ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse120952020.pdf
https://doi.org/10.30534/ijatcse/2020/120952020

Performance of Faster R-CNN to Detect Plastic Waste

Moechammad Sarosa[1], Nailul Muna[2], Erfan Rohadi[3]
[12]Electrical Engineering Departement, State Polytechnic of Malang, Malang, Indonesia
[3]Information Technology Department, State Polytechnic of Malang, Malang, Indonesia

ABSTRACT

Waste production in Indonesia reaches more than 65 million tons per year, which is a problem that has not been appropriately resolved. The government and organization must raise awareness of waste because improper waste management can cause pollution**.** The organization can manage waste by recycling to solve this problem. However, these efforts have encountered obstacles because public awareness of sorting waste is still deficient. Therefore, we need a sorting system for organic and inorganic waste. Reducing waste problems can be started by separating plastic waste from organic waste. This study aims to detect plastic waste and other separate plastic waste from organic waste. This plastic waste detection system will extract digital image features to detect plastic waste using the deep learning method. Deep learning is used because it is proven to have good performance. Faster Regional Convolutional Neural Network (Faster R-CNN) is a deep learning technique recently developed to recognize and classify computer vision. Faster R-CNN is an algorithm that utilizes the Convolutional Neural Network (CNN) in the object detection process. This study will detect and localize plastic waste in the image using the Faster R-CNN algorithm by utilizing the TensorFlow object detection framework. The system implements four network architecture models, namely Inception Resnet V2, Inception V2, Resnet 101, and Resnet 50. Based on the experiments that have been conducted, Faster R-CNN can provide good performance by obtaining an F1 Score of 93% on the Inception Resnet V2.

Key words: Deep Learning, Convolutional Neural Network (CNN), Object Detection, Plastic Waste Detection, Faster R-CNN.

1. INTRODUCTION

Indonesia is the fourth country with the most population globally after China, India, and the United States. Based on BPS data, Indonesia's population in 2018 is 265 million, and in 2025 it will increase to 285 million. The increasing population can increase the use of plastic, thus creating an increase in the amount of plastic waste produced. Each household in Indonesia has 0.52 kg of waste/person/day. Indonesia's economic growth is also a factor in the increasing use of plastics. In 2015, Indonesia was the second country to pollute the sea with plastic waste after China. The Indonesian government is committed to reducing the volume of plastic waste by 30% in 2025. But the progress has not been satisfactory until now [1].

High waste production, if not accompanied by proper management, will cause pollution. Waste management needs to be supported by adequate facilities and infrastructure [1]. Image processing is a method that is a solution to help these problems [2]. Image processing is an increasingly evolving field within the scope of research where the real-time observation system increases the opportunity for researchers to develop new research on a variety of problems [3]. Object detection is one of the most challenging issues in the field of computer vision [4]. The development of increasingly rapid computer vision is shown by the emergence of new methods and algorithms so that the process of object recognition/identification becomes faster and more accurate [5]. Object detection is related to identifying real-world objects such as people, animals, and other objects. The object detection algorithm uses various image processing applications to extract the purpose's desired part [6].

Image processing can apply the deep learning method, which is part of Artificial Intelligence (AI). Deep learning is a method that can follow humans' learning approaches through data patterns with varying abstraction [2][7]. Deep learning can build features without human supervision and has the capability of being faster and more accurate [2].

One model of deep learning is Convolutional Neural Networks (CNN). Object detection applies CNN to solve problems, and processes based on CNN have achieved sophisticated performance [8][9]. Girshick et al. propose CNN with the proposed region combined on R-CNN. The proposal is a selective search for extracting 2000 regions [10]. So that object search is not carried out on all parts of the image. R-CNN classifies only the selected portion of the image. The same author developed the R-CNN algorithm to overcome the R-CNN algorithm's limitations, where the input to the convolutional feature map is an image. Then the feature map identifies the area proposal. The research was further developed by replacing the region search with an object detection algorithm. This development is called the RCNN Faster proposed by Shaoqing Ren et al. [11].

