

Efficient Differentiation of Biodegradable and Non-Biodegradable Municipal Waste Using a Novel MobileYOLO Algorithm



Menaka Suman¹, Gayathri Arulanantham^{2*}

Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai 602105, India

Corresponding Author Email: gayathribala.sse@saveetha.com

<https://doi.org/10.18280/ts.400505>

ABSTRACT

Received: 1 March 2023
Revised: 28 June 2023
Accepted: 8 September 2023
Available online: 30 October 2023

Keywords:

MobileYOLO, MobileNetv2, real-time dataset, waste object detection

In the realm of waste management, the accurate identification of biodegradable and non-biodegradable items remains a critical challenge. An advanced real-time object detection method, termed “MobileYOLO”, was proposed, leveraging the strengths of the YOLO v4 framework. The MobileNetv2 network was integrated, and a section of the conventional computation was substituted with depth-wise separable convolutions utilizing the PAnet and head network. To enhance feature expressiveness capabilities during feature fusion, a refined lightweight channel attention mechanism, known as Efficient Channel Attention (ECA), was introduced. The Improved Single Stage Headless (ISSH) context module was incorporated into the micro-object identification branch to broaden the receptive field. Evaluations conducted on the KITTI dataset indicated an impressive accuracy of 95.7%. Remarkably, when compared to the standard YOLOv4, the MobileYOLO model exhibited a reduction in model parameters by 53.12M, a decrease in connectivity size by one-fifth, and an augmentation in detection speed by 85%.

1. INTRODUCTION

The rapid urbanization and subsequent rise in the urban populace have thrust the challenge of the “urban trash siege” to the forefront of environmental concerns in cities. This growth of urban sectors has been linked to a staggering increase in the generation of municipal waste, intensifying the urgency of the issue known as the “urban trash siege”. Research by International Lianhe Zaobao suggests that by 2050, there will be a 70% growth in global garbage volumes [1-3]. In such a context, strategies to achieve the reduction, recycling, and harmless disposal of urban waste have emerged as imperative challenges for urban governance.

Industrialized nations have identified waste reuse as a pivotal measure in this combat [4]. The success of waste reuse is contingent on the efficient classification of trash. It has been observed that hazardous materials such as waste fluorescent bulbs, discarded thermometers, obsolete pharmaceuticals, packaging, and waste image sheets are typically consigned to landfills. Concurrently, valuable recyclable commodities, encompassing materials like paper, plastic bags, and beverage bottles, are found amidst this refuse [5]. When valuable waste is discerned and segregated from non-recyclables, it not only ensures environmental protection but also catalyzes socio-economic progress. Illustratively, one tonne of waste paper has been converted into 850 kilograms of quality paper, saving 300 kilograms of wood in the process; similarly, a tonne of discarded plastic bottles has been transformed into 600 kilograms of unleaded gasoline or diesel [6].

Waste classification undoubtedly benefits societal structures, economies, and the environment. It has been recognized for conserving natural resources and bolstering the recycling movement, while concurrently curtailing pollution

[7]. Despite its significance, the predominant method of waste identification has been reliant on human efforts. Manual sorting, though widely practiced, is not only marked by inefficiencies but also poses potential health threats to the workforce [8, 9]. Given these challenges, a shift towards leveraging artificial intelligence for waste categorization has been advocated. The future of waste management is envisaged to be dominated by intelligent categorization and recycling.

Over recent years, advancements in Deep Learning (DL) have permeated a diverse spectrum of sectors [10]. Technologies underpinned by DL have been extensively deployed to address multifarious real-world challenges, enhancing system precision and efficacy in the process [11]. Such methodologies find applications across an array of sectors spanning healthcare, security, and automation, to name a few.

The automatic detection of littered areas has been explored to bolster cleaning efficiency and preclude secondary pollution. Hyperspectral data from aircrafts has been utilized to track waste dispersal across expansive terrains [12]. An unsupervised area proposal generation technique integrating selective search and non-maximum suppression has been applied to pinpoint the extent and position of littered zones [13]. Drawing from remote sensing imagery, models to autonomously discern areas riddled with waste have been developed. Preliminary experiments have demonstrated that the aforementioned model can adeptly identify littered regions with swiftness and precision, underscoring the efficiency and feasibility of deploying DL tools for such purposes.

For the real-time identification of waste, a 23-layer CNN model was employed, coupled with the generation of area proposals and the adaptation of the Visual Geometry Group-16 model for discerning plastic bottles amid challenging

backgrounds [14, 15]. SpotGarbage, an innovative smartphone application, was introduced to detect and loosely classify litter locations within geotagged images captured by users. Deep convolutional neural networks (DCNNs) have been harnessed for the detection and pinpointing of pavement flaws and waste. The employed waste classification technique is anchored on the ResNet-34 algorithm [16]. Enhancements to this algorithm included multi-feature fusion of input images, feature reuse within the residual unit, and the introduction of a novel activation function [17]. The refined trash classification technique demonstrated capabilities in efficiently and accurately sorting waste.

From an environmental perspective, the repercussions of inadequate waste management are profound, manifesting as pollution, soil and water source contamination, and heightened greenhouse gas emissions. On an economic front, suboptimal waste management practices can impose considerable financial strains on municipalities. Absence of meticulous identification and segregation during sorting and disposal can inflate associated costs, encompassing transportation, landfill utilization, and waste processing. Moreover, public health is invariably compromised with substandard waste management, evidenced by disease proliferation, contamination of food and water resources, and the surge of pests.

Utilizing the KITTI dataset, an accuracy of 95.7% was achieved by the proposed system during trials. When juxtaposed with YOLOv4, the MobileYOLO model manifested significant enhancements. A notable reduction in model parameters by 53.12M was observed, amplifying computational efficiency and tempering storage and processing demands. Furthermore, a marked diminution, approximately one-fifth, in connectivity size was witnessed, fostering efficient information dissemination within the system. This model also showcased an 85% surge in speed detection prowess, catalyzing swifter, and more precise object detection and tracking. Such strides in accuracy, parameter reduction, connectivity size, and speed detection underscore the potential of the proposed system, earmarking its applicability across varied object detection realms.

Subsequent sections delve deeper into this research. Section 2 elucidates works pertinent to this domain. An innovative algorithm for the streamlined identification of biodegradable and non-biodegradable entities within municipal waste is delineated in Section 3, with results elucidated in Section 4. The manuscript culminates in Section 5, where future avenues of research are contemplated.

2. RELATED WORK

Approaches to trash classification through DL have been discerned to fall into four predominant categories: Recurrent Neural Networks (RNNs), DCNNs, the amalgamation of CNN models with transfer learning, and one-stage object identification algorithms [18]. The architecture of one-stage detection networks is noted for its simplicity and is attributed with exemplary detection accuracy [19], suggesting vast potential for research and application within real-time object identification.

