Waste Classification Using Improved Deep Learning Method

Jianchun Qi, Minh Nguyen and Wei Qi Yan
Auckland University of Technology, 1010 New Zealand

ABSTRACT

In this book chapter, YOLOv8, the effective version in the YOLO series, is modified through data augmentation, context strategies, and an advanced attention mechanism. These modifications aim to primarily improve the quality of waste dataset and the classification accuracy of small objects within given waste classes, thereby boosting the overall performance of the model. The waste data is classified into four classes, and 1,000 waste images were labelled for model training. Upon evaluation, the classification accuracy of the improved model reached 85.6%. The effectiveness of these improvements was further substantiated through ablation studies.

Keywords: YOLOv8, data augmentation, context information, attention mechanism, waste classification

INTRODUCTION

Along with the progress of economy, the amount of waste generated also gradually increases. However, environmental protection and economic development are closely related to whether the waste is effectively disposed of. If waste is not properly disposed of, such as in random landfills, harmful chemical substances will be produced (Chen et al., 2020). It will pollute the air, soil, and water, damage the environment, and thus affect human health. Besides, if the recycling of waste is high, it will save costs and natural resources, thus successfully generating economic benefits. Therefore, taking active and effective measures for waste disposal and improving the recycling of wastes will have a positive impact on the natural environment and economic development. Thus, it is necessary to classify different waste types, such as hazardous waste and recyclable waste. At present, waste classification is mainly concentrated in the reuse collection stations, which are semi-automated and semi-manual that have the problems of low sorting efficiency and poor working environments.

Visual object detection has made remarkable progress in recent years, it is essential to use deep learning methods and computer vision as well as robots to replace manual labor for an automated classification. However, the accuracy of automated waste classification is lower than that of other classification tasks, such as fruit classification. It is speculated that there are two reasons for this difference in performance:

- The waste dataset is with low quality;
- There are a number of small wastes in the waste category, such as fruit pits and batteries, which appear as small objects in the same image with glass bottles.

The dataset in deep learning is an important factor influencing the training process of deep learning algorithms, and low quality of the dataset can seriously affect the accuracy of model predictions. A high-quality dataset should have three characteristics: Observable features, a sufficient amount of data, and uniform distribution. Therefore, how to improve the model performance by using the quality of the waste dataset is one of the research objectives of this book chapter. Besides, improving the classification accuracy of small waste objects is also an important task of this book chapter. The blurry and noisy features that can be extracted from small objects in images badly affect the accuracy of object classification, which also influence waste classification. Taken the Deformable DETR model as an example, this model was designed to improve the problem of slow convergence of DETR and inaccurate classification of small objects (Zhu et al., 2020). Although it achieves an accuracy 28.8% in small object classification, it is still much lower than the accuracy of medium-sized object classification and large object classification, which are 49.2% and 61.7%, respectively. The research work on small object classification has progressed slowly related to visual object classification, and even networks that are adopted at small object classification have a huge performance gap in detecting small and medium-sized or large targets. Finally, an attention mechanism is adopted that helps the network to focus on important information of the input images.

Therefore, in this book chapter, YOLOv8 is employed as the main model to classify wastes by using data augmentation methods, contextual information, and attention mechanism as a way to improve the accuracy of waste classification. The waste is classified into four classes according to the criteria, i.e., recyclable waste, dry waste, wet waste, and hazardous waste. Overall, the main contributions of this book chapter are listed as follows:

- The improved YOLOv8 model for waste classification model is achieved through data augmentation, contextual information, and attention mechanism with a benchmark. The classification accuracy is up to 85.6%;
- The effectiveness of the improved YOLOv8 model and the impact on the performance have been explored through ablation studies;
- The waste classification model is proposed and a garbage dataset meeting the experimental requirements is collected, thus improving the garbage classification.

In this book chapter, related work is introduced in Section 2, our research methodology is presented in Section 3, the analysis of the results is described in Section 4, and conclusion is shown in Section 5.

