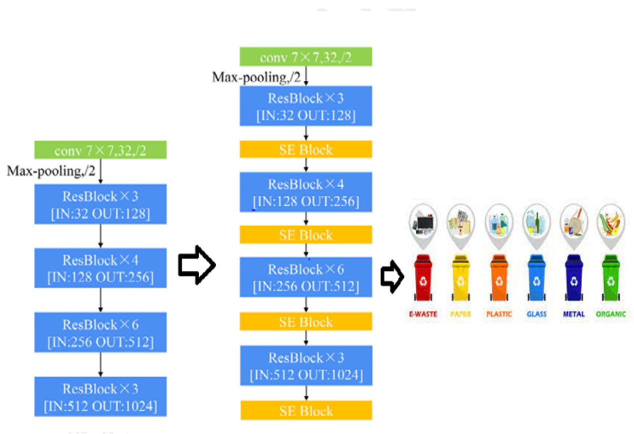
Global NEST Journal, Vol 26, No 3, 05304

A Deep Learning-Based Buffalo Optimizer based Squeeze and Excitation Network for Garbage Classification for a Sustainable Environment

Subhashini R[1,*], R.J. Kavitha[2], M. Sumithra[3]. and T. Jayasankar[4]
[1]Department of Information Technology, Sona college of Technology, Salem
[2]Department of Electronics and Communication Engineering, University College of Engineering, Panruti–607106
[3]Department of Information Technology, Panimalar Engineering College, Chennai
[4]Department of Electronics and Communication Engineering, University College of Engineering, Anna University (BIT Campus), Trichy, Tamil Nadu, India
Received: 18/08/2023, Accepted: 07/01/2024, Available online: 24/01/2024
*to whom all correspondence should be addressed: e-mail: subhashini.it@sonatech.ac.in
https://doi.org/10.30955/gnj.005304

Graphical abstract



Abstract

A Squeeze and Excitation Network is a deep-learning architectural component designed to enhance networks. The “squeeze” step reduces the spatial dimensions of the input feature maps, and the “excitation” step adaptively recalibrates channel-wise feature responses. This allows the network to focus on more educational features and ignore less useful ones. Garbage classification is a crucial task for sustainable environmental management. It involves categorizing waste into recyclables, organics, and non-recyclables, among other classes. Deep learning models, like the proposed Buffalo Optimizer-based SEN, can play a pivotal role in automating this classification process using computer vision techniques. Garbage Classification: There is still a long way to go until countries worldwide have successfully increased public awareness and implemented measures to prevent the rapid degradation of the natural environment. The annual global generation of e-waste is between 20 and 50 million growing components of municipal solid garbage. The disposal of electronic trash presents significant threats to environmental quality. As a result, pollution monitoring and control are crucial for maintaining a healthy ecosystem. In this research, picture-layered characteristics were extracted using an encoder composed of varying numbers of ResBlocks and the Squeeze-and-Excitation (SE) block, which was built on top of the UNet backbone network. UNet’s decoder structure was streamlined, and the number of network model parameters was cut in half. In the meantime, the multiscale feature fusion module was developed to enhance the network’s detection accuracy by optimizing its parameters with a bespoke loss function in place of the standard function. The African Buffalo Optimisation Algorithm is also used to fine-tune the hyper-parameters.

Keywords: Smart garbage management**,** deep learning, Image classification, garbage segregation buffalo optimizer-based squeeze and excitation network.

1. Introduction

The concept of ‘sustainable development’ was crystallized in 1987 in the generation to meet their own needs when possible; management means choosing one or more strategies that are meant to meet short-term goals while still leaving enough room for long-term choices that was hard to define the idea of environmental management because the word “environment” is so broad. People who work in different areas have given their different opinions on how to manage the environment Zhang X *et al.* (2023). Riordan has said that the two main ways to handle the environment are conservative and preservative. He says that the preservative method calls for people to not mess with the physico-biotic world and to fully adapt to it Wazid M *et al.* (2019).

Development that meets the needs of future generations to meet their own needs” was what the commission said sustainable development meant. While meeting immediate needs should come first, the main focus should be on meeting basic human wants and ending poverty. This was done to raise the standard of living for the poorest people in society and to avoid future costs that won't be covered. The idea has made people think about how to support economic and social growth in a way that doesn't hurt the environment or use up too many resources. This has put off the debate about whether to put development ahead of protecting the environment Gu M *et al.* (2019).

