

Attention-Enhanced YOLOv8-Based Plastic Resin Classification for Intelligent Recycling Support

Varshini B. C.

Department of Computer Science and Engineering
East West Institute of Technology
Bengaluru, Karnataka, India

Prof. B. Sowmya

Professor, Department of Computer Science and Engineering
East West Institute of Technology
Bengaluru, Karnataka, India

Abstract—Accurate identification of plastic resin types is an important requirement in modern recycling systems because different polymers require different processing and recovery procedures. Conventional manual sorting is labour-intensive and is often unreliable when waste objects are contaminated, damaged, unlabelled, or visually similar. This study presents an image-based plastic resin classification framework built using an enhanced YOLOv8 classification architecture. The system categorizes plastic waste into seven classes: high-density polyethylene (HDPE), low-density polyethylene (LDPE), polyethylene terephthalate (PET), polypropylene (PP), polystyrene (PS), polyvinyl chloride (PVC), and Other. A lightweight YOLOv8n-cls classifier is considered as the baseline, while YOLOv8s-cls is adopted for the enhanced model to obtain richer visual representations. Transfer learning, class-balanced focal loss, label smoothing, MixUp, CutMix, brightness and contrast variation, and other image augmentation strategies are incorporated to improve generalization. Softmax probabilities are used to provide confidence scores and top-*k* alternatives, while Gradient-weighted Class Activation Mapping (Grad-CAM) is employed to visualize image regions that influence the model decision. Experimental evaluation on 2,097 images demonstrates an improvement from a reported baseline accuracy of 79.05% to 88.00% for the enhanced framework. The proposed approach provides a practical basis for smart recycling bins, operator-assisted sorting interfaces, and camera-based recycling systems where both classification accuracy and model interpretability are desirable.

Index Terms—plastic waste classification, YOLOv8, deep learning, transfer learning, focal loss, MixUp, CutMix, Grad-CAM, recycling, computer vision.

I. INTRODUCTION

Plastic has become an essential material in packaging, healthcare, agriculture, electronics, transportation, and manufacturing because of its low cost, durability, flexibility, and low weight. Unfortunately, these same properties also contribute to its persistence in the environment after disposal. Inadequately managed plastic waste may accumulate in landfills, drainage systems, rivers, oceans, and open environments, creating long-term ecological and economic concerns.

Recycling can reduce the environmental burden associated with discarded plastics; however, efficient recycling strongly depends on correct material separation. Common polymers such as polyethylene terephthalate (PET), high-density polyethylene (HDPE), low-density polyethylene (LDPE), polypropylene (PP), polystyrene (PS), and polyvinyl chloride (PVC) have different physical and chemical properties

and therefore require different recycling processes. Mixing incompatible polymers may reduce the quality of recycled material and increase subsequent purification and processing costs.

In many recycling facilities, resin identification is still performed manually using information such as resin identification codes, product shape, transparency, texture, colour, and packaging labels. Manual inspection becomes difficult when objects are crushed, stained, partly occluded, or missing their labels. Furthermore, several resin categories exhibit very similar visual characteristics. For example, HDPE and LDPE are both polyethylene materials, while PET and PP bottles may have comparable shapes and surface appearances.

Image-based machine learning provides an economical alternative because classification can be performed using standard RGB cameras. Convolutional neural networks (CNNs) automatically learn hierarchical visual features including edges, textures, colour patterns, and object-level structures. The YOLO family is widely recognized for computationally efficient computer vision and can support classification, detection, and segmentation applications.

Although a lightweight YOLO classifier is attractive for real-time deployment, resin-level classification remains challenging when the visual differences between categories are subtle. This study therefore develops an enhanced YOLOv8-based pipeline combining a stronger classification backbone with transfer learning, hard-example-aware loss functions, advanced augmentation, and explainability.

The principal contributions of this work are summarized as follows:

- A seven-class plastic resin classification framework based on YOLOv8s-cls is developed for recycling assistance.
- A preprocessing and augmentation pipeline incorporating resizing, normalization, brightness and contrast adjustment, flipping, rotation, MixUp, and CutMix is introduced.
- Class-balanced focal loss and label smoothing are used to improve learning from difficult samples and reduce excessive model confidence.
- Softmax-based confidence estimation and top-*k* predictions are incorporated to support operator-assisted decision making.