Based on the problems, this study proposes a waste detection system using the Faster R-CNN algorithm. This system can detect and localize waste that implements image processing with Faster R-CNN, which utilizes CNN and deep learning techniques. There are four network architectures used, including Inception Resnet V2, Inception V2, Resnet 50, and Resnet 101. This study aims to detect waste objects with better performance so that the results of this study can be useful as a method for sorting plastic waste.

2. RELATED WORK

The deep learning method has achieved excellent results in a variety of machine learning applications, including computer vision, robotics, pattern recognition, recommendation systems, data analysis, voice recognition, self-driving cars, sound generation, natural language processing, bioinformatics, cellular advertising, visual arts processing [12], and scene labeling [13]. Deep learning can learn abnormal patterns automatically and saves a lot of detection time [14]. Deep learning methods, such as the Convolutional Neural Network (CNN), can produce hierarchical representations of an image that can create a classification [15]. CNN can identify visual objects from pixel images using small preprocesses [16]. The effect of applying the CNN feature for object detection systems has been studied by Girshick et al., who showed high-level performance in object detection by combining several low-level image features with a top-level context. To produce the proposed region, extract CNN features, and classification, researchers use different methods. Girshick enhances the algorithm of CNN with various functions provided to classify the proposed region. Region Proposal Network (RPN) and detection network are developments in CNN features. The CNN feature algorithm for object detection carried out a more considerable growth that leads to the detection and localization system. YOLO one-shot detector can predict bounding boxes and scores directly from the map feature. SSDs in different ways improve detection performance by using multi-scale convolutional predictors on several feature maps that produce better predictions of various aspects of size and ratio [8]. One of the algorithms from CNN is Faster R-CNN [11].

Abbas et al. conducted research focusing on Faster R-CNN training to recognize interest objects using custom datasets. RPN is an underlying part of the Faster R-CNN algorithm. RoI Pooling uses RPN output for classification. In this study, testing is carried out on the vehicle dataset image. To measure the performance of predicted results using the Computer Vision System Toolbox ™ and in the training process, using an NVIDIA ™ GPU supports CUDA with computational capabilities reaching computing 3.0 or more. Networks are trained using the train-FasterRCNNObjectDetector toolbox [17].

Liu et al. train hard negative samples on research conducted to improve the performance of Faster R-CNN. The model is obtained from the first training to enter the network. On the other hand, alternating training methods are used to build RPN and Fast R-CNN on the conventional layer of the Faster R-CNN algorithm. The results showed that the algorithm applied had a detection speed performance that was no better than the YOLO, SSD, and RFCN methods [18].

Wang et al. conducted a study for bottle detection on UAV images such as the Faster R-CNN, SSD, YOLOv2, and Faster RPN by building large-scale datasets. Label several bottles then distribute the bottles in the orientation towards the bounding box. The dataset can set the bottle detection benchmarks and build bounding boxes that can provide information on the bottle's object where the bottle is handy for recycling. So in this study, a plastic waste detection system was developed by implementing the Faster R-CNN algorithm, a sophisticated algorithm in object detection [9].

3. PROPOSED METHOD

The proposed method consists of several stages, including data annotation, data training, and data testing. Data annotation is done using LabelImg software. There are 300 data used in the data annotation process. Annotated data is divided into two parts, including 200 training data and 100 testing data. Then 200 data were conducted training using the pre-trained Inception Resnet V2, Inception V2, Resnet 50, and Resnet 101 models with the training method, namely Faster R-CNN. The model studied predicts the bounding box and class of each data in the data testing process. The data training process is the process of evaluating data that has been trained to determine the performance of the learned model. The proposed method is shown in Figure 1.