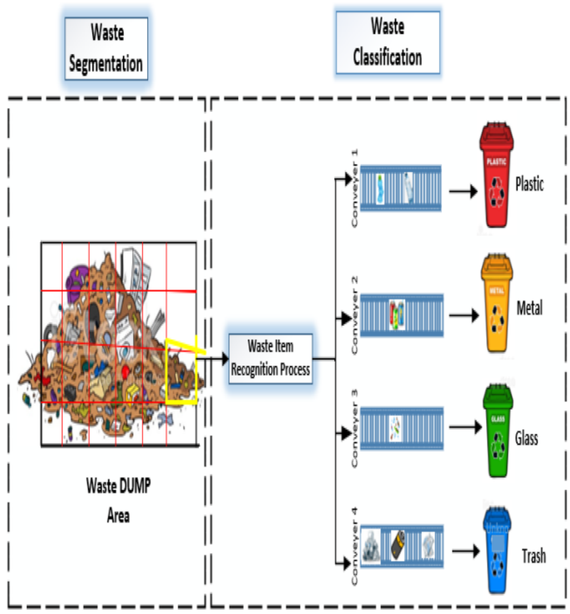


Figure 1. Waste management and classification

In 2004, Williams said that sustainable development could help solve problems with either growth or the environment. He said that the gap between the supply and demand of resources could be closed to some extent by lowering society's need for earth resources and increasing the supply of those resources Hon A *et al.* (2019). He has come up with different ways to achieve sustainability based on the concept. The first way is to make the earth's riches bigger and better. A weak durability is the name for it. This can be done by finding new sources of renewable resources, alternatives to non-renewable resources, better ways to use the resources we already have, and technology answers to issues like pollution and resource loss. Stopping people from using too many resources is the second way. It is possible to do this if people use fewer resources and live in harmony with nature. It's called "strong sustainability." "Moderate sustainability" is the name of a third way that takes parts of both the weak and strong methods to sustainability. Increasing the supply of resources while lowering demand is the main point of this plan. In this way, the resources are used and available at their best without sacrificing the benefits for future generations Daulay HY *et al.* (2020).

The three main parts of sustainable expansion are economic growth, social growth, and environmental safety. All of these parts work together to make the whole thing stronger. The goals of sustainable development are to end poverty, find the best ways to produce and consume goods and protect and control natural resources. To reach the goals of sustainable development, we also need good governance, sound economic policies, strong democratic institutions, a method that is based on people's needs, and respect for human rights Zhang X *et al.* (2023). The ecological balance has been upset by the effects of mass tourism on the natural surroundings and ecosystems. According to the famous economist and futurologist Herman Khan, the fast growth of tourism is "next only to atomic power in its potential Anuradha M *et al.* (2021)." This has led to damage to the environment. So, the idea of "sustainable development" needs to be applied right away to tourism, and the idea of "eco-tourism" is based on the idea of fairness between generations Golroudbari AA *et al.* (2023).

According to Healy, there is a lot of proof that the tourism business around the world is becoming more aware of how it affects the environment. Recently, a lot of work has been done at different levels to encourage tourism growth that is sustainable. All of these efforts around the world are meant to encourage sustainable tourism growth so that people in host communities can get more out of tourism resources and keep their cultural and environmental integrity by making sure that natural heritages and ecologically sensitive areas are protected. Boo wrote in 1990 that eco-tourism had been a big part of the economies of places like Belize, Costa Rica, Galapagos Island, Rwanda, and Montana Hon A *et al.* (20222).

The following would also be part of this: 1) Improve partnerships and teamwork between the public and private sectors at all levels; 2) Give local communities education and training programs along with technical help; and 3) Encourage a wider range of economic activities that make people want to go on eco-tourism trips. By taking these kinds of steps, communities may be able to grow and gain from eco-tourism. It also makes it easier for stakeholders to work together on tourism development and heritage preservation, which helps protect the environment, natural resources, and traditional heritage Jayanthi J *et al.* (2021).

So, if this is used for tourism, it is important to make sure that locals and other stakeholders can grow fairly. To do this, there needs to be a good connection between people and nature. The conservative method of environmental management. The first is living with the world in a way that is careful and planned. The second one is about protecting the earth for only a few good reasons. What this means is that smart use of natural resources should include changing how people do things so that the natural world doesn't get worse. Taking the right steps to stop and control environmental damage is what the third aspect is mostly about. In this way, environmental management is about making sure that people live in harmony with nature by using natural resources wisely and not upsetting the ecological balance and ecosystem stability. It was written in 2000. It was argued by Singh Beena in 1992 that environmental management should be focused on making the best use of natural resources so that they can last forever and meet people's basic needs Malik M *et al.* (2022).

Taking care of the environment doesn't have a single plan. To reach certain goals, it would be best to look at the different parts of the environment and see how they fit together during the growth process. It clearly recognizes natural resources and how they can be used to improve the quality of life and the balance of the environment.