RELATED WORK

In recent years, waste classification algorithms have been extensively developed (Rabano et al., 2018) (Abdu & Noor, 2022) (Yan, 2021). A system based on YOLOv3 for real-time classification of wastes in video streams was proposed, achieving an accuracy of 68% (De, Ladogana & Macchiarulo, 2020). YOLO utilized large-scale convolutional kernels and dense convolutional blocks to increase the perceptual field of the model, enhancing the feature sensitivity of both shallow and deep semantics to improve waste classification accuracy (Lun et al., 2023). Additionally, the ResNet-34 model was also applied to waste classification, and an automatic classification bin was designed, including its hardware structure (Kang et al., 2020). Although exceptional results were obtained for waste classification of normal objects, research work on the detection of small objects in waste classification is scant, and the accuracy remains significantly lower than that of normal size of waste objects, which greatly affects the overall accuracy of waste classification. The research challenges for small object detection can be divided into four aspects:

- Small objects in digital images have few effective features and low resolution.
- Small object images are not easily labeled and much susceptible to noise, resulting in the lack of small object datasets (Chen et al., 2022).
- Because of the location of small objects, the network model makes prediction with even one pixel offset, which makes the localization of small objects difficult and has a huge impact on IoU value (Chen et al., 2022).
- The small object image is minuscule, which is prone to aggregation. When the network model performs prediction, the correct small object borders may be ignored due to non-maximum suppression, and the model convergence is difficult (Heo et al., 2021).

In response to these difficulties, advanced deep learning methods are utilized to improve the accuracy of small object classification. Contextual learning, a general and effective strategy, leverages the dependencies between detection objects and scenes or between objects themselves to extract contextual feature information. An effective contextual method is proposed, which infers the inherent semantic and spatial layout relationships among visual objects, effectively addressing the under-detection of small objects (Fu et al., 2020) (Yan, 2019). The strategy of multiscale learning is also popular because it retains the rich detail in shallow features and the semantic information in deep feature maps, enhancing detection. For instance, to mitigate the issue of feature loss in small object models, a new real-time detection algorithm employs skip connections and upsampling methods to extract features, significantly improving small object detection performance (Nayan, Saha & Mozumder, 2020) (Nguyen & Yan, 2023). Additionally, the anchor-free mechanism proves effective for small objects, which occupy a small area in images; transitioning from anchor mechanisms to key point estimation reduces the complexity of hyperparameters and improves performance (Liu, Pan & Yan, 2022) (Fu et al., 2022).

Data augmentation is another essential technique that increases the informational value of limited data without increasing the quantity (Liu, Yan & Kasabov, 2023) (Qi, Nguyen & Yan, 2024). Convolutional neural networks, generally are invariant to image size and shifts, benefit from data augmentation, which helps them recognize the same object under various positions and scales. Currently, data augmentation falls into two categories: Single-sample enhancements like Flip, Rotation, Random Cropping, Elastic Deformation, Scaling, Color Transformation, Jitter, and Noise Injection; Multisampling enhancements like Mixup and Sample Pairing (Inoue, 2018) (Qi, Nguyen & Yan, 2023) (Zhang et al., 2018); and unsupervised, represented by methods like GAN and Auto Augmentation (Karras, Laine & Aila, 2019) (Cubuk et al., 2019). By employing data augmentation, deep learning models are able to avoid irrelevant features, thus improving overall performance.

The attention mechanism is becoming increasingly important in deep learning and is widely employed in the fields such as computer vision, natural language processing, and speech recognition. Integrating attention mechanisms into deep learning models significantly enhances both interpretability and performance. The classic attention mechanism, CBAM, computes attention maps across two dimensions—space and channel—and integrates them for adaptive learning, serving as a general, lightweight module (Woo et al., 2018). A newer attention module, Coordinate Attention, generates spatially selective attention maps by incorporating location information into channel attention, outperforming CBAM and enhancing performance in both object detection and segmentation tasks (Hou, Zhou & Feng, 2021).

METHODOLOGY

The role of data augmentation is closely related to the performance of visual object detection (Xia, Nguyen & Yan, 2023). Due to the small waste dataset and low waste dataset quality, using data augmentation can effectively improve the waste data quality and enhance the model performance in the following four aspects:

- Preventing the model from learning excessive feature information unrelated to the detection object, which can lead to overfitting (Guo et al., 2020);
- Improving the generalization ability of the model by increasing the number of training data samples;
- Reducing the proportion of unbalanced training data;
- Decreasing the sensitivity of the model and enhancing its robustness.

Therefore, in this book chapter, rotating, cropping, and colour transformation operations are mainly applied to improve the quality of labelled data. Besides, mosaic method is chosen to perform data augmentation on unlabeled data (Bochkovskiy et al., 2020). Mosaic is chosen because it utilizes four random images to obtain a new image after stitching, and each image corresponds a video frame.