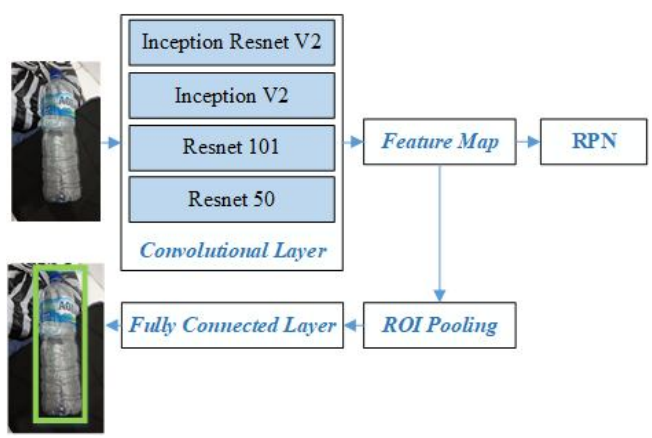

Figure 1: The proposed method for waste plastic detection

Figure 1 shows the process starts with a convolutional process. The convolution process uses four architectural models, including Inception Resnet V2, Inception V2, Resnet 50, and Resnet 101. The four architectures produce feature map output for the proposal region. The anchor boxes in each region define the Region Proposal Network (RPN). The RPN makes proposals at various scales. Process the detected object from bounding box with feature map. Bounding boxes with different sizes, RoI pooling layers, and aspect ratios are resized using max pooling. The collected feature maps are then classified so that the image output with the bounding box and the class name is obtained due to the classification.

3.1 Data Annotations

Data annotation is the process of labeling images. Data annotation is done by labeling the image's waste by manually tagging all the images, training data and test data. The annotation resulting from this process is in the form of a rectangular bounding box, which is four sides around the outside of the plastic waste area. Information obtained from plastic waste labeling results of bounding box coordinates and dimensions of length and width, object class names, and related image files is stored as .xml files in PASCAL VOC format. The .xml file is converted to a .csv file, so the annotation results are ready to be used as input for the TensorFlow Object Detection to produce the TFRecords file. The TFRecords file will generate a train.record file and a test.record file that is used to train the plastic waste recognition detection classifier. The advantage of using TFRecord is that the data processed from the disk is the data required at a certain time [19].

3.2 Faster R-CNN Algorithm

Ross B. Girshick proposed the Faster R-CNN which is a deep learning approach to object recognition and classification which is currently receiving a lot of attention from researchers. Faster R-CNN is a development of R-CNN and Fast R-CNN which was proposed in 2016. The overall performance of Faster R-CNN is improved the target object can be detected more quickly. The convolutional network on an R-CNN Faster can generate the proposed grid, which reduces the number of frames proposed box [20][21]. Faster R-CNN architecture is shown in Figure 2.

Figure 2 shows that Faster R-CNN consists of the Region Proposal Network (RPN) and Fast R-CNN. The RPN is the most critical part of the R-CNN Faster used to produce good regional proposals. The RPN has the role of a fully convolutional network. This feature is used to generate proposal regions and convolutional neural network layers to produce higher-dimensional features. This feature is distributed to RPN. Region proposals and higher-dimensional features are used as input to find the Region of Interest (RoI). This is used to convert higher dimension features to a size appropriate to the region proposal. The RoI collection features are incorporated into a fully connected layer consisting of a box-regression layer (reg) to get coordinates and a box-classification layer (cls) to get a score. This stage is to classify the areas of interest identified in the steps above into the appropriate class. The technique used here is Convolution Neural Networks (CNN) [18].

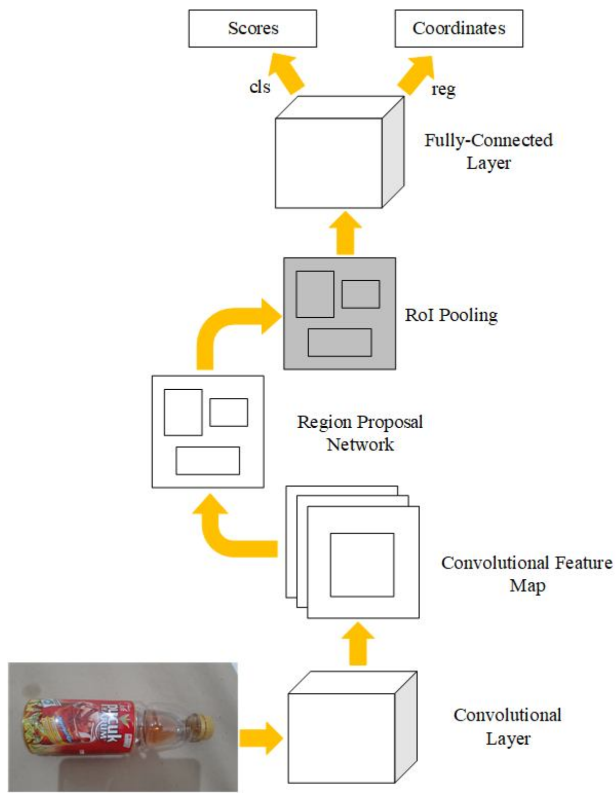

Figure 2: The Faster R-CNN architecture [17]