2. Related works

Using Kernel-based SVM and Deep Learning techniques, Altin *et al.* (2023) analyzed the estimated quantity of medical waste produced by a private hospital. For the Kernel-based SVM, we employed the Epanechnikov function, and for the Deep Learning technique, we used the Maxout activation function. The inputs to the model included the total number of operations, outpatients, inpatients, intensive care patients, and intensive care days. Root Mean Squared Error, Mean Absolute Error, and R-squared were also used to assess the quality of the suggested models. Although both algorithms achieve positive results, the Deep Learning approach is more effective. Kernel-based algorithms are expected to play an effective role in medical waste management planning by correctly interpreting the correlation between waste generation and model parameters.

Fractional Horse Herd Gas Optimization was developed by Ramya *et al.* (2023) as a reliable and efficient method for categorizing electronic trash on the IoT cloud platform. The suggested FrHHGO algorithm is used to route the photos of the E-waste acquired by the IoT nodes to their final destinations. In addition, the process of classifying E-waste involves extracting useful traits and then augmenting them. FrHHGO, short for Fractional Henry Gas Optimisation and Horse Herd Optimisation Algorithm, is the algorithm used to categorize E-waste. Extreme accuracy, 0.967, was achieved by the new technique, making it superior to different current ways. The minimum energy required was 0.301 J, and the latency was 0.666 s. By adopting FrHHGO-based ShCNN, the established E-waste categorization method improves social, environmental, and economic sustainability in developing nations.

When it comes to recognizing common garbage, a belt conveyor in a trash collecting system is no match for the deep learning algorithms presented by Li *et al.* (2023), namely CNN and Graph-LSTM. The six object categories of solution. The ability to represent long-term dependencies, together with enhanced efficiency, improved generalization, and convenient access, is a key benefit. The experimental consequences demonstrate that the projected system outperforms the approaches found in the literature, achieving an accuracy of 97.5% in real-world settings.

The best hybrid deep learning model for trash sorting was provided by Zhang *et al.* (2023). Utilizing Convolutional Neural Networks (CNNs) (AlexNet) for feature extraction, Deep Neural Networks (DBNs) for waste prediction from metropolitan city wastes, and Optuna for hyperparameter optimization, the suggested study achieves its goals. When associated with other state-of-the-art methods, this model performed better (R2 = 0.94, MPE = 0.02). This suggested optimized hybrid deep learning model improves performance to anticipate trash creation and categorize it over the individual learners' model.

Using a convolutional neural network (CNN), Chauhan *et al.* (2023) attempted to develop an image classifier that could identify objects and place them in appropriate waste categories. This paper offers a survey of the tools and techniques used to keep tabs on rubbish, sort it out, and get rid of it. Finally, a system and method for disposing of garbage efficiently are shown that may be used in the future to boost efficiency and cut costs. With the help of deep learning, one of the biggest obstacles to effective waste management is being removed. In comparison to AlexNet, VGG16, and ResNet34, the suggested technique performed better.

3. Proposed system

The disposal of electronic trash presents significant threats to environmental quality. As a result, pollution monitoring and control are crucial for maintaining a healthy ecosystem. In this research, picture-layered characteristics were extracted using an encoder composed of varying numbers of ResBlocks and the Squeeze-and-Excitation (SE) block, which was built on top of the UNet backbone network. UNet's decoder structure was streamlined, and the number of network model parameters was cut in half. In the meantime, the multiscale feature fusion module was developed to enhance the network's detection accuracy by optimizing its parameters with a bespoke loss function in place of the standard function. The African Buffalo Optimisation Algorithm is also used to fine-tune the hyper-parameters.

3.1. Dataset

The TrashNet dataset was used in this research and in our investigation. Glass, paper, cardboard, plastic, metal, and photographs of common waste make up the six categories of the dataset. The dataset consists of the following 2527 pictures: There are 501 glasses, 594 papers, 403 cardboard, 482 plastics, 410 metals, and 137 random trash per day. Figure 1 displays a few examples from the data set. Due to the study's emphasis on waste categorization and location, we did not include the daily garbage class in our model's training or evaluation. In boundary boxes were added to the photos in the PASCAL VOC format to indicate which classes they belonged to (Xmin: top left; Ymin: top right; Xmax: bottom left). There were 496 photos total, separated into two sets for training and testing/validation. Seventy-five out of a hundred of the photos were used for training, while the residual 25 per cent were used for testing and validation due to the relatively increased size of the training dataset.

3.2. Dataset augmentation

Dataset augmentation of the training set has been utilized to prevent overfitting; this is done by randomly cropping, flipping, rotating, and scaling images during the training phase. The localization and classification's fully linked layers were given regularisation with a dropout of 0.5 before Softmax to improve generalization.