- Grad-CAM is integrated to provide visual evidence of the image regions contributing to individual predictions.
- The enhanced framework achieves a reported test accuracy of 88.00%, compared with 79.05% for the baseline YOLOv8n-cls configuration.

II. RELATED WORK

Plastic identification has traditionally relied on material sensing, spectroscopy, machine vision, and more recently deep learning. Near-infrared (NIR) spectroscopy is particularly useful because different polymers exhibit characteristic spectral responses [1]–[3]. Despite its effectiveness, NIR-based systems require dedicated sensors, calibration procedures, and controlled operating conditions, potentially increasing the cost of deployment.

RGB-image-based methods provide a more accessible alternative. Traditional image-processing techniques rely on manually selected features, whereas deep neural networks learn task-relevant features directly from training data. Architectures such as ResNet, EfficientNet, MobileNet, and YOLO have demonstrated strong image recognition capability in waste classification and related applications [5], [7], [8].

ResNet introduced residual learning to support deeper network optimization [16], whereas EfficientNet proposed compound scaling of depth, width, and image resolution [17]. YOLO-based approaches are attractive for recycling systems because of their computational efficiency and potential for real-time inference. Recent investigations have extended YOLO architectures to household-waste recognition and automated sorting applications [9].

Nevertheless, resin-level plastic recognition remains more difficult than general waste-category classification because samples belonging to different polymers can have nearly identical shapes, colours, and surface textures. The present study addresses this challenge by combining a YOLOv8 classification backbone with transfer learning, focal loss, MixUp, CutMix, label smoothing, confidence-aware prediction, and Grad-CAM visualization.

Table I summarizes representative approaches and their practical limitations.

III. PROPOSED SYSTEM

A. System Overview

The proposed system receives an RGB image containing a plastic waste object and assigns the image to one of seven resin categories: HDPE, LDPE, PET, PP, PS, PVC, or Other.

Images may originate from a curated dataset, recycling-station camera, conveyor-belt camera, smart bin, or web-based upload interface. The major processing stages are:

- 1) image acquisition,
- 2) preprocessing,
- 3) augmentation,
- 4) YOLOv8-based feature extraction,
- 5) seven-class prediction,
- 6) probability estimation,
- 7) top- k ranking, and

TABLE I
SUMMARY OF EXISTING APPROACHES

Method	Approach	Limitation
NIR Sorting	Spectral sensing	Requires dedicated sensors and controlled acquisition conditions.
Basic CNN	RGB image classification	May be sensitive to illumination, background variation, and limited datasets.
ResNet50	Transfer learning	Strong feature extraction but comparatively high computational demand.
EfficientNet	Transfer learning	Efficient architecture but performance depends strongly on model tuning.
YOLOv8n-cl	Lightweight classification	Fast inference but limited representation for visually similar polymers.
Proposed Model	YOLOv8s-cl with enhanced training	Designed to improve resin-level image classification while retaining practical efficiency.

8) Grad-CAM visualization.

The final output contains the predicted polymer category, classification confidence, alternative high-probability classes, and an explanation map indicating the spatial regions that contributed strongly to the prediction.

B. Baseline and Enhanced Models

YOLOv8n-cl is selected as the lightweight baseline configuration. Although its compact architecture is computationally attractive, its feature capacity may be insufficient when resin categories have subtle visual differences.

The enhanced system therefore employs YOLOv8s-cl. The model is initialized using pretrained weights and subsequently fine-tuned on the seven-class plastic dataset. The final classification layer is adapted to produce seven outputs.

Compared with the baseline model, the training process incorporates:

- transfer learning,
- class-balanced focal loss,
- label smoothing,
- MixUp augmentation,
- CutMix augmentation,
- geometric and photometric transformations, and
- Grad-CAM explainability.

IV. PREPROCESSING AND MATHEMATICAL FORMULATION

A. Image Representation

An RGB input image is represented as

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

where H and W denote image height and width, respectively, and the third dimension represents the RGB colour channels.