3.3 Regional Proposal Network (RPN)

RPN on the Faster R-CNN is used to extract the proposal area of the target object. The RPN process starts from the network at the last convolutional layer that slides over the feature mapping. The feature map is fully connected to a window of the same size. Then map each slide window to a vector of length 512. Finally, the lower dimensional vectors are fed into the fully connected layer, the regression box layer (reg), and the box classification layer (cls). The cls layer produces 4k coordinates for each box. The reg layer provides a 2k score for each sliding window location that estimates an object (foreground) or not an object (background). This principle is shown in Figure 3 [18].

An anchor is defined at the midpoint of a sliding window with an area of n × n. Each sliding window can predict the possibility of a proposal simultaneously at each sliding window location. The reference box k can be obtained from aspect ratio, scale, and anchor. By default, three aspect ratios and three scales can be used to get the reference box k, which is nine at each sliding position. p is calculated for each anchor that measures the possibility of overlapping anchors with the object's boundary area. There are two anchor categories: positive and negative classes, with an overlapping ratio of Intersection-over Union (IoU) between anchor box and ground truth for classification [18]. Classification indicates the probability of 0 or 1, indicating whether the region contains objects or not. The classification rules are as follows [17] [18]:

1. If the anchor box has the largest IoU,,the corresponding anchor p^* is a positive sample.
2. If the IoU is higher than 0.7, the appropriate anchor is a positive sample.
3. If the IoU anchor box is lower than 0.3, the appropriate anchor is labeled as a negative sample.
4. If it is not a positive sample and a negative sample, it is not used.

Anchor boxes that cross the image boundary are also discarded.

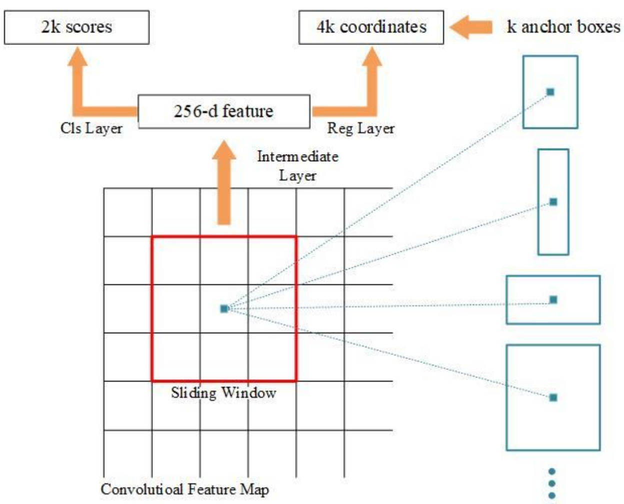

Figure 3: The principal of Region Proposal Network (RPN) [17][18]

The overlapping ratio of IoU is defined as

$$IoU = \frac{S_{AnchorBox} \cap S_{GroundTrut\ h}}{S_{AnchorBox} \cup S_{GroundTrut\ h}} \quad (1)$$

3.4 Loss Function

Loss on the Faster R-CNN can be defined as the output of classification and regression networks. The loss function is formulated as follows [18].

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}} \sum_{i=1}^{N_{cls}} L_{cls}(p_i, p_i^*) + \frac{\lambda}{N_{reg}} \sum_{i=1}^{N_{reg}} p_i^* L_{reg}(t_i, t_i^*) \quad (2)$$

Where i is an index of the i -th anchor box and p_i is the probability that an anchor i is an object. The corresponding probability p_i^* is set to 1 if the anchor is a positive sample. And the probability p_i^* is set to 0 if the anchor is a negative sample. t_i is a vector that represents the four coordinates of the bounding box parameter. t_i^* is a vector of coordinates of the corresponding ground-truth bounding box. N_{cls} and N_{reg} are the normalization coefficient of the L_{cls} is a loss function classification and L_{reg} is a regression loss function. λ is the weight parameter between L_{cls} and L_{reg}. The L_{cls} loss function classification is a logarithmic loss of the target category or not. L_{cls} can be defined as follows [22].