3.3. Improved network model: RAFF-NET

To achieve end-to-end classification training, RAFF-Net uses the UNet system as its backbone and is primarily composed of three components: an encoder, a decoder, and a multiscale feature fusion module connected to the encoder.

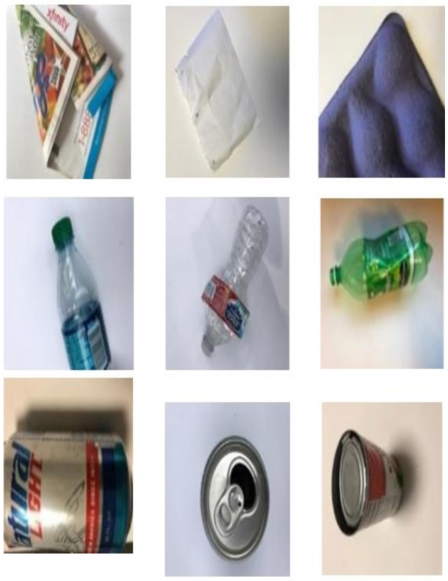


Figure 2. Images after training dataset.

The encoder's primary role is to improve classification accuracy by isolating relevant picture characteristics, eliminating distracting background noise, and saving the high-level semantic information features. Each map goes through two 33 convolutional layers with ReLU functions and a 22 layer, and the original UNet is downsampled four times in the encoder module. Because of ResNet's superior feature extraction capabilities, the enhanced ResNet50 was utilized as the encoder of the new model, and the SE block was fused among ResNet's residual modules to further improve the encoder's feature extraction performance on images with many layers.

The encoder got rid of the fully connected and average pooling layers from ResNet50, but it kept the initial 7x7 convolutional layer of residual modules (3, 4, 6, and 3). Four groups of residual modules were created, with IN and OUT representing the sum of input and output picture channels for each group. The encoder of the refined model reduced the number of channels from the original ResNet50 by half, shifting the residual module's position from channel 32 to channel 1024 in the final output. Meanwhile, the encoder took in images with a 512 x 512 aspect ratio, which is far larger than the standard 224 x 224 ratio for ResNet input images. In Figure 2(b), the SE block follows the sets of residual modules. The SE block's output feature map dimensions matched those of its preceding modules in terms of the number of channels and the aspect ratio of input data. For each SE block, we used a value of 4 for the hyperparameter r. The exact location of the SE blocks will be studied in the experimental phase as it has a bearing on the overall network performance. The initial picture size was 3 512 512, and the encoder downsampled it four times until settling on a size of 1024 16 16 using the tweaked ResNet and SE blocks.

The Identity Block make up the ResBlock in the ResNet50 architecture. The Identity Block has the following components: batch normalization, ReLU activation function, constant connectivity, and two 1 1 convolutional layers. Where W and H are the length and width of the feature map depicted in Figure 3, IN and OUT are the sum of feature map channels prior to and following processing by the integer-taking operation, reveals the difference between the Convolution Block and the Identity Block. The left node corresponds to the convolution dimensions are both variable and series to adjust the network's dimension; the shortcut path (shortcut path) performs an 11 convolution, and the main path (main path) adjusts the network's dimension to match.

Similar to UNet's upsampling method, the decoder module made use of a 2 2 deconvolution with a step size of 2, increasing the length and width of the feature map by 2 times with each upsampling until the original map size of 512 512 pixels was recovered. The original UNet decoder removes two 33 convolutional layers after splicing the feature maps for each layer, reducing computation and the sum of parameters.

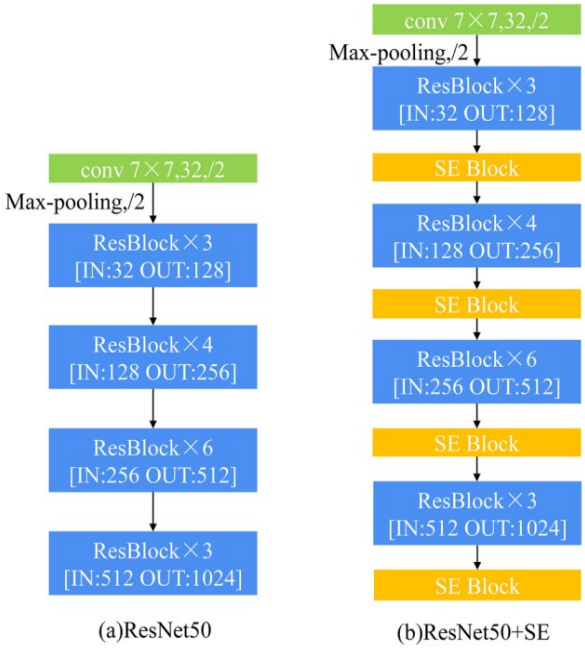


Figure 3. Encoder construction. SE: Squeeze-and-excitation.