Among the eminent one-stage object detection methods, YOLO stands out [20]. Within the YOLO framework, feature extraction is carried out using CNNs, obviating the need for a Region Proposal Network (RPN). Detection speed is enhanced as these features are directly introduced to a regression

network, which subsequently provides both the object bounding box and class probability [21]. Both RNN and DCNN algorithms employ an end-to-end network approach, analyzing entire images through regression. Such a strategy is identified to bolster real-time performance and subsequently enhance object detection accuracy.

YOLO, characterized as a proficient CNN model, excels in real-time object detection. The entire image is processed through a singular computational model, subsequently subdividing it into regions and determining boundary boxes and probabilities for each [22]. Among object detection algorithms, YOLO is recognized for its prowess, especially when handling extensive datasets, marking it as exceptionally fast and accurate [23]. It has been observed that YOLO efficiently discerns objects within dynamic frames or videos in real-time. Darknet, a framework specifically tailored for YOLO deployment, was introduced by the researchers [24]. YOLO approaches detection by segmenting frames into grids, striving to pinpoint objects within each. Utilizing a Non-max suppression filter, only the most prominent bounding boxes within a grid are retained, ensuring optimal outcomes [25]. Subsequent to the original YOLO's inception, several iterations were developed to augment its efficacy, notably YOLO v2 and YOLO v7.

In contrast, two-stage object detection methods such as Faster R-CNN and Mask R-CNN initiate by ascertaining an object's location through an area proposal network, followed by classification and prediction of the selected region. Despite the commendable accuracy attributed to these techniques, they necessitate substantial computational resources, rendering real-time requisites elusive [26]. Conversely, one-stage object detection algorithms, typified by SSD and YOLO, facilitate simultaneous classification and regression predictions on an object. Such methodologies have been demonstrated to diminish computational intricacies and expedite inference. SSD, for instance, employs varied-sized convolution layers in conjunction with a multi-scale feature map to discern objects. The SSD model, when augmented with a deconvolution layer as seen in DSSD, enhances the detection capability for smaller objects without compromising detection speed.

3. PROPOSED SYSTEM

3.1 Lightweight model for pavement garbage classification

A model, predicated on YOLOv5, which demonstrates proficiency in rapid object detection, is introduced in this article. To tailor this model for waste categorization, structural modifications are undertaken. The ghost module is incorporated into the backbone network to curtail the model's dimensions [27]. In an endeavor to further streamline the model, regular convolution layers are supplanted by depth-wise separable convolutional layers, leading to a reduction in the number of variables within the PANet network. To maintain detection accuracy while concurrently achieving a diminutive model size, the attention mechanism termed SELayer is assimilated into the backbone network. As illustrated in Figure 1, image features are initially extracted from the backbone network and subsequently channeled into the PANet network. In the concluding stages, the input feature maps undergo processes of up-sampling and down-sampling, ultimately yielding predictions for both category and bounding box. Notably, within the structure delineated in Figure 1, the

ghost module is substituted for the BottleneckCSP, with ‘N’ signifying the number of times the module has been stacked.

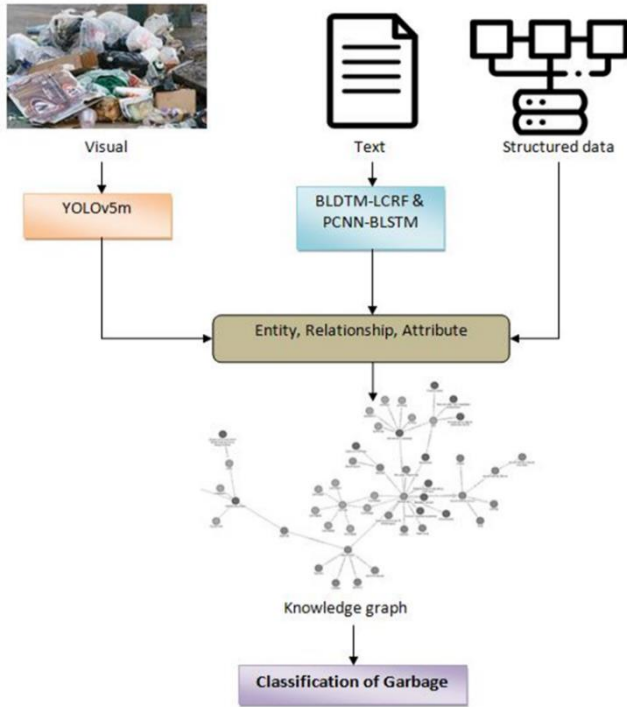


Figure 1. Architecture overview

3.2 Embedding of the ghost module

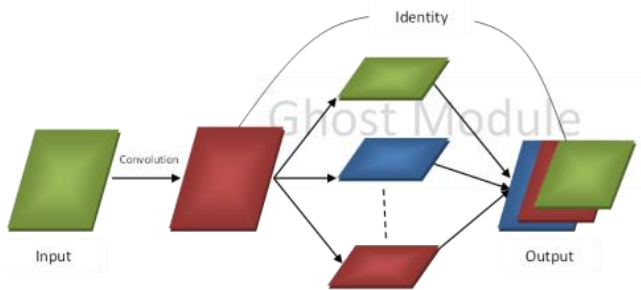


Figure 2. Schematic of the ghost module

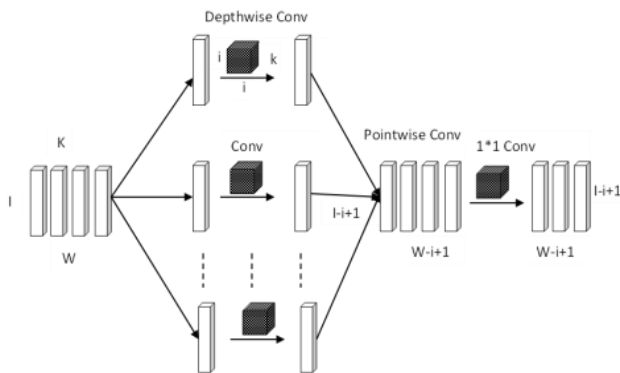


Figure 3. Depiction of the depth-wise separable convolutional model

The GhostNet, a pioneering lightweight neural network proposed by Huawei Noah’s Ark Laboratory, is predicated on the integration of the ghost module [28]. When integrated into

a convolutional neural network, the ghost module has been observed to diminish computational costs and model parameters. Such reductions arise because the ghost module is capable of constructing additional feature maps via highly economical operations. For addressing memory and performance concerns associated with the garbage classification model, the BottleneckCSP component of YOLOv5 was replaced with the ghost module having step sizes of 1 and 2, respectively. As illustrated in Figure 2, feature maps within the ghost module are manipulated to generate $H' \times W'$ n feature maps, post which a linear mapping of the n channel feature maps is conducted. Two ghost modules are stacked to form the Ghost Bottleneck. The initial ghost module aids in amplifying the feature map channels, while the subsequent ghost module reduces the channel count to align with the shortest pathway.

