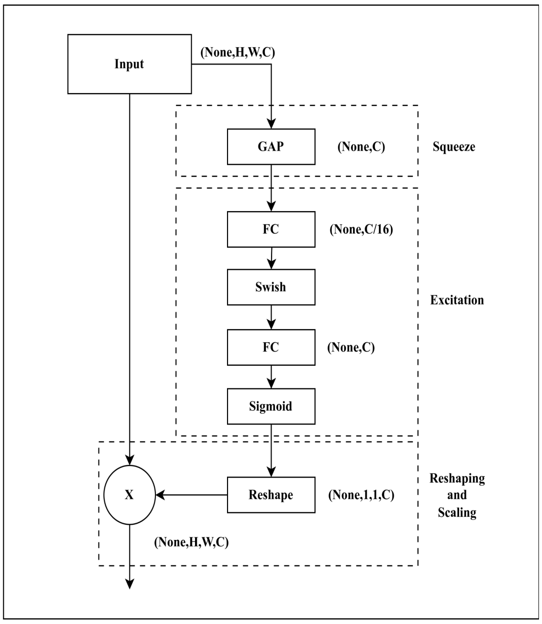


Fig.4. Residual Block

Fig.5. Squeeze and Excitation Block

E. System Flowchart

The video inference process is initiated by taking live frames with a camera. It views each frame in real time to detect and classify objects in each frame and displays results to the user. Initially, an OpenCV video capture object is created. This app utilizes the camera to initiate reading frames. All frames get resized to simplify processing by the computer system. The resized frame is passed through the YOLO object recognition system to detect and classify objects in the frame. Objects detected get clipped from the frame, and each clipped photo is processed and tagged by a Convolutional Neural Network (CNN) system. Results, in the form of bounding around objects and class names, get superimposed over the original frame along with the confidence score. This frame is saved and transmitted to the web client to be viewed. This process continues until stopped by the user. Following this, the video capture object is destroyed, and processing is complete.

The pipeline connecting YOLO and CNN is a fundamental component of the video inference process. YOLO is responsible for detecting and localizing objects within each frame. It produces a series of bounding boxes with the coordinates of detected objects. A corresponding region is cropped from the original frame for each detected object. The cropped image is converted to the grayscale color space according to the expectations of the CNN model. Next, the cropped image is resized to the required input size for the CNN model. Afterward, the pixel values of the cropped image are normalized to align with the training data. The preprocessed image is finally fed into the CNN model for class prediction. Annotating the original frame with bounding boxes and labels based on the predicted class label and confidence score ensures that the strengths of both YOLO (real-time object detection) and our Custom CNN (accurate classification) provide accurate yet efficient results in the video inference.