The encoder and decoder were connected via a skip link at the relevant feature layer to aid in the recovery of fine-grained location data that had been lost as a result of upsampling in the encoder. With this method, the decoder may make better use of the encoder's precise position information in the shallow network, allowing for more effective picture information recovery. In order to achieve the same sum of channels as the decoder feature layer, our encoder feature layer underwent 1 1 convolution, which was not the case with the original UNet's skip connection method. The encoder's feature map was processed by the residual block, the SE block, and the 1 x 1 convolution layer in that order, while the upper decoder layer's feature map was upsampled and the 1 x 1 convolution layer was used to process it. The procedure began with a splice on the feature map, continued through a batch normalization layer and ReLU function, a 22 deconvolution layer operation, and another splice. A ResBlock module in the subsequent layer received the encoder's output result, and an upsampling module in the decoder enabled recovery from the encoder, where n is the number of ResBlocks in the encoder.

Adding a multiscale feature fusion unit allowed us to further deepen the system and collect finer-grained features at each layer, allowing for more precise picture categorization. Although the higher-level features have greater semantic info, their resolution is very restricted. The lower-level features have better resolve and include more location and part information, but they also have less semanticity and more noise owing to less convolution. The model's framework is depicted in Figure 4.

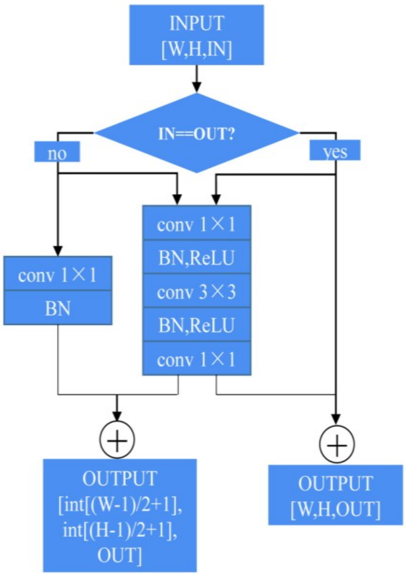


Figure 4. Resblock structure. ReLU: units.

As a result, they lack the ability to notice subtleties. The multiscale feature fusion module of our enhanced model achieved a multiscale layer, which achieved the goal of contrast to the original UNet model, which only performed decoder. In the multiscale feature fusion module, five feature maps were generated for each level (the last encoder level and four decoder levels) using 11 convolution and upsampling at different scales, with feature map dimensions of 1512512. These five maps were then fused, with the fusion map obtained using 11 convolution and a function, yielding a two-channel 512512 probabilistic map as the final output. A likelihood value for each species (background) was assigned to each pixel in the likelihood map. To generate the model's categorization effect map, just extract the index of the pixel's highest probability value.

3.3.1. Loss function

Variations in subject extension strength, for example, might result in a wide range of picture sizes throughout the capturing process. In cases where a classification error is readily apparent, the input pictures' computed loss value is low. The cross-entropy loss function has the benefit of more accurately gauging the deviation between the expected and actual values. Consequently, the function was used as the new model's loss function. To build the weighted loss function, a balancing factor was implemented.

The model established during this study will be refined to calculate camera-object distance and municipal garbage size. One area of use is enabling the mobile robotic scheme to modify its gripper stroke and location in response to recognized items. The best placement of the camera on the mobile robotic scheme is another challenge.

$$L_{total} = \sum_{n=1}^{Q} \omega_n l_n + \omega_{mix} l_{mix} \tag{1}$$

l_n symbolizes the loss last-level encoder, Q = 5; l_{mix} is the Q-level output's inability to retain multiscale feature fusion. We measured the loss in terms of shadows by using the usual cross-entropy for each event.:

$$1 = \sum_i -\left[y_i log a_i + (1-y_i)\log(1-a_i) \right] \tag{2}$$

y_i symbolizes the true charge of the pixel class of the ith picture, and a_i is the foretold value of the pixel category for the ith picture. ω_n and ω_{mix} are the loss. The setting is intended as shadows:

$$\omega = \begin{cases} \dfrac{\sum_{i=0}^{N} |T_i|}{N * |T_i|}, p \in T_i \\ 1, \quad\quad p \notin T_i \end{cases} \tag{3}$$

N signifies the sum of all labels, p characterizes the sum of pixel opinions in the example, and T_i indicates the number of altogether labels.