Compared with Cutmix and Mixup, Mosaic not only increases data diversity but also enriches the image background, effectively adjusts the batch size, and provides better calculations of variance and mean values. Moreover, it avoids the interference from non-informative pixels during the training process, thus enhancing model performance. The examples of data augmentation strategies applied in this book chapter are illustrated in Figure 1 and Figure 2.

Figure 1. Visualization of data augmentation. The first row is the original image. The remaining ones from second column to the bottom are: Rotating, cropping, and color Transformations.

Figure 2. Visualization of data augmentation: Mosaic.

Network structure

Compared with YOLOv5 (Zhu et al., 2021), YOLOv8 has made a series of significant improvements. In the backbone network, YOLOv8 continues the CSP concept by utilizing the SPPF module but replacing the C3 module with the C2F module, applies two 3×3 convolutions to reduce the resolution by using a factor 4 and achieves lightweight. Then, YOLOv8 removed all the convolutional machine structures in the PAN-FPN upsampling stage of YOLOv5. In the neck and head stages, YOLOv8 introduces Decoupled-Head, eliminates the objectness branch, and shifts from anchor-based to anchor-free. YOLOv8 also adopts Binary CrossEntropy Loss (BCE Loss) for classification, and takes use of CIoU Loss and Distribution Focal Loss (DFL) for regression. Additionally, YOLOv8 utilizes a dynamic Task-Aligned Assigner for the matching (Redmon et al., 2016). In this book chapter, YOLOv8 model has been selected as the baseline, the model performance is enhanced by focusing on three aspects. Figure 3 illustrates the overall structure of the improved YOLOv8 model.

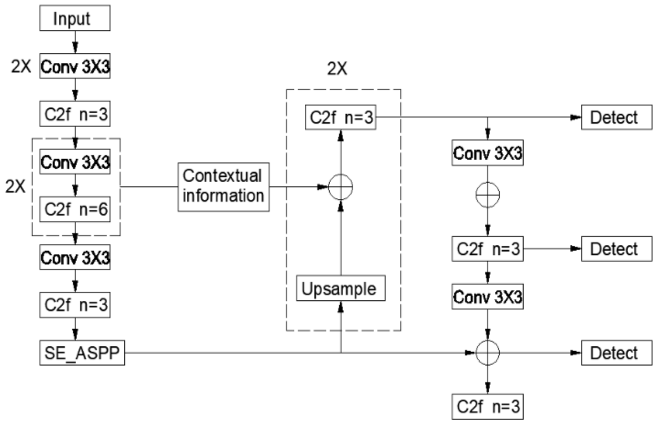


Figure 3. The structure of the improved YOLOv8 model.

Hence, a contextual information module is incorporated into the model to extract higher-level abstract features from the pixels surrounding small objects like fruit pits and batteries in waste images. In waste classification, small object detection typically relies on low-level features, which lack semantic richness despite their high resolution and detailed information. Conversely, while high-level features provide stronger semantic information, they suffer from low resolution and weak detail perception.

Therefore, to efficiently leverage the advantages of both, a feature fusion method is chosen to be implemented, as depicted in Figure 4. Feature fusion combines the information from both shallow and deep features, achieves a complementary balance that enhances the robustness and accuracy of the improved model in detecting small waste objects.

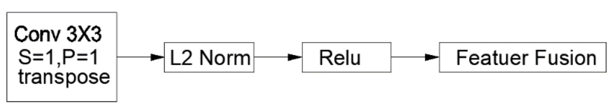


Figure 4. The structure of contextual information.

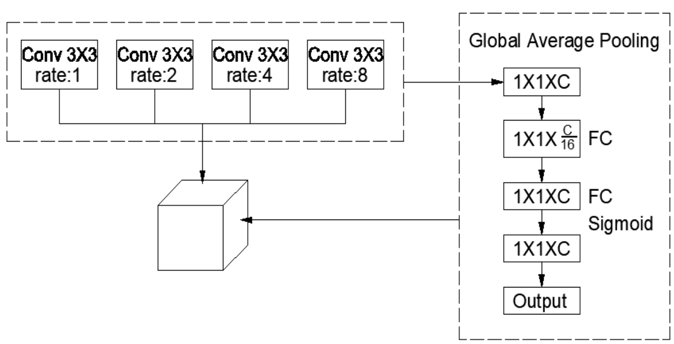


Figure 5. The structure of SE_ASPP module.