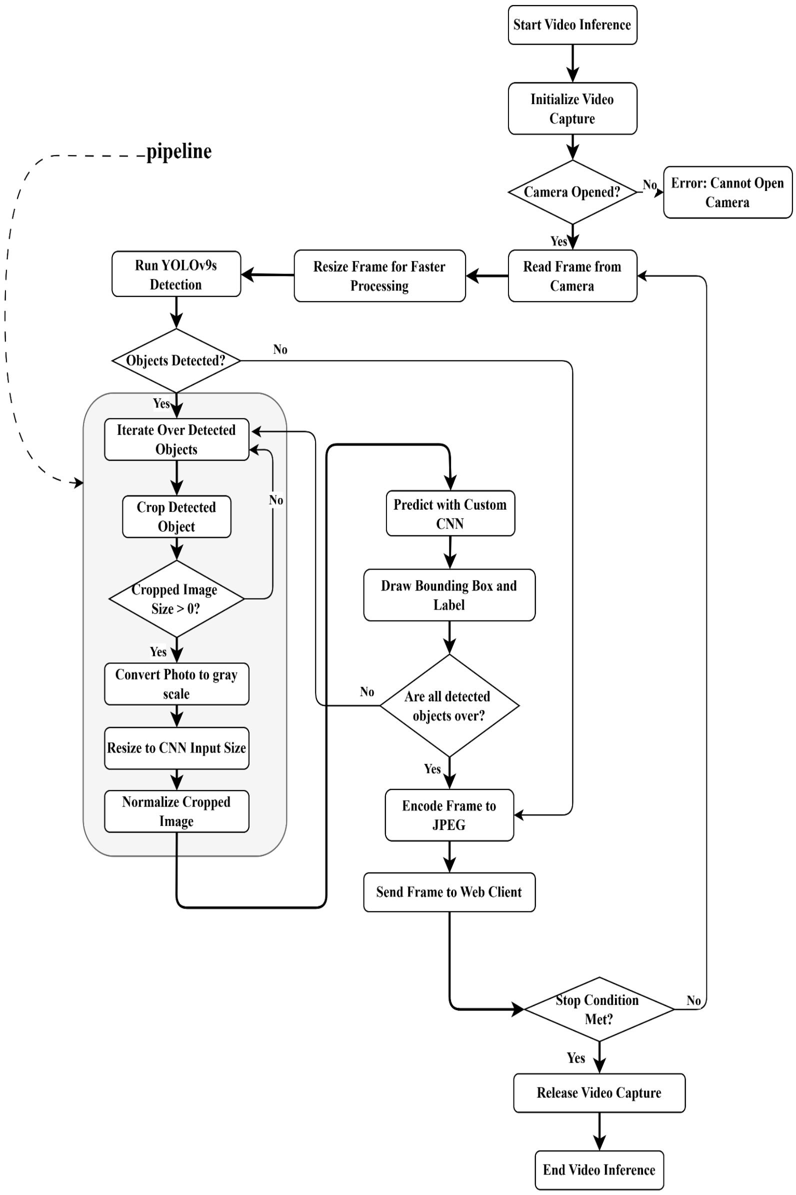


Fig.6. Video Inference Flowchart

4. RESULTS AND DISCUSSIONS

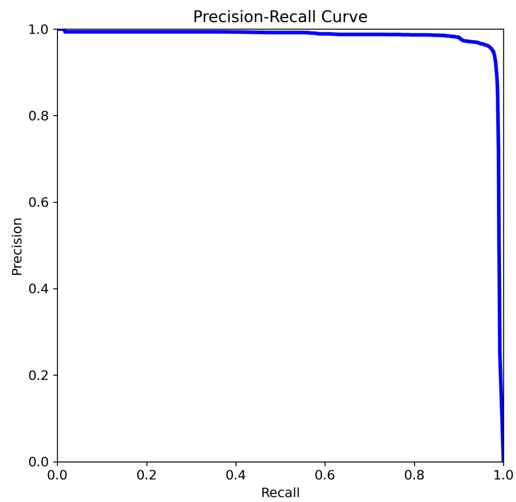
A. YOLOv9 Detection Performance

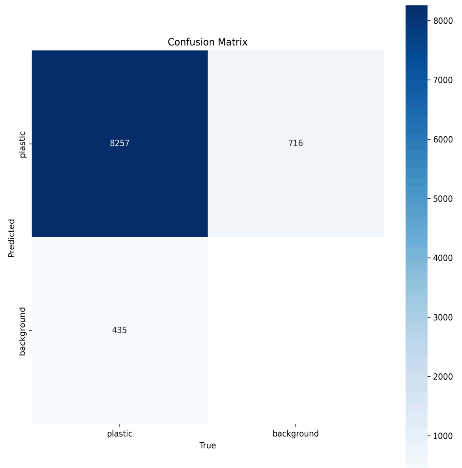
The YOLOv9s model was evaluated for real-time municipal plastic waste detection. **Table 1** summarizes the quantitative performance metrics results, demonstrating high detection reliability with strong localization accuracy across varying IoU thresholds, confirming its suitability for real-world municipal waste sorting tasks.

Table 1. Result Summary of YOLOv9s

Precision	Recall	f1-score	map@50	map@50-95
0.95774	0.97038	0.96402	0.97892	0.90924

Fig.7. illustrates the trade-off between precision and recall, with a consistently high area under the curve indicating robust detection performance across confidence thresholds.