In object detection methodologies, the Ghost Module stands out as a significant technique that curtails computational costs and memory demands. This module introduces what are termed “ghost channels”, partitioning the input channels into two distinct subsets: primary and ghost. While the primary set, having a reduced number of channels, is directly linked to the subsequent layers, the ghost set independently assimilates supplementary data. Through a linear combination, these ghost channels are integrated with the primary channels, bolstering parameter efficiency. Despite its reduced complexity, this streamlined module retains its performance, rendering it apt for real-time object detection, especially in devices with stringent resource constraints. Within convolutional neural networks, the efficacy of the Ghost Module is exemplified as it provides an optimized approach to object detection tasks. In the research outlined, the depth-wise separable CNN takes precedence over conventional convolution layers, further reducing the parameters within the PANet network. Consequently, the described model’s lightweight nature is accentuated. Figure 3 illustrates the mechanics of depth-wise separable convolution, which can be discerned in two phases: depth-wise convolution and pointwise convolution.

$$P_{\text{conv}} = (I - i + 1) \times (W - i + 1) \times N_c \times i \times i \times K \quad (1)$$

$$P_{KW \text{ conv}} = ((I - i + 1) \times (W - i + 1) \times N_c \times i \times i \times K + N_c \times 1 \times 1 \times K \times (I - i + 1) \times (W - i + 1)) \quad (2)$$

Computational costs associated with the Depth-wise Separable Convolutional Model are exemplified in Eq. (1) and Eq. (2):

Eq. (1):

- P_{conv} denotes the computational burden of the conventional convolutional procedure.
- I and W symbolize the spatial metrics of the input feature map (width and height, respectively).
- N_c is representative of the input channel count.
- i signifies the kernel size (dimensions of the convolutional kernel).
- K stands for the output channel tally.

Eq. (2):

- $P_{KW \text{ conv}}$ embodies the computational expenditure of the Depth-wise Separable Convolutional operation.
- The inaugural term signifies the computational load of the

depth-wise convolution.

- The subsequent term conveys the computational heft of the point-wise convolution.

Subsequently, the ratio $P_{KW_{conv}}/P_{conv}$ is computed, providing insight into the comparative computational loads.

$$\frac{P_{KW_{conv}}}{P_{conv}} = \frac{1}{N_c} + \frac{1}{i^2} \quad (3)$$

When executing depth-wise separable convolution, an elevated count of output channels is observed. This translates to the assertion that in Eq. (3), the NC count markedly

overshadows h . Consequently, the training duration for standard 2D convolution is h^2 times that of the depth-wise separable convolution.

3.3 Network architecture and comparative models

Conventional object detection networks, exemplified by YOLOv4, typically operate in a single step. Such frameworks enable the concurrent prediction of item categories and bounding boxes, culminating in marked enhancements in detection timeframes. The architectural representation of YOLOv4 is depicted in Figure 4.

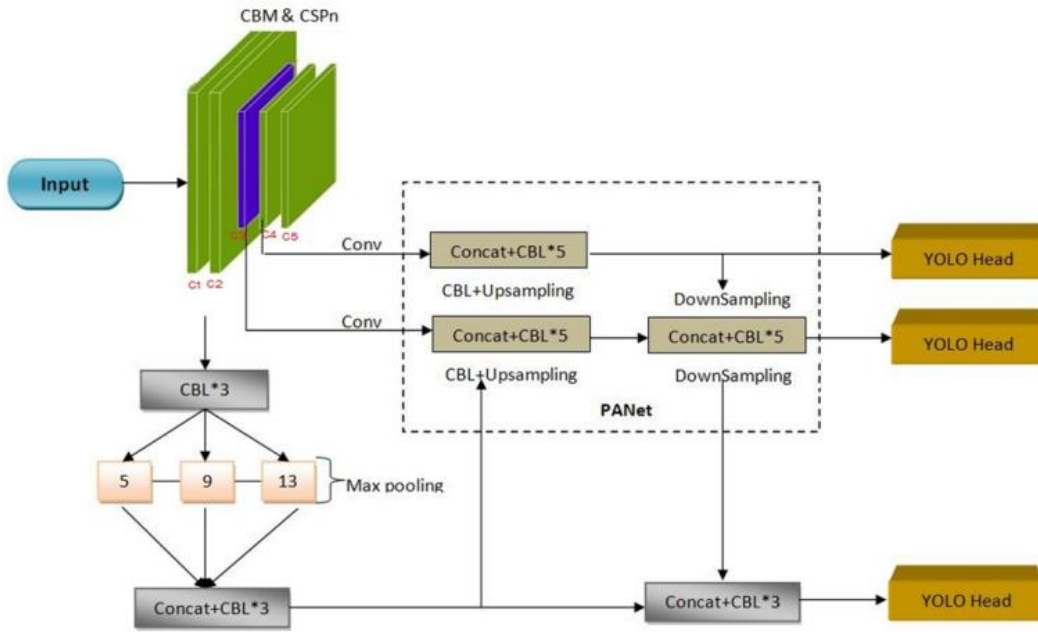


Figure 4. Structure of the YOLOv4 model

The architecture of YOLOv4 is strategically compartmentalized into three pivotal sections: the backbone networks, neck networks, and head networks. Serving as the foundation or backbone network is the CSPDarknet53 feature extraction mechanism, composed of a quintet of modules ranging from CSP1 to CSP5. Each of these modules is constructed by alternating the stacking of CSPx blocks with CBM in lieu of CBL blocks [29]. The character “x” in CSPx denotes the number of CSP blocks. Initially, the CSP divides the input into a bifurcated branch structure. One of these branches undergoes a CBM operation, subsequently traverses the residual structure, and again partakes in a CBM operation. Post this sequence, a concatenation process is employed, resulting in the output of both branches. Within this context, CBM represents the sequence of convolution, Batch Normalization, and Mish activation functions. Conversely, CBL signifies convolution, BN, and Leaky ReLU activation functions. The CSPnet topology is integrated into the YOLOv4’s backbone network. A significant residual side, enhancing learning capacity, emerges from the counteraction of CBM with the Res unit stack [30]. Through the employment of the Mish function, which possesses neither upper nor lower bounds and is characterized by non-monotonicity and infinite-order continuity, the model is facilitated in executing regularisation and upholding gradient stability in the network. The subsequent equation delineates the loss function:

$$Loss = Loss(coord) + Loss(cls) + Loss(conf) \quad (4)$$

$$Loss(coord) = \lambda_{coord} \sum_{x=0}^{H \times H} \sum_{y=0}^h X_{xy}^{obj} (2 - w_x \times i_x) \quad (5)$$

$$Loss(cls) = - \sum_{x=0}^{Zl \times H} X_{xy}^{obj} \sum_{c \in classes} \{ \hat{p}_x(c) \lg p_x(c) + (1 - \hat{p}_x(c)) \lg(1 - p_x(c)) \} \quad (6)$$

$$Loss(conf) = - \sum_{x=0}^{H \times H} X_{xy}^{noobj} \sum_{y=0}^M X_{xy}^{obj} \{ \hat{C}_x \lg C_x + (1 - \hat{C}_x) \lg(1 - C_x) \} - \lambda_{noobj} \sum_{x=0}^{H \times H} \sum_{y=0}^C X_{xy}^{naobj} \{ \hat{C}_x \lg C_x + (1 - \hat{C}_x) \lg(1 - C_x) \} \quad (7)$$

where,

- H stands for grid size;
- X designates the x th feature map;
- y represents the predicted bounding box;
- w indicates width, while l defines height;
- obj and $no\ obj$ signify the constants of object presence and absence, respectively.
- The predicted and actual category classifications are respectively symbolized by C_x and \hat{C}_x .