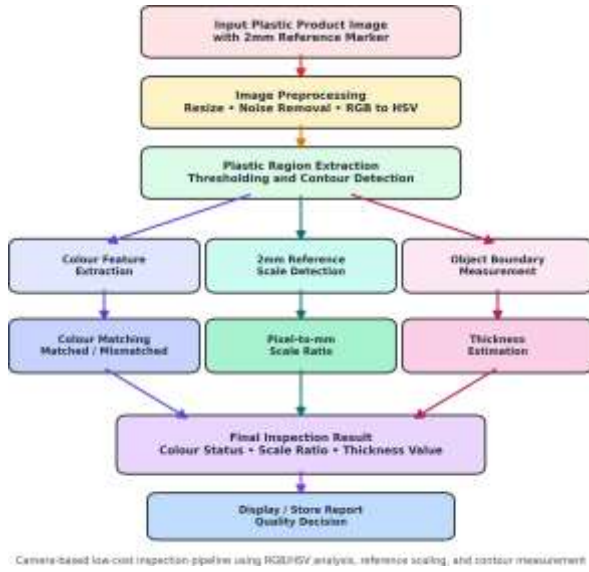


Fig. 1. Proposed YOLOv8-based plastic resin classification pipeline.

B. Image Resizing

Each input image is mapped to the spatial dimensions required by the model:

$$I_r = R(I, H_m, W_m), \quad (2)$$

where $R(\cdot)$ denotes the resizing operation and $H_m \times W_m$ is the network input resolution.

C. Pixel Normalization

Pixel intensity values are converted from the range $[0, 255]$ to $[0, 1]$ using

$$x' = \frac{x}{255}, \quad (3)$$

where x represents the original pixel intensity.

Channel-wise standardization can subsequently be expressed as

$$x'' = \frac{x' - \mu}{\sigma}, \quad (4)$$

where μ and σ denote the mean and standard deviation of the corresponding channel.

D. Brightness and Contrast Adjustment

To increase robustness against illumination variation, an image may be transformed as

$$I_c = \alpha I_r + \beta, \quad (5)$$

where α controls contrast and β adjusts brightness.

E. MixUp Augmentation

MixUp generates a synthetic sample from two training images x_i and x_j :

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j, \quad (6)$$

with the corresponding target

$$y \sim \lambda y_i + (1 - \lambda) y_j, \quad (7)$$

where $\lambda \in [0, 1]$, typically sampled from a Beta distribution [13].

F. CutMix Augmentation

CutMix replaces a region of one image using a patch taken from another training sample [14]:

$$\tilde{x} = M \odot x_i + (1 - M) \odot x_j, \quad (8)$$

where M denotes a binary spatial mask and $\odot$ represents element-wise multiplication.

The target is combined according to the proportion of pixels retained from each image:

$$y \sim \lambda y_i + (1 - \lambda) y_j. \quad (9)$$

V. PREDICTION AND LOSS FORMULATION

A. Feature Extraction

The preprocessed input I_p is passed through the YOLOv8s feature extractor:

$$F = f_{\theta}(I_p), \quad (10)$$

where θ denotes trainable network parameters and F is the learned feature representation.

B. Classification Layer

The extracted representation is transformed into class logits:

$$z = WF + b, \quad (11)$$

where W is the learnable weight matrix, b represents the bias, and

$$z = [z_1, z_2, \dots, z_C]. \quad (12)$$

For this study,

$$C = 7. \quad (13)$$

C. Softmax Probability

The posterior probability associated with class c is computed as

$$P(y = c | I_p) = \frac{\exp(z_c)}{\sum_{k=1}^C \exp(z_k)}. \quad (14)$$

The final predicted category is

$$\hat{y} = \arg \max_{c \in \{1, \dots, C\}} P(y = c | I_p). \quad (15)$$

For operator-assisted classification, the system additionally reports the k highest-probability outputs:

$$TopK(I_p) = \text{argsort}(P(y | I_p))_{1:k}. \quad (16)$$

D. Class-Balanced Focal Loss

Standard cross-entropy treats all correctly and incorrectly classified samples similarly. Focal loss reduces the contribution of easy examples and places additional emphasis on difficult samples [12].

For a single sample,

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t), \quad (17)$$

where p_t represents the predicted probability associated with the correct class, α_t is a class-dependent weighting factor, and γ is the focusing parameter.