$$L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1-p_i^*)(1-p_i)] \quad (3)$$

The regression loss function is defined as follows.

$$L_{reg}(t_i, t_i^*) = R(L_{cls}(t_i, t_i^*)) \quad (4)$$

Robust loss function or which can be written with R can be defined as follows.

$$R(x)\begin{cases} 0.5x^2, |x| < 1 \\ |x| - 0.5, other \end{cases} \quad (5)$$

Regression to determine the four coordinate bounding boxes is as follows:

$$t_x = \frac{x - x_a}{w_a}, t_w = \log\left(\frac{w}{w_a}\right) \quad (6)$$

$$t_y = \frac{y - y_a}{h_a}, t_h = \log\left(\frac{h}{h_a}\right) \quad (7)$$

$$t^*_x = \frac{x^* - x_a}{w_a}, t^*_w = \log\left(\frac{w^*}{w_a}\right) \quad (8)$$

$$t^*_y = \frac{y^* - y_a}{h_a}, t^*_h = \log\left(\frac{h^*}{h_a}\right) \quad (9)$$

Where x, y, w, and h are the midpoint, width, and height of the predicted box x_a, y_a, w_a, and h_a are the midpoints, widths, and heights of the anchor box. And x^*, y^*, w^*, and h^* are the midpoints, widths, and heights of ground-truth bounding boxes. The fine tuning process of Ground-Truth, proposal, and anchor box is illustrated in Figure 4.

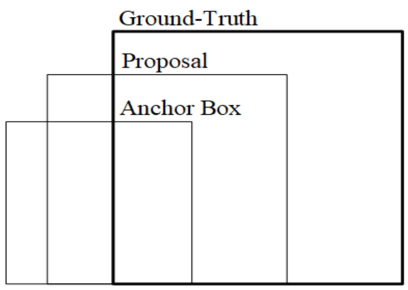

Figure 4: The Process fine tuning

The coordinates of the various aspect ratios and scaling of the anchors are shown in Figure 5.

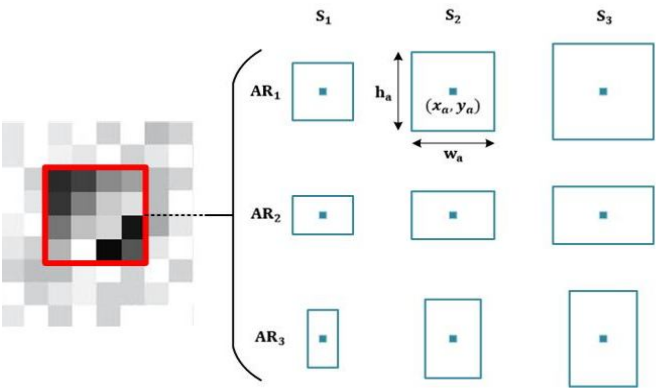

Figure 5: Different Aspect-Ratio and Scaling of the anchor box [17]

4. EXPERIMENTS AND RESULTS

This part is evaluating the performance of the waste detection system. In this paper, the pre-trained model used is faster_rcnn_inception_v2_coco_2018_01_28. The performance is represented using the detection result.

4.1 Dataset

In this study, the dataset was collected by collecting images of plastic waste. There are two types of datasets used, namely training data and testing data. The training data used consists of 200 images and 100 images for testing data. The sample data used is shown in Figure 5.

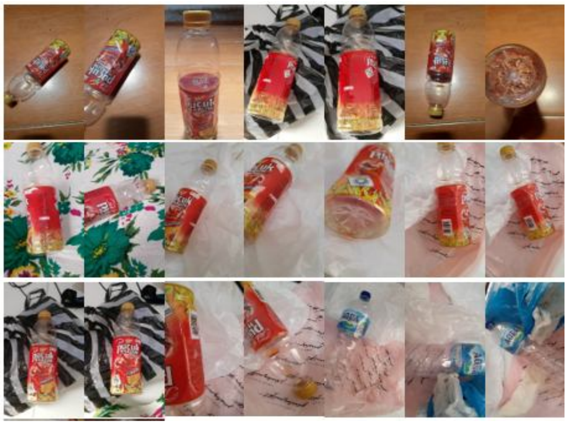

Figure 5: Some samples dataset used in this research

Based on Figure 5, the plastic waste used is plastic bottle waste from several brands of beverage bottles in Indonesia. The picture is taken using the phone camera.