- The confidence pertaining to the predicted object is represented by $p_x(c)$, and $\hat{p}_x(c)$ corresponds to the confidence of the actual object.
- Penalty coefficients `_coord` and `_noobj` are similarly defined.

The SPP structure is recognized for its significant augmentation of the receptive field, thereby inheriting enhanced contextual attributes. Responsibility for transporting the P3, P4, and P5 feature layers is attributed to the PANet. Subsequent to the facilitation of a bottom-to-top data fusion enhancement, these distinct feature layers are subjected to a down-sampling process, executed in the inverse sequence of the former step. This methodology harnesses insights from low-level feature positioning while concurrently abbreviating the path of information transmission. The loss function of YOLOv4, denoted as “Loss”, amalgamates the regression loss function, probability loss function, and classification loss functions.

3.4 Architecture of the MobileYOLO model

In the present study, the MobileYOLO algorithm has been introduced, derived from the foundational structure of YOLOv4. Figure 5 provides a graphical representation of the overall architecture, highlighting the lightweight optimizations incorporated within MobileYOLO. A reduction in the model parameters of YOLOv4 was achieved by replacing its conventional communication system with the more lightweight Mobilenetv2. Furthermore, the adoption of Depthwise Separable Convolution in lieu of traditional pooling layers, both within the feature extraction network and the detection head network, contributed to this parameter minimization. To enhance the neural network, a concise attention module, termed ECANet, was integrated. This module is postulated to augment the prominence of pertinent features, attenuate redundant features, and amplify the network’s feature fusion characterization capabilities.

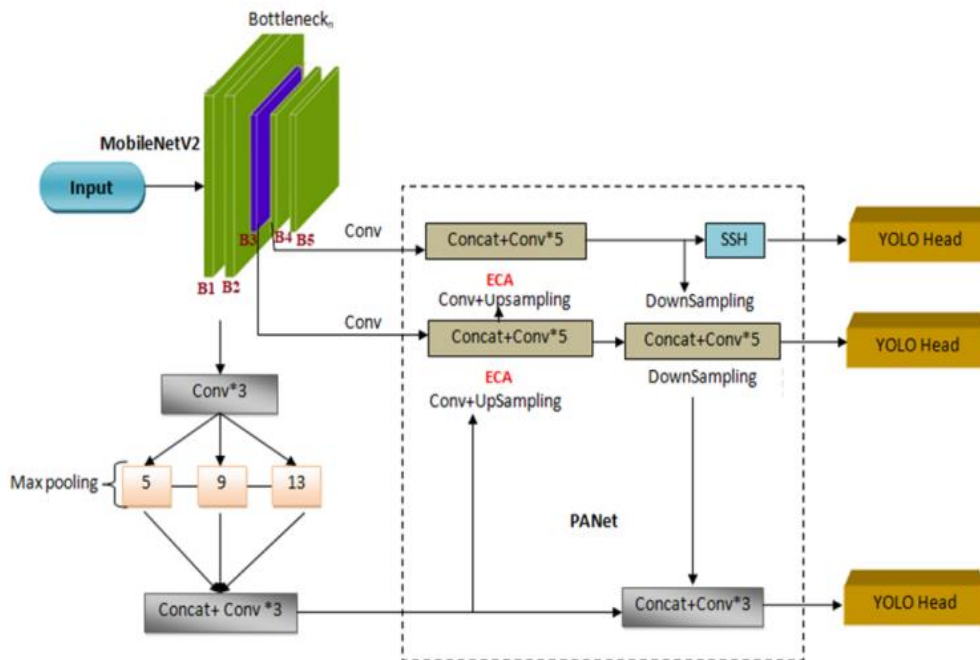


Figure 5. Architecture of the MobileYOLO model

3.5 Structure of the MobileNetV2 network

For the objective of extracting object features specific to waste management scenarios, MobileNetV2 has been utilized as an advanced backbone network. The adoption of this network is noted to significantly reduce the quantity of model parameters. Distinct from its predecessor, MobileNetV1, MobileNetV2 incorporates an inverted residual module furnished with linear bottleneck blocks. This inclusion is reported to markedly elevate the efficiency of image classification and detection tasks, especially when executed on mobile platforms.

The propensity for the loss of characterization trait information becomes pronounced when operations employ the ReLU activation function on layers proximate to the base convolutional layer. However, it is observed that the deployment of the Linear transfer function on high-dimensional convolution operations minimizes data loss. This phenomenon can be attributed to the richer data content inherent to the high-dimensional fully connected layer,

compared to its lower-dimensional counterparts [31]. In an attempt to circumvent excessive feature information loss, the ReLU6 activation function was integrated into the high-dimensional convolution layer of MobileNetV1.

The topology of bottleneck sections within the MobileNetV2 network is illustrated in Figure 6. The primary objective of the inverted residual module is discerned to amplify the network’s feature extraction potential and facilitate efficient multi-layer information relay.

When assessing a multitude of algorithms on ImageNet-compiled data, the TOP-1 predictive accuracy serves as a measure to gauge the congruence between the primary rated classification and the factual outcome, as detailed in Table 1. Subsequent to the integration of the ECA module, a marginal increment in both the parameters and computational magnitude is detected. Nevertheless, ECA retains the potential to deliver augmented detection accuracy, all the while imposing only a trifling computational demand, thereby preserving its lightweight attributes and ensuring proficient performance.