For a multi-class classification problem,

$$L_{focal} = - \sum_{c=1}^C \alpha_c y_c (1 - p_c)^\gamma \log(p_c). \quad (18)$$

E. Label Smoothing

Label smoothing reduces excessive confidence by replacing hard one-hot labels with softened targets:

$$y_c^{LS} = (1 - \epsilon)y_c + \frac{\epsilon}{C}, \quad (19)$$

where ϵ is the smoothing coefficient.

VI. PROPOSED ALGORITHM

VII. DATASET AND EXPERIMENTAL SETUP

The dataset is organized according to the standard image classification directory structure containing separate training, validation, and test folders. Each folder contains seven class directories corresponding to HDPE, LDPE, Other, PET, PP, PS, and PVC.

The complete dataset contains 2,097 images. Of these, 1,468 images are used for training, 419 for validation, and 210 for final testing. The class-wise distribution is shown in Table II.

The implementation is developed using Python, PyTorch, Ultralytics YOLOv8, OpenCV, Pillow, NumPy, Pandas, and Matplotlib. Grad-CAM is incorporated for explanation generation. The YOLOv8s-cls model is initialized using pretrained parameters and fine-tuned using the plastic resin dataset.

Algorithm 1 Enhanced Plastic Resin Classification

- 1: Collect images belonging to HDPE, LDPE, PET, PP, PS, PVC, and Other categories.
- 2: Remove corrupted, duplicated, severely blurred, or incorrectly labelled samples.
- 3: Divide the dataset into training, validation, and test subsets using stratified sampling.
- 4: **for** each training image **do**
- 5: Resize the image to the required network dimensions.
- 6: Normalize image intensity values.
- 7: Apply brightness, contrast, geometric, and other augmentation operations.
- 8: Apply MixUp or CutMix when selected.
- 9: **end for**
- 10: Initialize a pretrained YOLOv8s-cls model.
- 11: Configure the classification output for seven categories.
- 12: Fine-tune network parameters using focal loss and label smoothing.
- 13: Evaluate performance using the validation subset.
- 14: Save the model producing the best validation result.
- 15: **for** each test image **do**
- 16: Generate class logits.
- 17: Convert logits to probabilities using Softmax.
- 18: Obtain $\hat{y} = \arg \max P(y = c | I_p)$.
- 19: Determine confidence and top- k alternatives.
- 20: Generate Grad-CAM visualization.
- 21: **end for**
- 22: Compute accuracy, precision, recall, F1-score, and confusion matrix.

TABLE II
DATASET DISTRIBUTION

Class	Train	Validation	Test
HDPE	210	60	30
LDPE	210	60	30
Other	210	60	30
PET	210	60	30
PP	210	60	30
PS	209	59	30
PVC	209	60	30
Total	1468	419	210

VIII. EVALUATION METRICS

Model performance is evaluated using accuracy, precision, recall, F1-score, macro-average F1-score, and the confusion matrix.

Accuracy is defined as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (20)$$

Precision is

$$Precision = \frac{TP}{TP + FP}. \quad (21)$$

Recall is

$$Recall = \frac{TP}{TP + FN} \quad (22)$$

The F1-score is the harmonic mean of precision and recall:

$$F1 = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (23)$$

For C classes, the macro-averaged F1-score is

$$F1_{macro} = \frac{1}{C} \sum_{c=1}^C F1_c \quad (24)$$

Top- k accuracy can be expressed as

$$TopKAccuracy = \frac{1}{N} \sum_{i=1}^N I[y_i \in TopK(\hat{p}_i)], \quad (25)$$

where N represents the number of test samples, y_i is the correct category, and $TopK(\hat{p}_i)$ denotes the set of k classes receiving the largest predicted probabilities.

IX. RESULTS AND DISCUSSION

The baseline YOLOv8n-cls configuration produced a reported test accuracy of 79.05%. Although the compact network provides efficient inference, its errors were concentrated primarily among visually similar resin categories.

The enhanced YOLOv8s-cls framework achieved a reported test accuracy of 88.00%, corresponding to an improvement of 8.95 percentage points relative to the baseline. The result suggests that the additional representational capacity of YOLOv8s, combined with hard-example-aware optimization and stronger augmentation, provides improved discrimination for resin-level classification.

Table III provides a comparison among the evaluated architectures.