4.2 Implementations

The waste detection system is created using the TensorFlow Object Detection framework. The training and test process is run on a computer with Windows 10 64 bit OS specifications supported by an Intel Core i5 5600U dual-core 2.3 GHz up to 2.8 GHz and 8 GB DDR4 RAM, equipped with NVIDIA GeForce GTX 930MX.

4.3 Evaluating of Training Process

In this study, the training process was carried out in the 4000 steps. From the training process, it can be seen that the computational training time for each architectural model. Computing time is the time required for the computation process done by the computer when solving problems using the algorithm used. The training computation time is shown in Table 1.

Table 1: The Computation Time of the Training Process

Network Architecture	Computation Time
Inception Resnet V2	58h 30m
Inception V2	6h 14m
Resnet 50	15h 59 m
Resnet 101	22h

Based on the Table 1, Inception v2 has the fastest computation time compared to other network architecture models. Inception V2 applies the smart factorization method to reduce computation complexity by changing 5x5 convolution to 3x3. It has faster computation time than other network architecture models both in the training process and in the testing process.

From the training process that has been done can be obtained the value of the total loss. Total loss is the total of all errors that occur during the dataset training process. The total loss is used to determine how well the model can learn the object features at each step during the training process. Loss functions describe the errors that the model makes. The results of the total loss can be seen in Table 2.

Table 2: The Total Loss Result

Network Architecture	Total Loss
Inception Resnet V2	0,1213
Inception V2	0,2792
Resnet 50	0,2192
Resnet 101	0,1343

Table 2 shows the smallest average total loss is Inception Resnet V2, Resnet 101, Resnet 50, and Inception V2, respectively. The smaller the total loss value, the better the model will learn the object's features.

4.4 Evaluating of Detection Performance

After the training process is carried out, the results of the training will be implemented in the waste detection system. Samples of plastic waste detections are shown in Figure 6.

Based on the detection results in Figure 6, the data obtained were divided into 3 data groups, namely True Positive (TP), False Positive (FP), and False Negative (FN). True Positive is when the system can detect and localize objects correctly. A false Positive is an object that can detect an object but on the wrong object. False Negative is the system can not recognize the object in the image that has been inputted. The results of the detection and localization of waste objects on the system are shown in Table 3.

Figure 6: Sample of plastic waste detection results

Table 3: Waste Detection Result

Network Architecture	TP	FP	FN
Inception Resnet V2	86	3	11
Inception V2	65	28	7
Resnet 50	67	25	8
Resnet 101	80	12	8

Based on Table 3, t the highest detection result value is in the Inception Resnet V2 architectural model, which TP has 86, FP has 3, and FN has 11 with a total of 200 training data and 100 test data. From the data obtained can be known as a precision, recall, and F1 Score through the calculation as follows:

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FP} \tag{11}$$

$$F1Score = 2 \times \frac{Precision \times Recall}{Precisiom + Recall} \tag{12}$$

The following values obtained from the calculation of Precision, Recall, and F1 Score are shown in Table 4.

Table 4: Performance of Faster R-CNN for Waste Detection System

Network Architecture	Precision	Recall	F1 Score
Inception Resnet V2	97%	89%	93%
Inception V2	70%	90%	79%
Resnet 50	73%	89%	80%
Resnet 101	87%	91%	89%

Table 4 the Inception Resnet V2 network model is a combination of Inception and Resnet architectures, where the architecture has parallel feature extraction. So the Inception Resnet V2 architectural model allows the model to learn multi-level features from multiple files to improve performance so that the Inception Resnet V2 model has a higher F1 score than other architectural models than the Inception or Resnet model only.