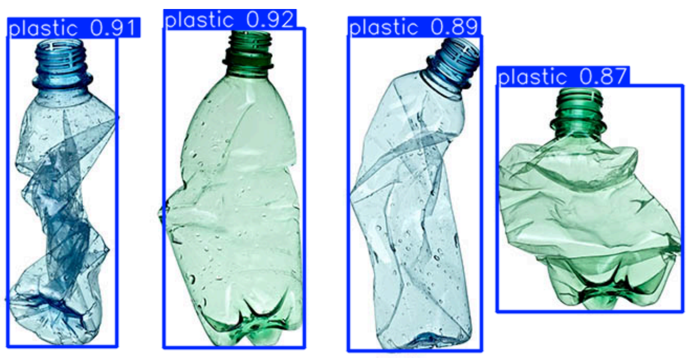


Fig.7. Precision Recall Curve **Fig.8.** Confusion Matrix

Fig.9. YOLOv9s Inference result

B. CNN Classification Performance

The custom Convolutional Neural Network (CNN) was used to classify plastics into PET, PP, HDPE, and PS categories. The training and validation loss curves show a steady convergence over time, suggesting that the model learned effectively without signs of overfitting. Supporting this, the confusion matrix reveals that the model correctly distinguishes between the different plastic types: HDPE, PET, PP, and PS, with very few misclassifications, indicating strong and reliable performance across all categories, as in **Table 2**.

Table 2. Result Summary of CNN

Metrics	HDPE	PET	PP	PS	Weighted average
Precision	0.984	0.9849	0.9783	0.9891	0.9841
Recall	0.982	0.9826	0.9909	0.9812	0.9842
F1-Score	0.9827	0.9837	0.9846	0.9851	0.9840
Accuracy	0.9840				

1. Comparison with Other CNNs

The convolutional neural network (CNN) models were benchmarked in a Kaggle environment with an NVIDIA T4 GPU, which was used to accelerate the computationally intensive training process. The primary objective was to classify grayscale images of plastic waste into four distinct categories: HDPE, PET, PP, and PS. All input images were standardized to a resolution of 224x224 pixels and fed to the models in batches of 32. A consistent training protocol was applied across all models, involving 10 epochs of training. The models were optimized using the Adam algorithm, with a learning rate that decayed exponentially to ensure stable convergence. Categorical cross-entropy was utilized as the loss function, and model performance was evaluated based on weighted averages of precision, recall, and F1-score, along with classification accuracy and inference time. This standardized approach was applied to both a custom-designed CNN and several established architectures, including MobileNetV2, ResNet50, and EfficientNetB0, to facilitate a direct and unbiased comparison of their performance, and the results are given in Table 3. The comparative analysis demonstrated that our custom CNN achieved performance comparable to larger, pre-trained models like MobileNetV2 and ResNet50. Notably, our model is significantly more efficient, containing only 192,189 parameters and exhibiting a faster inference time of 2.176 ms.

Journal of Advanced College of Engineering and Management

Table 3. Comparative Performance of CNN Models

Models	Parameters	Precision	Recall	F1-Score	Accuracy	Inference time(ms)
Custom CNN	192,189	0.75424	0.75339	0.75336	0.75339	2.176
ResNet 50	1,469,252	0.75214	0.74896	0.74949	0.74896	2.177
MobileN etV2	1,466,690	0.75174	0.74835	0.74863	0.74835	2.237
Efficient Net -B0	4,054,115	0.6847	0. 6733	0.6720	0.67331	2.370