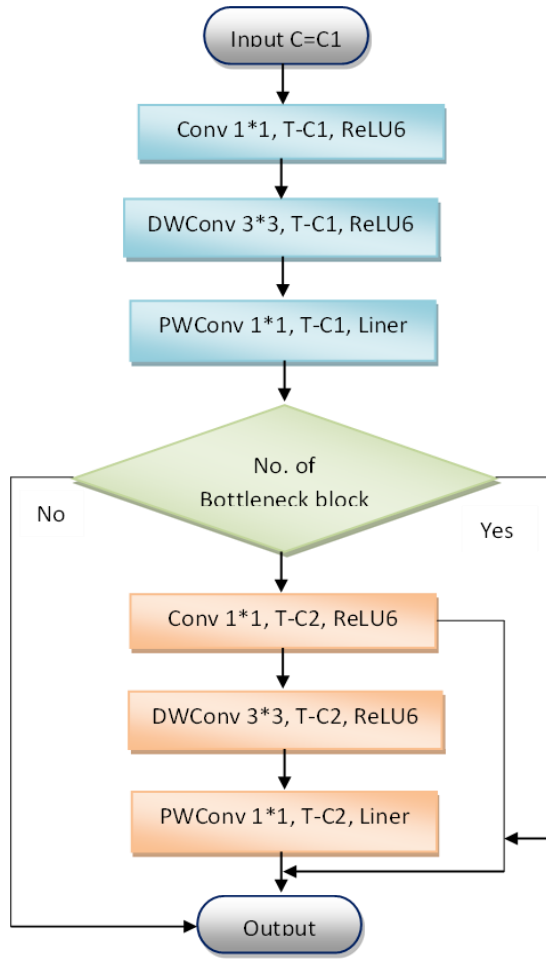


Figure 6. Structure of the bottleneck block in MobileNetV2

Table 1. Performance indicators of various attention systems

Models	Param Value	FIOP Value	s/G Accuracy	TOP-1
ResNet-50	42.51	7.36		76.91
+SE	47.06	7.37		77.58
+CBAM	47.06	7.37		78.44
+AA	45.43	8.11		78.76
+ECA	42.51	7.37		78.67

In the quest to circumvent information loss during the dimensionality reduction process, feature information acquisition was found to obviate the need for manual cross-validation adjustments. Subsequently, the Sigmoid activation function was employed, yielding dimension c with a novel parameter. The reallocation of weights across different channel characteristics ensued, wherein non-essential features were observed to be more robustly suppressed, amplifying the significance of pertinent information. ECA, in comparisons drawn with attention modules, was discerned to mirror the efficiency and lightweight properties akin to SENet.

$$H = \varphi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{L}{\gamma} \right\rfloor_{ODD} \quad (8)$$

For multiple channels symbolized as c , the notation $\lfloor * \rfloor$ represents the odd number proximate to the value, as depicted in Figure 7.

Within the delineated context module, convolutions of size 3×3 were leveraged, aiming to emulate the functionality of 5×5 and 7×7 convolution kernels. Such an approach was discerned

to act as a surrogate for larger convolutions. Evident from Figure 8, the kernel was subsequently aimed at diminishing the aggregate parameter count. With the ISSH context module's integration into each triad of PAnet outputs, an escalation in the overall parameter number was noted. This study saw a reduction in the SSH module output to 76×76 , enhancing the feature capturing capacity for diminutive objects.

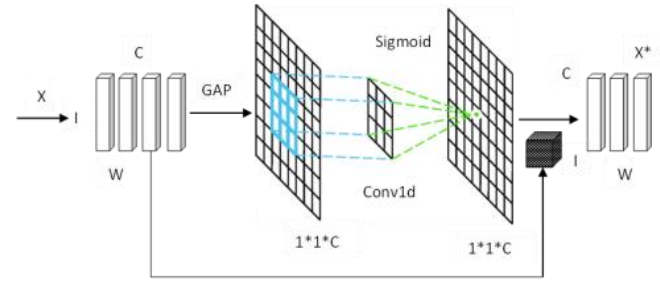


Figure 7. ECA structure

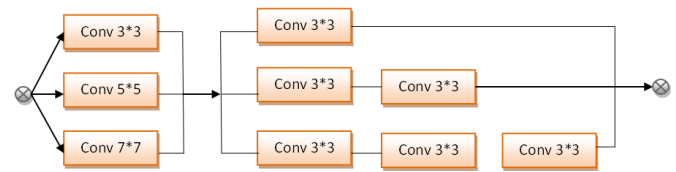


Figure 8. ISSH structure

From a managerial perspective, the deployment of this algorithm was correlated with the optimization of waste management protocols. Such implementations allowed municipalities to channel resources judiciously, predicated on the precision-driven classification of waste constituents. This precision not only translated to economic efficiency but also fostered targeted waste treatment methodologies, abating environmental ramifications and fostering sustainable modalities. Additionally, superior waste segregation facilitated by the algorithm was linked to public health enhancements, curtailing disease propagation and contamination threats. These implications are instrumental in refining waste management infrastructures, proffering multifaceted benefits to both urban centers and the environment.

4. RESULT AND DISCUSSION

Open-source datasets concerning waste remain limited in the current landscape. To construct such datasets, self-image acquisition and online searches are often employed by researchers. Notably, in 2020, a dataset focusing on household refuse was provided by Huawei during a competition named the "Trash Categorization Challenge Cup". Comprising 14,964 images, this dataset was segmented into four primary categories and 44 subcategories. For the purpose of this study, attention was primarily centered on conventional street waste. From the provided Huawei dataset, images exemplifying typical pavement garbage were exclusively extracted. Furthermore, an additional 294 images capturing fallen leaves, a common pavement occurrence, were integrated. The culmination of this endeavor resulted in a curated dataset of standard pavement refuse, encompassing 2,442 images. As detailed in Table 1, this newly established dataset was subsequently divided into seven distinct categories: beverage

bottles, pericarps, dry cells, disposable fast-food containers, plastic bags, cigarette butts, and fallen leaves.

Utilizing the PyTorch DL framework, experiments were conducted on the Ubuntu 16.04 system. The computational setup for the experiment consisted of an Intel core processor with a 32GB memory and an NVIDIA 1080Ti graphics card. The transfer learning methodology was applied for model training. A bifurcated training strategy was adopted. Initially, pretrained weights from the VOC07+12 datasets were integrated. During this preliminary phase, the backbone network was maintained in a static state to mitigate excessive variability, ensuring that initial training did not inadvertently compromise the integrity of the original weights. Parameters were set with a batch size of 32 and an input resolution of 608×608. The training spanned 60 epochs in total. The loss convergence of the MobileYOLO application on the KITTI dataset is illustrated in Figure 9. Epoch-wise loss values were documented, highlighting a consistent decrease in validation loss curves. Following 65 iterations, stabilization of loss was observed at 1.55, indicating the model’s optimal state.

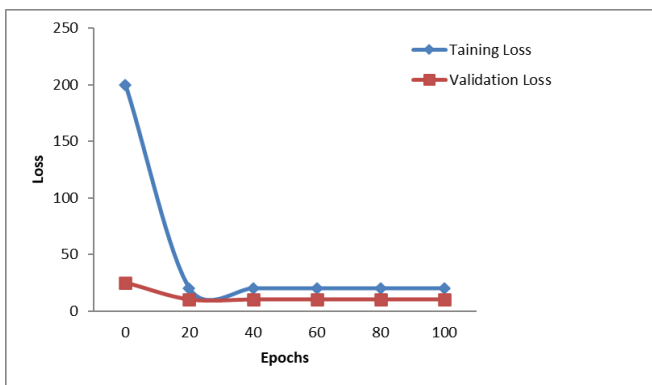


Figure 9. Loss convergence of MobileYOLO

4.1 Evaluation metrics and methodology

The collective interpretation of these metrics provides a comprehensive average for each specific AP encompassed within the dataset. Such an approach aids in the assessment of the efficacy of the proposed technique. AP, predominantly employed as an intuitive criterion, gauges the detection model’s precision and is evaluated based on two dimensions: precision and recall. The F1 score, on the other hand, serves as an indicator that quantifies the model’s accuracy by computing the harmonic mean of precision and recall. Algorithms

employed to determine prediction accuracy, recall value, F1 score, and AP are detailed below:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$AP = \int_0^{-1} P(R) dR$$