3.3.2. Hyper-Parameters Tuning

At this stage, all of the models are fine-tuned and improved through the application of transfer learning. At initially, bottleneck layers are not frozen during network tuning. In this way, the models that have already been trained on various datasets serve as a warm-up for our fine-tuned parameters, facilitating the spread of the gradient across the network. After this is done, we will employ transfer learning by preserving the high-level illustrations learnt in the bottleneck levels. Therefore, the gradient can back-propagate via these layers, allowing for fine-grained weight updates, but only at a very low learning rate. We now introduce hyper-parameter tuning for optimal precision. NBC The primary goal of our research was to optimize the most important hyper-parameters that impact object detector training. Hyper-parameters are things like the weight of the l2 regularizer, the x, y, and scales of the box coder, the SSD anchor generators' scales and aspect ratios, the feature extractor's depth multiplier and least depth, and so on. For the projected construction detector, the alteration is:

- *Batch size* = 3 *images*;
- *Learning rate using momentum African Buffalo optimizer*;
- *Initial learning rate*: $1 \times 10 - 2$;
- ***Decay***: **0.95**;
- *Decay steps*: 10,000;
- *Momentum optimizer*: 0.9;
- *Final learning rate*: $1 \times 10 - 5$.

Faster R − CNN parameters for RPN:

- *IoU* = 0.7;
- *Proposed maps*: 100;
- *Crop size*: 17.

A) African Buffalo Optimization

Because of the plentiful grass and secure environment, the buffalo are forced to remain in their current home. However, if the present location's pasture is insufficient, the "waaa" sound is used to investigate potential new pastures. Using these may more efficiently look for profitable feeding areas. This is represented mathematically by equations (4) and (5).

$$mk+1 = mk + lp1\left(bgmax - wk\right) + lp2\left(bgmax.k - wk\right) \quad (4)$$

where mk denotes a "maaa" sound with a exact orientation to a buffalo k (k = 1,2,3, … n) , Learning parameters lp1 and lp2 are [0,1]; we may infer that the buffalo has moved from its present position mk to a new site with a higher memory capacity, which is consistent with its migratory lifestyle. The true correction of herd motion is achieved by mathematical modelling (5).

$$wk+1 = \left(wk + mk\right) / \acute{o} \quad (5)$$

During each cycle, the fitness value for each buffalo is calculated; the best value is given to bgmax (the best global solution), while the best value for each individual is given to bgmax. k (the best local solution). Each buffalo mechanically updates its site, and it moves to be near its best conceivable neighbour, as determined by rules (4) and (5). This revision towards the optimal key.

Algorithm 1: The ABO procedure

Step 1: ***initialization*** *Random i initialize kth buffalos' location on solution space*
Step 2: *Evaluation of buffalos' fitness value and assigning the herd's best to* bg_{max} *and individual buffalo's best to* $bg_{max,k}$
Step 3: *Update (Exploitation move)*
Update the buffalos' fitness value according to equation (4)
Step 4: *Update (exploration move)*
Update the movement of buffalo according to (5)
Step 5: *Is* bg_{max} *Updating Yes, go to step 6, No go to Step 1.*
Step 6: *Check the validation of stopping criteria; if satisfied Yes, go to step 7. No, go to step 2.*
Step 7: *Return the best solution so far.*

4. Results and discussion

In addition, we generated a bespoke minor set of images with twenty-five different photos of glass bottles, completely different from the ones the model is tested against, in order to evaluate our models' generalizability beyond the TrashNet data set in real time. Only a tiny collection of photos have been examined for testing thus far because the bottle detection technique in municipal waste requires the recognition of items with varying degrees of chromatic aberration in shape, position, and clarity. Bottles are the single most common item found in landfills, and our tiny collection is comprised entirely of bottles, both plastic and glass.

The most important detections from the case study are shown in Figure 5. The glass and plastic bottles in the picture are upside down, sideways, crooked, misshapen, partially obscured, etc.

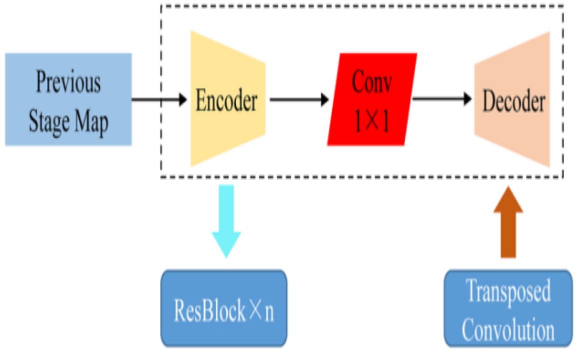


Figure 5. Grouping of low and deep feature map.