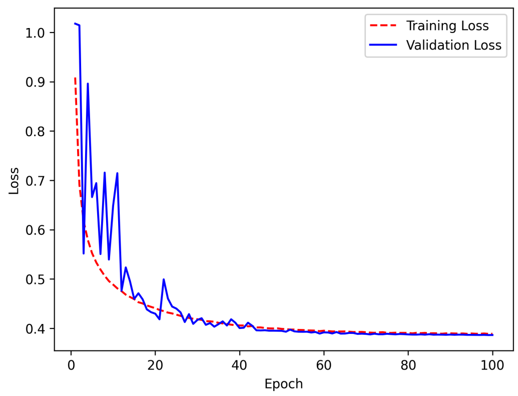


Fig.10. Training graph of CNN

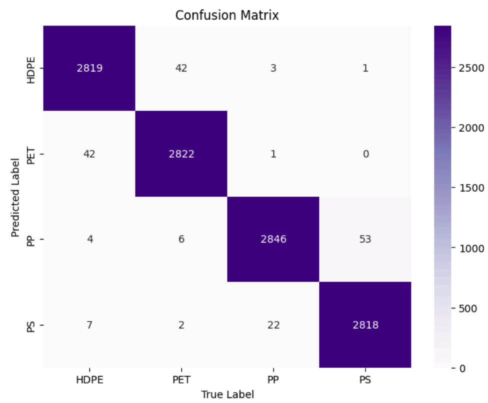


Fig.11. Confusion Matrix of CNN

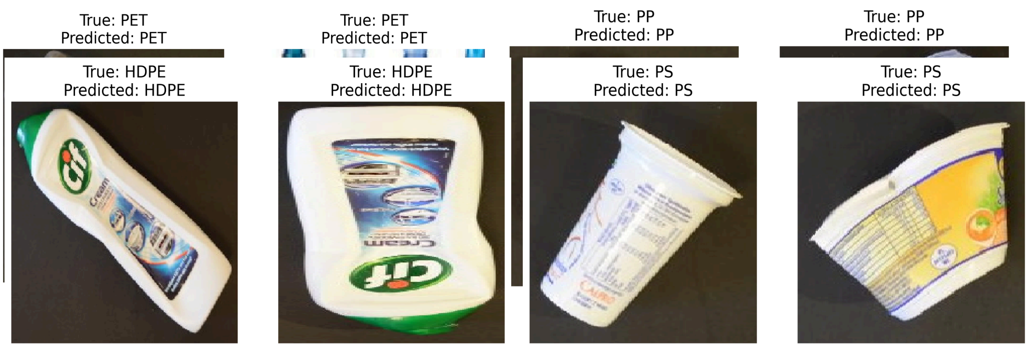


Fig.12. CNN's inference on images from different classes

C. Integrated System Output

The YOLOv9 detection and CNN classification models were integrated into a real-time web-based interface. Fig.13. shows a glimpse of real-time video inference through our integrated systems.

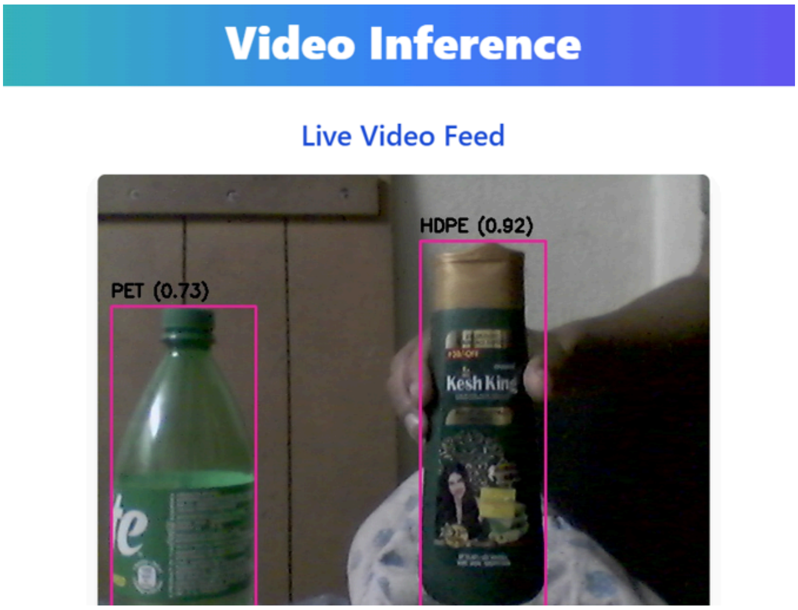


Fig.13. Integrated Inference of YOLOv9 and Custom CNN

5. CONCLUSION

In this study, a real-time plastic waste detection and classification system integrating YOLOv9 for object detection and a custom Convolutional Neural Network for municipal plastic type classification was built. The YOLOv9 model achieved a precision of 95.77%, a recall of 97.04%, and an mAP@50-95 of 90.92%, demonstrating robust detection performance across varying object scales and conditions. The CNN model achieved a test accuracy of 98.4%, enabling reliable classification of plastics into PET, PP, HDPE, and PS categories.