Finally, the SPPF module in YOLOv8 is replaced with the SE_ASPP module. SE_ASPP combines Atrous Spatial Pyramid Pooling (ASPP) and the channel attention mechanism SENET. Generally, the receptive field is closely linked to object classification. A larger receptive field typically improves network performance, but an excessively large receptive field can make the model difficult to converge. If the model needs a large receptive field while maintaining the resolution of the feature map (to preserve image details), the dilated convolution becomes essential. Thus, the ASPP module is beneficial as it can effectively balance the receptive field and resolution. It utilizes multiple parallel dilated convolution layers with different dilation rates to sample the input features, allowing the model to construct different receptive fields from branches of varying scales, extract the input features, and use them to generate the final feature results.

Additionally, the use of channel attention mechanism SENET not only facilitates the effective transfer of key feature information, enhancing information reuse and amplifying useful information but also minimizes redundant feature information. Figure 5 illustrates the specific structure of SE_ASPP.

RESULT ANALYSIS

The hardware configuration for our experiments includes an Nvidia GeForce GTX 3090 graphics card and an Intel i7 processor. The experimental software encapsulates Python 3.8.16 and Torch 1.13.1. The detailed parameters of the experiment are presented in Table 1. In this book chapter, the enhanced YOLOv8 model significantly improved the accuracy of waste classification. Additionally, ablation experiments were conducted to verify the feasibility and effectiveness of the improved model.

Table 1. The detailed parameters of the experiment

Classes	Parameters
Initial learning rate	0.01
Optimizer	SGD
Momentum	0.9
Weight decay	0.0005
Batch size	16
Epoch	300

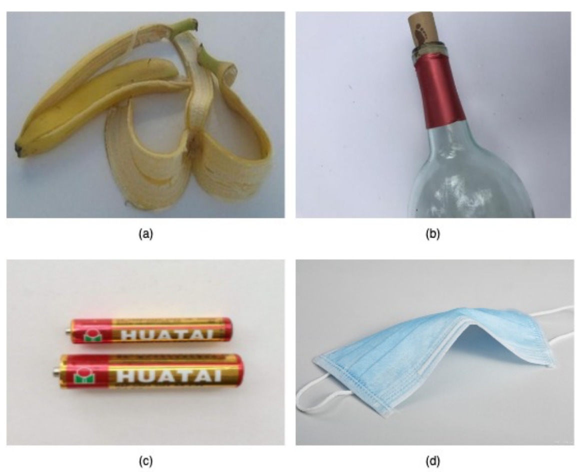


Figure 6. The samples of waste dataset, (a) A banana is assigned to the class "Wet", (b) A glass bottle is grouped to the class "Recyclable", (c) A battery is classified to the class "Hazardous", (d) A mask is set to the class "Dry".

The dataset in this book chapter is a waste dataset. The waste images were selected from a large pool and divided into four classes, totaling 1,000 images. According to the waste classes, the images were manually labelled as "Recyclable," "Hazardous," "Wet," and "Dry." Additionally, each class contains different types of wastes. For example, the recyclable category includes glass, cardboard, and plastic. To enrich the dataset, multiple perspectives of the same object are also annotated. Table 2 presents the specific details of the dataset, Figure 6 displays the images from the dataset.

Table 2. The number of samples in our waste dataset

Classes of Samples	**Numbers**
Recyclable waste	253
Recyclable waste	251
Wet waste	252
Dry waste	244
Total	1,000

In this experiment, Average Precision (AP), Mean Average Precision (mAP), and F1 score are the metrics which are taken to evaluate the accuracy and performance of the model. Figure 7 depicts the confusion matrix of the model.