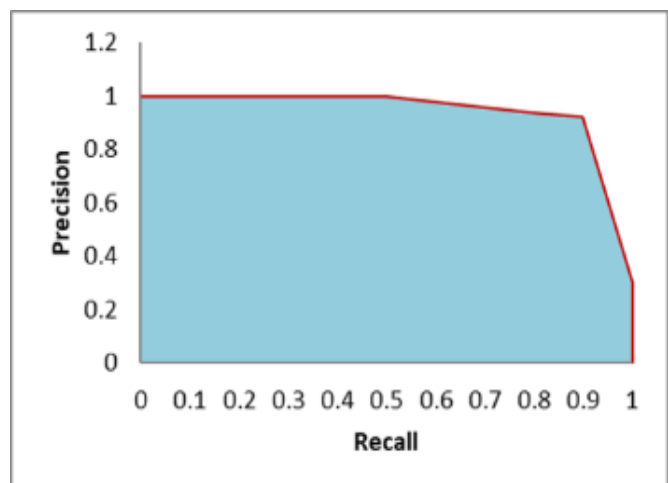
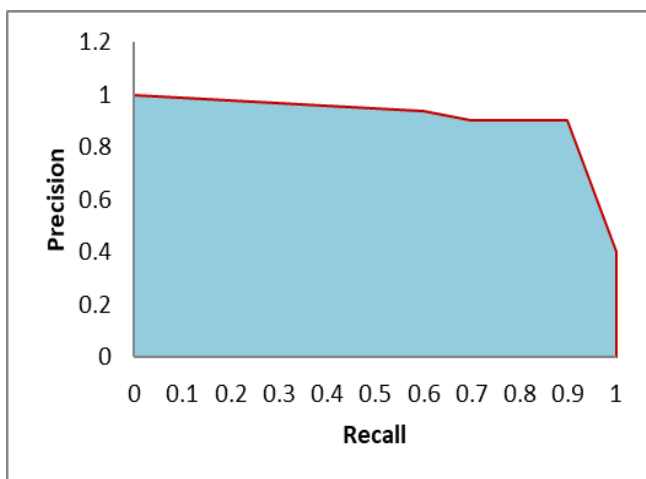
For the evaluation graph, recall is suggested to be positioned on the abscissa, while accuracy is recommended to be situated on the ordinate. A direct correlation has been observed between the enhancement in detection accuracy and the magnitude of the scrutinized region.

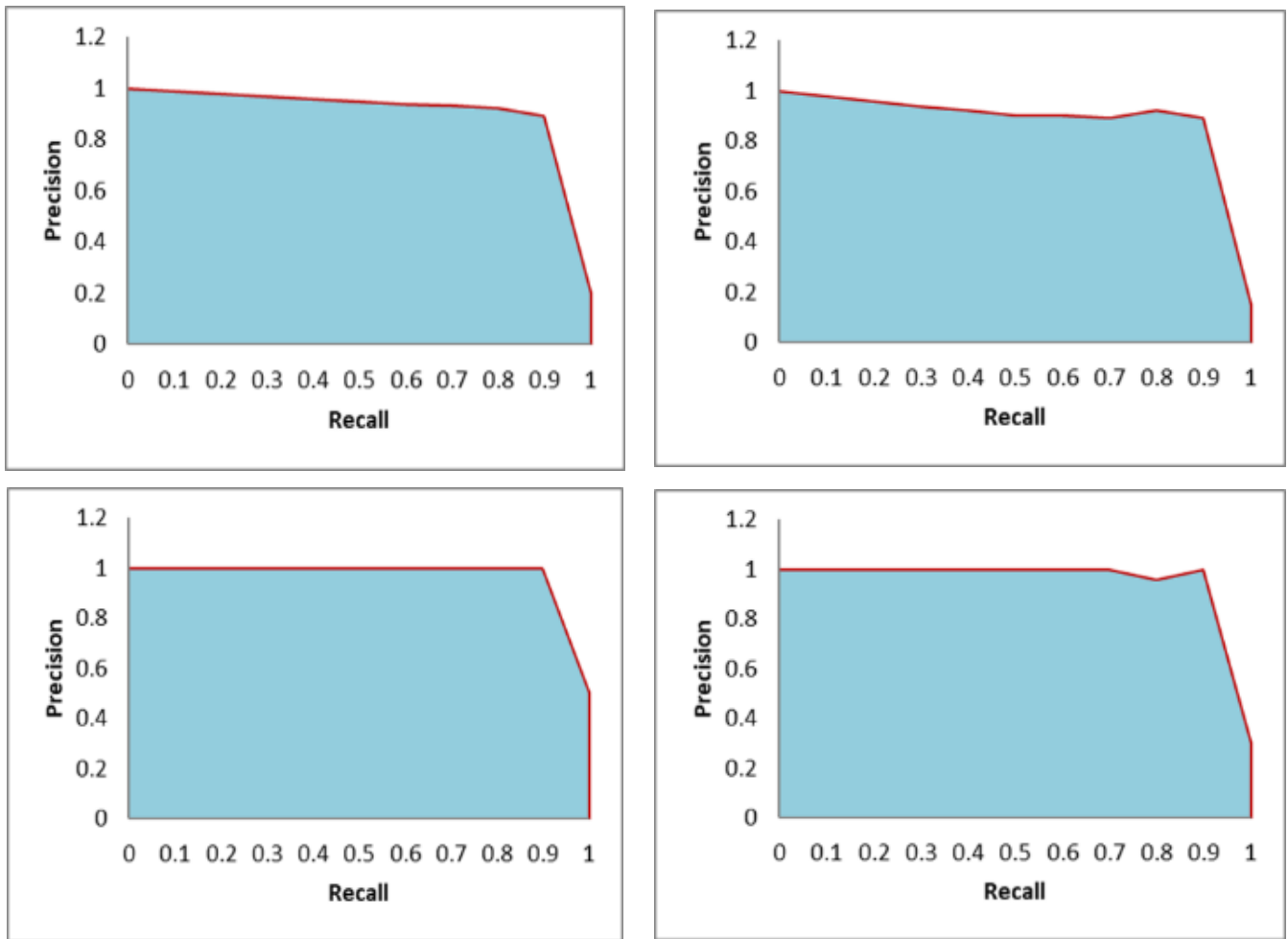
4.2 Experimental results

The model was applied to evaluate and obtain experimental findings post training the MobileYOLO network. As depicted in Figure 10, the average Precision-Recall (P-R) curve of the detection models was obtained using YOLOv4 and MobileYOLO, employing the KITTI data set. The P-R curve serves as a robust comparative tool to assess the detection effectiveness of diverse techniques. The final results of the prediction, inclusive of the detection category and the associated confidence level, are illustrated at the header of the graphic adjacent to each detection box, as shown in Figure 11. Visual inspection of the results indicates that the proposed MobileYOLO exhibits a detection efficacy equivalent to that of the YOLOv4 algorithm. Additionally, the incidence of false positives and missed detections is low, indicating a high level of applicability.