On our custom dataset, the implication time evaluation was done with a resolution increase from 300x300 to 500x500. A resolution of 500500 is needed to see even the smallest bottles and cans. While the localization was precise and the regular confidence was 75.54 percent, one object (a plastic bottle) had the wrong label and should have been classified as a can (Figure 6).

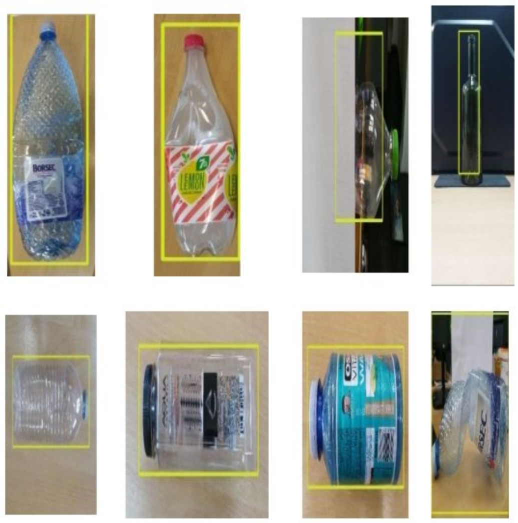


Figure 6. Custom dataset forecasts for one object in image.

Ubuntu 20.04 served as the experimental platform for this research. To conduct the tests, we used Python 3.7, PyTorch 1.8. All systems were trained using the dual-card-distributed hybrid training approach on two NVIDIA GPUs, and all epochs were set to 300.

4.1. Evaluation Metrics

4.1.1. Precision

Precision is distinct as the proportion of positive sample predictions that turn out to be right..

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

When both the label value and the prediction are positive, we say that the forecast is true positive (TP), and when both are positive, we say that the prediction is *false positive* (*FP*).

4.1.2. Recall

The term "recall" refers to the likelihood that an expected good outcome really is a positive sample.

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

4.1.3. Mean Average Precision

The mAP characterizes the average precision (AP) be around over all categories.

$$mAP = \frac{1}{N} \sum AP_i \tag{8}$$

where N is the entire sum of classes and APi is the average precision in class i, and where mAP@0.5 is the value of the IoU parameter representing the mean accuracy when the threshold is set to 0.5. The experimental comparison of existing representations to the projected model is presented in Tables 1 and 2.

Table 1: Experimental investigation of various models on 70%-30%

Method	Recall	mAP@0.5	Precision	FLOPs
LSTM	0.930	0.964	0.921	15.8
CNN	0.939	0.950	0.930	15.8
RNN	0.947	0.956	0.937	15.8
Proposed model	0.956	0.972	0.959	16.0

Table 2: Analysis of proposed model on 80%-20%

Method	Recall	mAP@0.5	Precision	FLOPs
CNN	0.939	0.950	0.930	15.8
RNN	0.956	0.972	0.959	15.9
LSTM	0.945	0.967	0.953	15.9
Proposed model	0.962	0.968	0.961	16.9

Table 1 above describes the experimental examination of different models on a 70%–30% basis. In the examination of the CNN approach, the precision and recall rates were 0.930 and 0.939, respectively, followed by a mAP@0.5 range of 0.950 and a FLOPs value of 15.8. The RNN approach then achieved 0.937 precision and 0.947 recall rate, followed by 0.956 mAP@0.5 range and 15.8 FLOPs, respectively. Then the LSTM method reached the precision as 0.921 and recall rate as 0.930 and then mAP@0.5 range of 0.964 and then FLOPs as 15.8 correspondingly. Then the Proposed model reached the precision of 0.959 and recall rate of 0.956, and the mAP@0.5 range of 0.972, and then FLOPs as 16.0, correspondingly (Figure 7).

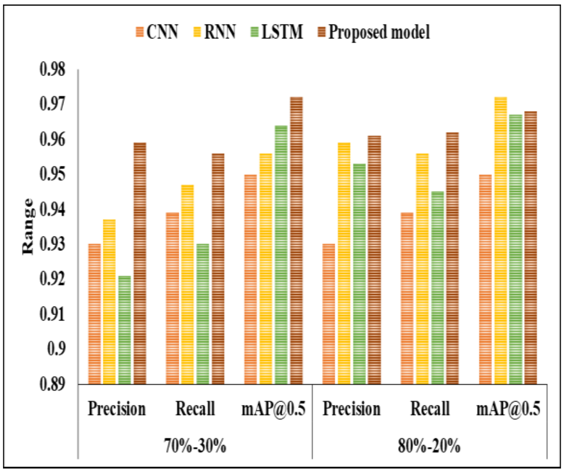


Figure 7. Analysis of various model for waste classification

In Table 2 represent that the Analysis of proposed model on 80%-20%. In the analysis of the CNN method reached the precision of 0.930, the recall rate was 0.939 the mAP@0.5 range of 0.950, and then FLOPs at 15.8 correspondingly. Then the RNN method reached the precision as 0.959 and recall rate as 0.956 and then mAP@0.5 range of 0.972 and then FLOPs as 15.9 correspondingly. Then the LSTM method reached the precision as 0.953 and recall rate as 0.945 and then mAP@0.5 range of 0.967 and then FLOPs as 15.9 correspondingly. Then the Proposed model reached the precision of 0.961 recall rate of 0.962 and the mAP@0.5 range of 0.968, and then FLOPs of 16.9 correspondingly.