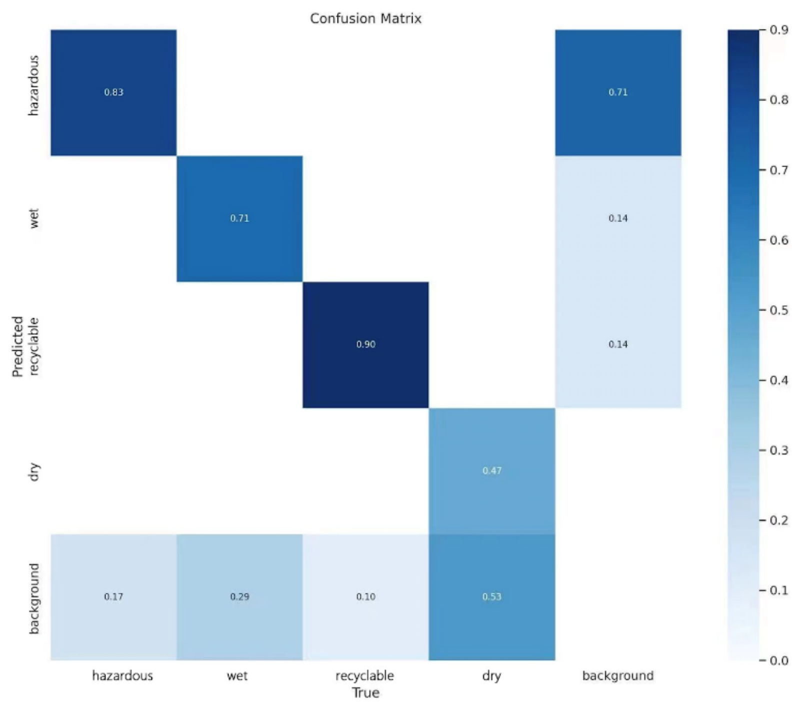
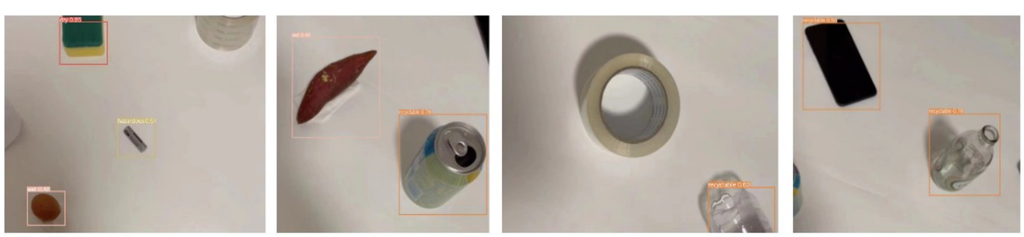


Figure 7. The confusion matrix of waste classification

To make the classification results visible, a classified waste is marked. The waste classification results are shown in Figure 8. In these results, multiple classes of classified wastes were represented in colour with bounding boxes, including "Wet," "Dry," "Recyclable," and "Hazardous." Most of the classification results are correct; however, the object in Figure 8(c) is a missed classification and should belong to the class "dry".



(a) (b) (c) (d)

Figure 8. Waste classification results. (a) The visual objects like egg shell, battery, and sponge dishcloth, were classified to "Wet", "Hazardous", and "Dry" respectively. (b) The visual objects sweet potato and can, were classified to "Wet" and "Recyclable" respectively. (c) The visual objects like plastic bottles were classified as "Recyclable". (d) The visual objects like phone and glass bottle were classified to the classes like "Recyclable" and "Recyclable" respectively.