4.3 Ablation experiment

In the field of DL, ablation studies are frequently employed. The primary objective of such an experiment is to assess the influence of various network nodes on the overall model integrity. The specific results of these modules, referenced earlier, are provided in Table 2, where “X” denotes the improvement strategy associated with each event.





a) PR curve of YOLOv4

b) PR curve of MobileYOLO

Figure 10. P-R graphs depicting the outcomes of the detection process



Figure 11. Visualization of results

Table 2. Ablation study results

Model	MobileNetV2	Convolution of DW	ECA	ISSH	mAP	Params	Size	FPS
YOLOv4 (0)					0.934	64.38	245.7	47.7
YOLOv4 (1)	Yes				0.916	39.2	98.5	68.9
YOLOv4 (2)	Yes	Yes			0.888	12.12	46.3	82.8
YOLOv4 (3)			Yes		0.944	64.37	245.8	47.3
YOLOv4 (4)				Yes	0.942	64.46	246.2	47.2
YOLOv4 (5)	Yes	Yes	Yes		0.898	12.13	46.4	82.0
MobileYOLO (6)	Yes	Yes	Yes	Yes	0.901	12.26	46.8	80.4

Table 3. Comparative results

Method	Version	Size of Input	mAP	Params	Size	FPS
SSD0300	VGG-16	300×300	0.842	24.28	92.3	48.3
YOLOv3	Darknet-53	608×608	0.876	61.54	235.8	42.2
RetinaNet	ResNet-50	500×500	0.888	38.35	167.5	36.6
YOLOv4-tiny	CSPdarknet-53	608×608	0.829	6.30	22.6	95.9
YOLOv4	CSPdarknet-53	608×608	0.936	65.39	246.5	47.8
Mobile YOLO	MobileNetV2	608×608	0.909	12.27	46.9	80.5

Findings from Experiments 1 and 2 indicate that the incorporation of MobileNetV2 and deep separable volume enables a proactive reduction in model parameters. This reduction subsequently leads to an enhancement in detection speed, albeit at the expense of accuracy. The Efficient Channel Attention (ECA) module and the Single Stage Headless (SSH) module were subjected to empirical comparison and testing in Experiments 3 through 6. It was observed that a moderate increase in the number of parameters could potentially contribute to a 2.1% increase in the mean Average Precision (mAP). Noteworthy enhancements were made to the YOLOv4 in this study. Despite the 2.6% decrease in mAP when compared with YOLOv4, MobileYOLO significantly reduced the initial count of model parameters to 52.11 million. Concurrently, the overall model size was reduced to approximately one-fifth of its original size, and the frame rate witnessed a surge of 70 percent.

4.4 Comparison of various algorithms

To evaluate the proposed method's effectiveness and scientific validity, it was subjected to a comparison with the prevailing one-stage object detection techniques. The results of these comparisons are encapsulated in Table 3.

The performance of MobileYOLO on mAP is observed to be 2.6% inferior to that of YOLOv4, yet it exhibits a significant increase of 32.7 in Frames Per Second (FPS) and a substantial reduction in both model parameters and weight. MobileYOLO, being more affordable, amalgamates the functionalities of YOLOv4 and YOLOv4-tiny. Despite its trade-offs in detection speed and accuracy, it provides enhanced security and is thus more suited for applications such as waste management.

5. CONCLUSIONS

This study presents a real-time object detection method predicated on the YOLOv4 algorithm. Empirical results demonstrate that MobileYOLO achieves an accuracy rate of 90.7% on the KITTI dataset. In contrast to YOLOv4, all parameters of the proposed MobileYOLO system have been reduced by 53.12 million, leading to a considerable decrease in computational power demands and associated SSH training expenses. This reduction may be beneficial for the deployment of embedded devices. However, the MobileYOLO methodology proposed herein is a supervised learning method, necessitating a substantial quantity of training samples. Future studies should focus on reducing the number of experimental samples and exploring semi-supervised approaches.

REFERENCES

[1] Aishwarya, A., Wadhwa, P., Owais, O., Vashisht, V.

- (2021). A waste management technique to detect and separate non-biodegradable waste using machine learning and YOLO algorithm. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, pp. 443-447. <https://doi.org/10.1109/Confluence51648.2021.9377163>
- [2] Nafees, A., Khan, S., Javed, M.F., Alrowais, R., Mohamed, A.M., Mohamed, A., Vatin, N.I. (2022). Forecasting the mechanical properties of plastic concrete employing experimental data using machine learning algorithms: DT, MLPNN, SVM, and RF. *Polymers*, 14(8): 1583. <https://doi.org/10.3390/polym14081583>
- [3] Abdulrahman, H., Hewahi, N. (2021). Waste classifications using convolutional neural network. In 2021 International Conference on Data Analytics for Business and Industry (ICDABI), IEEE, pp. 260-265. <https://doi.org/10.1109/ICDABI53623.2021.9655829>
- [4] Ravishankar, A., Murthy, A., Sharma, M., Chitra, R.K., Anitha, R. (2021). Automated waste segregation using convolution neural network. In 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), IEEE, pp. 1-4. <https://doi.org/10.1109/SMARTGENCON51891.2021.9645825>
- [5] Behera, S.K., Barathwaj, Y., Vasundhara, L., Saisudha, G., Haariharan, N.C. (2021). AI based waste classifier with thermo-rapid composting. *arXiv e-prints*, arXiv-2108.01394. <https://doi.org/10.48550/arXiv.2108.01394>
- [6] Menaka, S., Harshika, J., Philip, S., John, R., Bharathiraja, N., Murugesan, S. (2023). Analysing the accuracy of detecting phishing websites using ensemble methods in machine learning. In 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), IEEE, pp. 1251-1256. <https://doi.org/10.1109/ICAIS56108.2023.10073834>
- [7] Sivakumar, M., Renuka, P., Chitra, P., Karthikeyan, S. (2022). IoT incorporated deep learning model combined with smartbin technology for real-time solid waste management. *Computational Intelligence*, 38(2): 323-344. <https://doi.org/10.1111/coin.12495>
- [8] Belsare, K.S., Singh, M. (2022). Various frameworks for IoT-enabled intelligent waste management system using ML for smart cities. In *Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI, Singapore: Springer Nature Singapore, 2022: 797-817*. https://doi.org/10.1007/978-981-19-2069-1_55
- [9] Dessai, H., Miranda, D. (2021). Garbage segregation using images. In 2021 2nd International Conference for Emerging Technology (INCET), IEEE, pp. 1-4. <https://doi.org/10.1109/INCET51464.2021.9456408>
- [10] Saranya, A., Bhambri, M., Ganesan, V. (2021). A cost-effective smart e-bin system for garbage management using convolutional neural network. In 2021 International Conference on System, Computation,