5. Conclusion

This research gives a comprehensive and in-depth analysis of object detectors based on convolutional neural networks, specifically as they relate to the classification of municipal garbage. The primary goal was to improve the presentation of existing pre-trained models. Data augmentation, CNN model, dataset size, loss optimization, hyper-parameters modifying, evaluation so on all have a role in how well a CNN object detector performs. Therefore, in this study, a complicated series of procedures was performed to design a precise and real-time architecture. This network's results on the Trash Net dataset are compared to those of other convolutional neural networks. When compared to other CNN Image Classifiers, our fine-tuned perfect achieved the greatest accuracy possible for object detectors. All of the models utilized in this study improved performance, despite the object indicator construction being more complicated than a CNN for accuracy being tested in a different way. The concept will eventually be put into action in a real-world setting, most likely with the help of a Raspberry 3+ board. Asynchronous threading and an extra Intel neural network processor were used to reduce the detection time.

References

Altin, F. G., Budak, İ. and Özcan, F. (2023). Predicting the amount of medical waste using kernel-based SVM and deep learning methods for a private hospital in Turkey, *Sustainable Chemistry and Pharmacy,* **33**, 101060.

Anuradha M., Jayasankar T., Prakash N. B., Sikkandar M. Y., Hemalakshmi G. R., Bharatiraja C. and Britto A. S. (2021). IoT enabled cancer prediction system to enhance the authentication and security using cloud computing, *Microprocessors and Microsystems*, **1**;80:103301.

Chauhan, R., Shighra, S., Madkhali, H., Nguyen, L. and Prasad, M. (2023). Efficient Future Waste Management: A Learning-Based Approach with Deep Neural Networks for Smart System (LADS), *Applied Sciences,* **13(7)**, 4140.

Chen G. and Alampalli S. Proceedings of the 9th International Conference on Structural Health, Monitoring of Intelligent Infrastructure: Transferring Research into Practice.

Daulay HY. (2020). Getting the History of Hand Washing as A New Normal Activity Part, *InThe 5 Th International Seminar On Social Studies and History Education (ISSSHE)* **12(204)**.

Hon A. (2022). You've Been Played: How Corporations, Governments and Schools Use Games to Control Us All, *Swift Press;* **23**, 15.

Jayanthi J., Lydia E. L., Krishnaraj N., Jayasankar T., Babu RL. and Suji R. A. (2021). An effective deep learning features-based integrated framework for iris detection and recognition, *Journal of ambient intelligence and humanized computing*, **12**, 3271–3281.

Kljajin M. Foreword/Uvodna rije, *Tehnicki Vjesnik-Technical Gazette*. 13, **20(2)**, III

Li, N. and Chen, Y. (2023). Municipal solid waste classification and real-time detection using deep learning methods. *Urban Climate,* **49**, 101462.

Ludwig H., Schmidt T., Kühn M. (2023). Voice user interfaces in manufacturing logistics: a literature review, *International Journal of Speech Technology*, **1**, 1–3.

Ramya, P. and Ramya, V. (2023). E-waste management using hybrid optimization-enabled deep learning in IoT-cloud platform, *Advances in Engineering Software,* **176**, 103353.

Wazid M., Das A. K., Shetty S., JPC Rodrigues J. and Park Y. (2019). LDAKM-EIoT: Lightweight device authentication and key management mechanism for edge-based IoT deployment, *Sensors,* 14, **19(24)**, 5539.

Zhang X., Eltouny K., Liang X. and Behdad S. (2023). Automatic Screw Detection and Tool Recommendation System for Robotic Disassembly, *Journal of Manufacturing Science and Engineering*, **145(3)**, 031008.

Zhang X., Eltouny K., Liang X. and Behdad S. (2023). Automatic Screw Detection and Tool Recommendation System for Robotic Disassembly, *Journal of Manufacturing Science and Engineering*, **145(3)**, 031008.

Zhang, H., Cao, H., Zhou, Y., Gu, C. and Li D. (2023). Hybrid deep learning model for accurate classification of solid waste in the society, *Urban Climate,* **49**, 101485.