TABLE III
MODEL PERFORMANCE COMPARISON

Model	Classes	Accuracy
Basic CNN	7	72.00%
ResNet50	7	76.00%
EfficientNet-B0	7	82.00%
Proposed YOLOv8s-cls	7	88.00%

The class-wise performance reported for the proposed model is presented in Table IV.

The results indicate that PS and PVC are comparatively easier to recognize, whereas HDPE and LDPE remain more difficult. This behaviour is reasonable because HDPE and LDPE belong to the polyethylene family and may exhibit similar colours, textures, surface finishes, and product geometries.

PET and PP also produce occasional confusion because transparent or semi-transparent containers from both categories can exhibit similar visual patterns. Consequently, an RGB-only model may not always capture material-level differences that are obvious through spectral analysis.

TABLE IV
CLASS-WISE PERFORMANCE OF PROPOSED MODEL

Class	Precision	Recall	F1
HDPE	0.86	0.87	0.86
LDPE	0.85	0.83	0.84
PET	0.88	0.90	0.89
PP	0.87	0.86	0.86
PS	0.91	0.93	0.92
PVC	0.90	0.90	0.90
Other	0.89	0.87	0.88
Macro Average	0.88	0.88	0.88
Weighted Average	0.88	0.88	0.88

The reported confusion matrix is provided in Table V.

The confusion patterns further demonstrate the difficulty of separating HDPE from LDPE and PET from PP using visual appearance alone. Future systems could therefore combine RGB classification with resin-code recognition or spectroscopy to increase confidence for ambiguous samples.

A. Explainability Using Grad-CAM

Prediction accuracy alone is insufficient for a recycling support system when operators need to understand why a classifier has produced a particular output. Grad-CAM generates a heat map from gradient information and highlights image regions that strongly influence the selected class [15].

This visualization can reveal whether the network is responding to the plastic object or relying on irrelevant background features. Such information is valuable for detecting spurious correlations and increasing confidence in human-assisted sorting scenarios.

B. Web-Based Recycling Support Interface

The trained model is incorporated into a Flask-based interface through which users can upload an image of a plastic item. The system produces the predicted resin category, confidence value, ranked alternatives, and recycling-related guidance.

Figures 2 and 3 may be replaced with the corresponding screenshots from the developed application.



Fig. 2. Example interface output for PVC classification with confidence and ranked predictions.

TABLE V
CONFUSION MATRIX OF PROPOSED YOLOV8S-CLS MODEL

True/Pred.	HDPE	LDPE	Other	PET	PP	PS	PVC
HDPE	26	2	0	1	0	0	1
LDPE	3	25	0	0	1	1	0
Other	0	1	26	1	1	0	1
PET	1	0	0	27	2	0	0
PP	0	1	1	2	26	0	0
PS	0	0	1	0	0	28	1
PVC	0	0	1	0	0	1	28

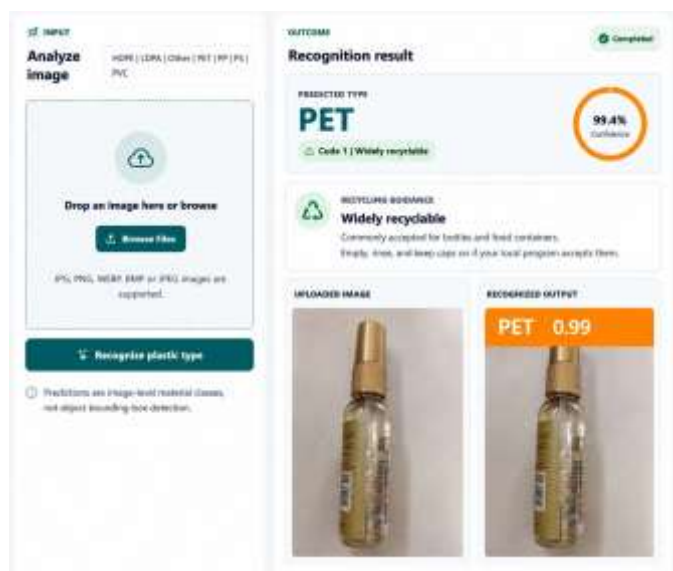


Fig. 3. Example interface output for PET classification with recycling guidance.

X. LIMITATIONS AND FUTURE WORK

The proposed classifier is primarily based on RGB appearance. Therefore, polymers having nearly identical visual properties may remain difficult to distinguish. The dataset is also relatively limited in comparison with the variability encountered in industrial recycling environments.