6. FUTURE WORK

Future work focuses on improving the system by expanding the dataset to include more plastic types and real-world waste images under diverse conditions, which would enhance generalization. Future work includes refining our CNN's architecture to improve feature extraction and make it more robust and efficient for plastic classification. There is room for more experiments in the model's architecture to achieve better results. We can implement a hardware system for plastic type classification on a conveyor belt. The real-time performance of our integrated model can be enhanced by converting it into ONNX, TensorRT, which will enable optimized execution across various hardware platforms with reduced inference latency. Additionally, model quantization and pruning can be done to reduce model size and speed up inference without losing accuracy. Such an optimized model can be used with Raspberry Pi or Coral TPU, which are low-power devices, making the system more practical for real-world applications.

7. ACKNOWLEDGMENTS

We are sincerely thankful to the Department of Electronics and Computer Engineering, Thapathali Campus, IOE, for encouraging us to do this work as a part of our minor project and providing us with the necessary resources.

We express our heartfelt gratitude to Asst. Professor Suwarna Lingden from Thapathali Campus, for his valuable guidance, feedback, and inspiration, which played a key role in the successful completion of this work.

Also, we would like to thank J. Bobulski and J. Piatkowski for providing us with access to the WaDaBa database, which helped us a lot in our dataset preparations.

Lastly, we thank our colleagues and families for their moral support and motivation during this academic endeavor.

REFERENCES

1. W. Rahman, R. Islam, A. Hasan, N. Bithi, M. Hasan, and M. M. Rahman, “Intelligent waste management system using deep learning with IoT,” Journal of King Saud University - Computer and Information Sciences, Sep. 2020.
2. M. Hassan et al., “Advanced data augmentation techniques for improving dataset diversity in plastic waste classification,” *Journal of Computer Vision and Machine Learning*, 2020.
3. E. Gothai, R. Thamilselvan, P. Natesan, M. Keerthivasan, K. Kabinesh, and D. K. Ruban, “Plastic waste classification using CNN for supporting 3 R's principle,” in 2022 *International Conference on Computer Communication and Informatics (ICCCI*), Coimbatore, India, pp. 01–07, 2022.
4. O. Tamin, “Machine learning for plastic waste detection: State-of-the-art, challenges, and solutions,” IEEE, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/9990703
5. J. Choi, B. Lim, and Y. Yoo, “Advancing plastic waste classification and recycling efficiency: Integrating image sensors and deep learning algorithms,” *Applied Sciences*, vol. 13, 2023.
6. E. Ramos, A. G. Lopes, and F. Mendonça, “Application of machine learning in plastic waste detection and classification: A systematic review,” Processes, vol. 12, no. 8, 2024. [Online]. Available: https://www.mdpi.com/2227-9717/12/8/1632

7. L. Guezouli, G. Aridj, and L. Saher, "CNN-based plastic waste detection system," *BOHR International Journal of Smart Computing and Information Technology*, vol. 4, pp. 79–85, 04 2023.
8. A. El Zaar, A. Aoulalay, N. Benaya, A. El Mhouti, M. Massar, and A. El Allati, "A Deep Learning Approach to Manage and Reduce Plastic Waste in the Oceans," E3S Web of Conferences, vol. 337, p. 00065, 2022. [Online]. https://www.e3s-conferences.org/articles/e3sconf/pdf/2022/03/e3sconf_icegc2022_00065.pdf
9. A. Zhang, Z. Lipton, M. Li, and A. Smola, "Dive into deep learning," 06 2021.
10. J. Bobulski and J. Piatkowski, "*PET waste classification method and plastic waste database - WaDaBa*," 01 2018, pp. 57–64.
11. J. Bobulski and M. Kubanek, "Deep learning for plastic waste classification system," *Applied Computational Intelligence and Soft Computin*g, vol. 2021, 05 2021.
12. C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," in *Computer Vision – ECCV 2024*, A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, and G. Varol, Eds. Cham: Springer Nature Switzerland, 2025, pp. 1–21. Available: https://github.com/WongKinYiu/yolov9.
13. P. Hidayatullah, N. Syakrani, M. R. Sholahuddin, T. Gelar, and R. Tubagus, "YOLOv8 to YOLO11: A Comprehensive Architecture In-depth Comparative Review," *arXiv preprint arXiv:2501.13400*, 2025. [Online]. Available: https://arxiv.org/abs/2501.13400