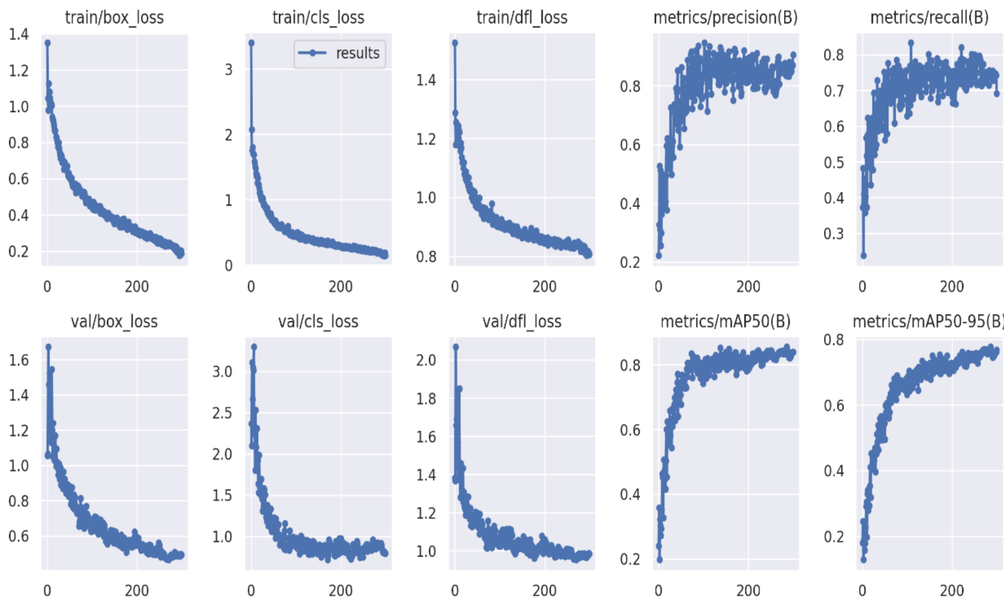


Figure 9. The mean average precision and loss of the waste classification

Furthermore, the parameters such as mAP and Recall rates for waste classification are detailed in Table 3. After 300 training epochs, the total mAP of the model is 0.856, with the mAP values of classification for Hazardous, Recyclable, Wet, and Dry classes being 0.927, 0.955, 0.820, and 0.720, respectively. Figure 9 illustrates the stability of the improved model.

Table 3. The parameters of each waste class

Classes	Box	R	mAP 50	mAP 50:90
Recyclable waste	0.874	0.848	0.927	0.884
Recyclable waste	0.912	0.900	0.955	0.905
Wet waste	0.866	0.761	0.820	0.674
Dry waste	0.717	0.467	0.720	0.650
Total	0.842	0.744	0.856	0.778

Table 4. The comparison of waste classification

Models	mAPs	F1 score
Faster R-CNN	0.639	0.611
SSD	0.665	0.616
YOLOv5	0.717	0.676
YOLOv7	0.759	0.720
YOLOv8	0.802	0.768
Ours	0.856	0.790

To further explore the performance of the model, the improved YOLOv8 model is compared with five advanced models by using our dataset, as shown in Table 4. Our model achieved the highest mAP 0.856, which is an improvement by 0.054 over the original YOLOv8 model. In comparison, the mAP values of YOLOv5 and YOLOv7 are 0.717 and 0.759, respectively. Moreover, the mAP values of SSD and Faster R-CNN are relatively low, not exceeding 0.700, with Faster R-CNN having the lowest mAP at 0.639, which is 0.026 smaller than the mAP value of SSD.

Moreover, ablation experiments were conducted on the proposed model to verify the validity and necessity of the proposed improved features (akin to control variables). The results of the ablation studies confirm that the improved YOLOv8 model is feasible and valid. The results of the eight ablation experiments are detailed in Table 5. The original YOLOv8 model without any modifications show the lowest mAP 0.802. Consequently, data augmentation, feature fusion, and SE_ASPP modules are separately added. Among these, the SE_ASPP module contributed to the largest performance improvement, with an increase in mAP of 0.022. Feature fusion increased the mAP of YOLOv8 by 0.015, while data augmentation, though it provided the smallest improvement, increased the mAP by 0.003. This demonstrates that data augmentation, feature fusion, and SE_ASPP are all crucial for enhancing the performance of the model.

Table 5. The mAP of the model in the ablation experiment

Model	Data Augmentation	Feature Fusion	SE_ASPP	mAP
YOLOv8				0.802
	√			0.814
		√		0.817
			√	0.824
	√		√	0.845
	√	√		0.834
		√	√	0.851
	√	√	√	0.856

The mAP value for the combination of data augmentation and feature fusion is 0.834. If feature fusion is ignored, the mean AP value is 0.845. The mAP value increases to 0.851 if feature fusion and SE_ASPP are applied. These results are lower than the mAP 0.856 if all three features are utilized, indicating that optimal results are obtained by applying all three features together.

CONCLUSION

In conclusion, YOLOv8 is enhanced by incorporating data augmentation, contextual information methods, and the SE_ASPP module to improve the accuracy of waste classification. 1,000 waste images are labelled across four categories: “hazardous”, “wet”, “recyclable”, and “dry”. The mAP of this model improved by 5.4% to 85.6%. Additionally, the ablation study verifies the effectiveness of the model, demonstrating that the model is efficient and stable for the waste classification. However, the experiments are limited, primarily by the lack of a diverse dataset. Thus, enhancing the waste dataset remains an important task for future work.

REFERENCES

Abdu, H., Noor, M. (2022) A survey on waste detection and classification using deep learning. IEEE Access, 10, 128151-128165.