- Automation and Networking (ICSCAN), IEEE, pp. 1-6. <https://doi.org/10.1109/ICSCAN53069.2021.9526547>
- [11] Jimeno, F.N., Briz, B.J.A., Artiaga, M.R.P., Angelia, R.E., Limsangan, N.B. (2021). Development of smart waste bin segregation using image processing. In 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), pp. 1-6. <https://doi.org/10.1109/HNICEM54116.2021.9732038>
- [12] Wang, J., Ayari, M.A., Khandakar, A., Chowdhury, M.E.H., Uz Zaman, S.A., Rahman, T., Vaferi, B. (2022). Estimating the relative crystallinity of biodegradable polylactic acid and polyglycolide polymer composites by machine learning methodologies. *Polymers*, 14(3): 527. <https://doi.org/10.3390/polym14030527>
- [13] Deepa, S., Umamageswari, A., Menaka, S. (2023). A novel hand gesture recognition for aphonic people using convolutional neural network. In *Computer Vision and Machine Intelligence Paradigms for SDGs: Select Proceedings of ICRTAC-CVMIP*, Singapore: Springer Nature Singapore, 2021: 235-243. https://doi.org/10.1007/978-981-19-7169-3_22
- [14] Akanksha, Gupta, M., Tripathi, M.M. (2021). Smart robot for collection and segregation of garbage. In 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM), IEEE, pp. 169-173. <https://doi.org/10.1109/ICIPTM52218.2021.9388369>
- [15] Karthika, K.P., Joshiba, J., Indhumalar, M., Saranya, G. (2021). Machine learning powered smart dumpster monitoring and clearance system. In 2021 6th International Conference on Communication and Electronics Systems (ICES), IEEE, pp. 1598-1602. <https://doi.org/10.1109/ICES51350.2021.9489123>
- [16] Sruthy, V., Akshaya, Anjana, S., Ponnaganti, S.S., Pillai, V.G., Preetha, P.K. (2021). Waste collection & segregation using computer vision and convolutional neural network for vessels. In 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), IEEE, pp. 1043-1048. <https://doi.org/10.1109/ICCCIS51004.2021.9397092>
- [17] Divakar, S., Bhattacharjee, A., Priyadarshini, R. (2022). An IoT-based smart garbage segregation system using deep learning. In *Internet of Things and Its Applications: Select Proceedings of ICIA*, Singapore: Springer Nature Singapore, 2020: 121-132. https://doi.org/10.1007/978-981-16-7637-6_12
- [18] Vinod, D., Bharathiraja, N., Anand, M., Antonidoss, A. (2021). An improved security assurance model for collaborating small material business processes. *Materials Today: Proceedings*, 46: 4077-4081. <https://doi.org/10.1016/j.matpr.2021.02.611>
- [19] Varudandi, S., Mehta, R., Mahetalia, J., Parmar, H., Samdani, K. (2021). A smart waste management and segregation system that uses internet of things, machine learning and android application. In 2021 6th International Conference for Convergence in Technology (I2CT), IEEE, pp. 1-6. <https://doi.org/10.1109/I2CT51068.2021.9418125>
- [20] Akanksha, S., Bindushree, K., Rachana, R., Sharadhi, M.K., Saravana, M.K. (2022). Efficient garbage management system using machine learning. *Perspectives in Communication, Embedded-systems and Signal-processing-PiCES*, 5(10): 97-101. <https://doi.org/10.5281/zenodo.6026635>
- [21] Thiruneelakandan, A., Kaur, G., Vadnala, G., Bharathiraja, N., Pradeepa, K., Retnadhas, M. (2022). Measurement of oxygen content in water with purity through soft sensor model. *Measurement: Sensors*, 24: 100589. <https://doi.org/10.1016/j.measen.2022.100589>
- [22] Nguyen, X.C., Nguyen, T.T.H., La, D.D., Kumar, G., Rene, E.R., Nguyen, D.D., Chang, S.W., Chung, W.J., Nguyen, X.H., Nguyen, V.K. (2021). Development of machine learning-based models to forecast solid waste generation in residential areas: a case study from Vietnam. *Resources, Conservation and Recycling*, 167: 105381. <https://doi.org/10.1016/j.resconrec.2020.105381>
- [23] Yarlagadda, J., Ramakrishna, M. (2019). Fabrication and characterization of S glass hybrid composites for Tie rods of aircraft. *Materials Today: Proceedings*, 19: 2622-2626. <https://doi.org/10.1016/j.matpr.2019.10.104>
- [24] Mehrdad, S.M., Abbasi, M., Yeganeh, B., Kamalan, H. (2021). Prediction of methane emission from landfills using machine learning models. *Environmental Progress & Sustainable Energy*, 40(4): e13629. <https://doi.org/10.1002/ep.13629>
- [25] Chhabra, M., Sharan, B., Gupta, K., Astya, R. (2022). Waste classification using improved cnn architecture. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4157549>
- [26] Soundarya, B., Parkavi, K., Sharmila, A., Kokiladevi, R., Dharani, M., Krishnaraj, R. (2022). CNN based smart bin for waste management. In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), IEEE, pp. 1405-1409. <https://doi.org/10.1109/ICSSIT53264.2022.9716437>
- [27] Moradi, M.H., Sohani, A., Zabihigivi, M., Wagner, U., Koch, T., Sayyaadi, H. (2022). Machine learning and artificial intelligence application in land pollution research. *Current Trends and Advances in Computer-Aided Intelligent Environmental Data Engineering*, 273-296. <https://doi.org/10.1016/B978-0-323-85597-6.00008-2>
- [28] Azhaguramyaa, V.R., Janet, J., Narayanan, V.V.L., Sabari, R.S., Santhosh, K.K. (2021). An intelligent system for waste materials segregation using IoT and deep learning. In *Journal of Physics: Conference Series*, IOP Publishing, 1916(1): 012028. <https://doi.org/10.1088/1742-6596/1916/1/012028>
- [29] Anitha, R., Maruthi, R., Sudha, S. (2022). Automated segregation and microbial degradation of plastic wastes: a greener solution to waste management problems. *Global Transitions Proceedings*, 3(1): 100-103. <https://doi.org/10.1016/j.gltp.2022.04.021>
- [30] Joshi, L.M., Bharti, R.K., Singh, R. (2022). Internet of things and machine learning-based approaches in the urban solid waste management: trends, challenges, and future directions. *Expert Systems*, 39(5): e12865. <https://doi.org/10.1111/exsy.12865>
- [31] Mookkaiah, S.S., Thangavelu, G., Hebbar, R., Haldar, N., Singh, H. (2022). Design and development of smart internet of things-based solid waste management system using computer vision. *Environmental Science and Pollution Research*, 29(43): 64871-64885. <https://doi.org/10.1007/s11356-022-20428-2>