Real waste streams can contain partially broken objects, dirt, liquid residue, stains, labels, overlapping materials, complex backgrounds, and substantial illumination variation. If such conditions are insufficiently represented during training, performance may decrease when the model is deployed outside the laboratory environment.

Future research may therefore consider:

- collecting a substantially larger dataset directly from recycling facilities;
- evaluating the system using independent external test datasets;
- incorporating automatic resin identification code recognition;
- combining RGB imagery with near-infrared or hyperspectral measurements;

- extending image classification to multi-object detection and instance segmentation;
- introducing uncertainty thresholds that refer low-confidence samples for manual inspection;
- evaluating inference latency, memory usage, and energy consumption on embedded hardware;
- integrating the classifier with conveyor-belt actuators for automated physical sorting.

XI. CONCLUSION

This study presented an enhanced YOLOv8-based framework for image-level identification of seven plastic resin categories: HDPE, LDPE, PET, PP, PS, PVC, and Other. The method combines a YOLOv8s-cls backbone with transfer learning, focal loss, label smoothing, MixUp, CutMix, Softmax-based confidence estimation, and Grad-CAM visualization.

Relative to the reported 79.05% accuracy of the lightweight YOLOv8n-cls baseline, the enhanced model achieved a reported accuracy of 88.00%. The model provides not only the predicted resin class but also confidence information, ranked alternatives, and a visual explanation, making it more suitable for operator-assisted recycling applications than a classification-only pipeline.

Although RGB-based recognition cannot completely resolve polymers with extremely similar appearances, the findings demonstrate the potential of deep-learning-assisted resin classification for low-cost recycling applications. Future development should focus on larger real-world datasets, multimodal sensing, resin-code recognition, edge deployment, and integration with automated sorting systems.

REFERENCES

- [1] S. Serranti and G. Bonifazi, "Techniques for separation of plastic wastes," in *Use of Recycled Plastics in Eco-Efficient Concrete*. Woodhead Publishing, 2019, pp. 9–37.
- [2] M. Rani, N. Marchesi, L. Federici, C. Rovelli, I. Alessandri, F. Vassalini, G. Ducoli, L. Borgese, and A. Zacco, "Miniaturized near-infrared spectrometer in plastic waste sorting," *Materials*, vol. 12, no. 17, p. 2740, 2019.
- [3] Q. Duan and J. Li, "Classification of common household plastic wastes combining multiple methods based on near-infrared spectroscopy," *ACS ES&T Engineering*, vol. 1, no. 7, pp. 1065–1073, 2021.
- [4] T. Sousa, A. Matos, and J. Oliveira, "Deep learning for plastic waste classification," in *Proc. Int. Conf. Intelligent Systems and Data Science*, 2022.

- [5] S. Majchrowska, A. Mikołajczyk, M. Ferlin, Z. Klawikowska, M. A. Plantykowski, A. Kwasigroch, and K. Majek, "Deep learning-based waste detection in natural and urban environments," *Waste Management*, vol. 138, pp. 274–284, 2022.
- [6] S. Singh, S. Hait, and A. Thakur, "Hyperspectral imaging-based classification of post-consumer thermoplastics," *Resources, Conservation and Recycling*, 2023.
- [7] J. Choi, J. Lim, and S. Yoo, "Advancing plastic waste classification and recycling efficiency: Integrating image sensors and deep learning algorithms," *Applied Sciences*, vol. 13, no. 18, p. 10224, 2023.
- [8] S. Kunwar, B. R. Owabumoye, and A. S. Alade, "Plastic waste classification using deep learning: Insights from the WaDaBa dataset," *arXiv preprint arXiv:2412.20232*, 2024.
- [9] A. Arishi, "Real-time household waste detection and classification using YOLOv8 deep learning model," *Applied Sciences*, 2025.
- [10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 779–788.
- [11] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," 2023.
- [12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Computer Vision*, 2017, pp. 2980–2988.
- [13] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations*, 2018.
- [14] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Computer Vision*, 2019, pp. 6023–6032.
- [15] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Computer Vision*, 2017, pp. 618–626.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [17] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Machine Learning*, 2019, pp. 6105–6114.
- [18] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019.