Al-Sarayreh, M., Reis, M., Yan, W., Klette, R. (2019) A sequential CNN approach for foreign object detection in hyperspectral images. International Conference on Information, Communications and Signal.

Bochkovskiy, A., Wang, C., Liao, M. (2024) YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934 2020.

Chen, G., Wang, H., Chen, K., Li, Z., Song, Z., Liu, Y., Chen, W., Knoll, A. (2022) A survey of the four pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal. IEEE Trans. Syst. Man Cybern. Syst, 52, 936-953.

Chen, S., Huang, J., Xiao, T., Gao, J., Bai, J., Luo, W., Dong, B. (2020) Carbon emissions under different domestic waste treatment modes induced by garbage classification: Case study in pilot communities in Shanghai, China. Science of the Total Environment, 717, 137193.

Cubuk, E., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V. (2019) AutoAugment: Learning augmentation policies from data. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, California, pp. 113-123.

De Carolis, B., Ladogana, F., Macchiarulo, N. (2020) YOLO TrashNet: Garbage detection in video streams. In IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), Ljubljana, pp. 1-7.

Fu, J., Sun, X., Wang, Z., Fu, K. (2022) An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images. IEEE Trans. Geosci. Remote Sens, 59, 1331-1344.

Fu, K., Li, J., Ma, L., Mu, K., Tian, H. (2009) Intrinsic relationship reasoning for small object detection. arXiv:2009.00833 2020.

Fu, Y., Nguyen, M., Yan, W. (2022) Grading methods for fruit freshness based on deep learning. Springer Nature Computer Science.

Gowdra, N., Sinha, R., MacDonell, S., Yan, W. (2021) Maximum Categorical Cross Entropy (MCCE): A noise-robust alternative loss function to mitigate racial bias in Convolutional Neural Networks (CNNs) by reducing overfitting. Pattern Recognition.

Guo, C., Fan, B., Zhang, Q., Xiang, S., Pan, C. (2020) AugFPN: Improving multi-scale feature learning for object detection. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12595-12604.

Heo, B., Yun, S., Han, D., Chun, S., Choe, J., Oh, S. J. (2021) Rethinking spatial dimensions of vision transformers. In IEEE/CVF International Conference on Computer Vision, pp. 11936-11945.

Hou, Q., Zhou, D., Feng, J. (2021) Coordinate attention for efficient mobile network design. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13713-13722.

Inoue, H. (2018) Data augmentation by pairing samples for images classification. arXiv preprint arXiv:1801.02929.2018

Kang, Z., Yang, J., Li, G., Zhang, Z. (2020) An automatic garbage classification system based on deep learning. IEEE Access, 8, 140019-140029.

Karras, T., Laine, S., Aila, T. (2019) A style-based generator architecture for generative adversarial networks. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, California, pp. 120322-140332.

Liu, J., Pan, C., Yan, W. (2022) Litter detection from digital images using deep learning. Springer Nature Computer Science.

Liu, X., Yan, W. Q., Kasabov, N. (2023) Moving vehicle tracking and scene understanding: A hybrid approach. Multimedia Tools and Applications, 83, 51541-51558.

Liu, Z., Yan, W., Yang, B. (2018) Image denoising based on a CNN model. International Conference on Control, Automation and Robotics.

Lun, Z., Pan, Y., Wang, S., Abbas, Z., Islam, M. S., Yin, S. (2023) Skip-YOLO: Domestic garbage detection using deep learning method in complex multi-scenes. International Journal of Computational Intelligence Systems, 16, 139.

Nayan, A., Saha, J., Mozumder, A. (2020) Real time detection of small objects. Int. J. Innov. Technol. Explor. Eng, 9, 837.

Nguyen, M., Yan, W. Q. (2023) From faces to traffic lights: A multi-scale approach for emotional state representation. In IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys), pp. 942-949.

Pan, C., Yan, W. (2018) A learning-based positive feedback in salient object detection. International Conference on Image and Vision Computing New Zealand.

Pan, C., Yan, W. (2020) Object detection based on saturation of visual perception. Multimedia Tools and Applications, 79 (27-28), 19925-19944.

Pan, C., Liu, J., Yan, W., Zhou, Y. (2021) Salient object detection based on visual perceptual saturation and two-stream hybrid networks. IEEE Transactions on Image Processing.

Qi, J., Nguyen, M., Yan, W. (2022) Small visual object detection in smart waste classification using Transformers with deep learning. International Conference on Image and Vision Computing New Zealand (IVCNZ).

Qi, J., Nguyen, M., Yan, W. (2022) Waste classification from digital images using ConvNeXt. Pacific-Rim Symposium on Image and Video Technology (PSIVT).

Qi, J., Nguyen, M., Yan, W. Q. (2024) CISO: Co-iteration semi-supervised learning for visual object detection. Multimedia Tools and Applications, 83, 33941-33957.

Qi, J., Nguyen, M., Yan, W. (2024) NUNI-Waste: novel semi-supervised semantic segmentation waste classification with non-uniform data augmentation. Multimedia Tools and Applications, 25, 1-19.

Rabano, S. L., Cabatuan, M. K., Sybingco, E., Dadios, E. P., Calilung, E. J. (2018) Common garbage classification using MobileNet. In IEEE International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), pp. 1-4.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016) You Only Look Once: Unified, real-time object detection. In IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788.

Wang, L., Yan, W. (2021) Tree leaves detection based on deep learning. International Symposium on Geometry and Vision.

Woo, S., Park, J., Lee, J. Y., Kweon, I. S. (2018) CBAM: Convolutional block attention module. In European Conference on Computer Vision (ECCV), pp. 3-19.

Xia, Y., Nguyen, M., Yan, W. (2022) A real-time Kiwifruit detection based on improved YOLOv7. International Conference on Image and Vision Computing New Zealand (IVCNZ)

Xia, Y.; Nguyen, M.; Yan, W. Q. (2023) Kiwifruit counting using KiwiDetector and KiwiTracker. In Intelligent Systems Conference, pp. 629-640.

Xia, Y., Nguyen, M., Yan, W. (2023) Multiscale Kiwifruit detection from digital images. PSIVT.

Xiang, Y., Yan, W. (2021) Fast-moving coin recognition using deep learning. Springer Multimedia Tools and Applications.

Xiao, B., Nguyen, M., Yan, W. (2021) Apple ripeness identification using deep learning. International Symposium on Geometry and Vision.

Xiao, B., Nguyen, M., Yan, W. (2023) Apple ripeness identification from digital images using transformers. Multimedia Tools and Applications, Springer Science and Business Media LLC.

Xiao, B., Nguyen, M., Yan, W. (2023) Fruit ripeness identification using transformers. Applied Intelligence, Springer Science and Business Media LLC.

Xiao, B., Nguyen, M., Yan, W. (2023) A mixture model for fruit ripeness identification in deep learning. Handbook of Research on AI and ML for Intelligent Machines and Systems, pp.1-16, Chapter 16, IGI Global.

Xiao, B., Nguyen, M., Yan, W. (2023) Fruit ripeness identification using YOLOv8 model. Multimedia Tools and Applications.

Xue, Y. Yan, W. (2023) YOLO models for fresh fruit classification from digital videos. Handbook of Research on AI and ML for Intelligent Machines and Systems, pp. 421-435, Chapter 17, IGI Global.

Yan, W. (2023) Computational Methods for Deep Learning: Theoretic, Practice and Applications. Springer Nature.

Yan, W. (2019) Introduction to Intelligent Surveillance: Surveillance Data Capture, Transmission, and Analytics. Springer Nature.

Yang, X., Zhao, W., Wang, Y., Yan, W., Li, Y. (2024) Lightweight and efficient deep learning models for fruit detection in orchards. Scientific Reports 14, 26086

Zhang, H.Y., Cisse, M., Dauphin, Y.N., Lopez-Paz, D. (2018) Mixup: Beyond empirical risk minimization. In International Conference on Learning Representations, pp. 1-13.

Zhao, K., Yan, W. (2021) Fruit detection from digital images using CenterNet. International Symposium on Geometry and Vision.

Zhao, K., Nguyen, M., Yan. (2024) Evaluating accuracy and efficiency of fruit image generation using generative AI diffusion models for agricultural robotics. IEEE IVCNZ'24

Zhu, X., Lyu, S., Wang, X., Zhao, Q. (2021) TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In IEEE/CVF International Conference on Computer Vision, pp. 2778-2788.

Zhu, Z., Su, J., Lu, L., Li, B., Wang, X., Dai, J.F. (2020) Deformable DETR: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159.