5. CONCLUSION

This paper presents a deep learning method that utilizes the Faster R-CNN algorithm to detect and localize plastic waste. The data used consisted of 200 training data and 100 testing data. There are four network architecture models used, namely Inception Resnet V2, Inception V2, Resnet 50, and Resnet 101. Based on the experiments, Faster R-CNN has a good performance reaching 93% on the Inception Resnet V2 architecture. In the future, this research can be developed by implementing a system on automatic waste collection and sorting machines

REFERENCES

1. P. Purwoko and T. Wibowo, "Fiscal Incentives and Disincentives to Reduce Plastic Waste in Indonesia," in The 6th Asian Academic Society International Conference (AASIC), Thailand, 2018.
2. A. P. Anadkat, B. V. Monisha, M. Puthineed, A. K.

Patnaik, S. R and R. Syed, "Drone based Solid Waste Detection using Deep Learning & Image Processing," in Alliance International Conference on Artificial Intelligence and Machine Learning (AICAAM), Bangalore, India, 2019.

3. K. M. Angelo, "A novel approach on object detection and tracking using adaptive background subtraction method," in International Conference on Computing Methodologies and Communication, Chidambaram, India, 2018.
4. M. J. Shafie, B. Chywl, F. Li and A. Wong, "Fast YOLO: A Fast You Only Look Once System for Real-time Embedded Object Detection in Video," Canada, 2017.
5. B. N. Prastowo, N. A. S. Putro, O. A. Dhewa and A. M. H. Yusuf, "Pengenalan Personal Menggunakan Citra Tampak Atas pada Lingkungan Cashierless Store," Yogyakarta, 2019.
6. A. Raghunandan, Mohana, P. Raghav and H. V. R. Aradhya, "Object Detection Algorithms for Video Surveillance Applications," in International Conference on Communication and Signal Processing, Chennai, India, 2018.
7. M. Akour, O. A. Qasem, H. Alsghaier and K. Al-Radaideh, "The Effectiveness ofUsing Deep Learning Algorithms in Predicting Daily Activities," International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE), vol. 8, no. 5, pp. 2231 - 2235, 2019.
8. S. Alfasly, B. Liu, Y. Hu, Y. Wang and C.-T. Li, "Auto-Zooming CNN-Based Framework for Real-Time Pedestrian Detection in Outdoor Surveillance Videos," IEEE Access , vol. 7, pp. 105816 - 105826, 2019.
9. J. Wang, W. Guo, T. Pan, H. Yu, L. Duan and W. Yang, "Bottle Detection in the Wild Using Low-Altitude Unmanned Aerial Vehicles," in 21st International Conference on Information Fusion (FUSION), Cambridge, UK, 2018.
10. F. A. Hermawati, H. Tjandrasa and N. Suciati, "ombination of Aggregated CChannel Features (ACF) Detector and Faster R-CNN to Improve Object Detection Performance in Fetal Ultrasound Images," International Journal of Intelligent Engineering & System, vol. 11, no. 6, pp. 65-74, 2018.
11. K. Ouni, A. Ammar, A. Koubaa, T. Khursheed and B. Benjdira, "Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3," in International Conference on Unmanned Vehicle Systems (UVS), Muscat, Oman, 2019.

12 M. E. Fouki, N. Aknin and . K. E. E. Kadiri , "Multidimensional Approach Based on Deep Learning to Improve the Prediction Performance of DNN Models," International Journal of Emerging Technologies in Learning (iJET), vol. 14, pp. 30-41, 2019.

13. S. Angadi and V. S. Reddy, "Hybrid Deep Network Scheme for Emotion Recognition in Speech," International Journal of Intelligent Engineering & Systems, vol. 12, no. 3, pp. 59-67, 2019.
14. H. Huang, H. Deng, J. Chen, L. Han and W. Wang, "Automatic Multi-task Learning System for Abnormal Network Traffic Detection," International Journal of Emerging Technologies in Learning (iJET), vol. 13, pp. 4-20, 2018.
15. A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam and M. Nielsen, "Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network," in International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg, 2013.
16. N. A. Al-johania and L. A. Elrefaei, "Dorsal Hand Vein Recognition by Convolutional Neural Networks: Feature Learning and Transfer Learning Approaches," International Journal of Intelligent Engineering & System, vol. 12, no. 3, pp. 178-191, 2019.
17. S. M. Abbas and D. S. N. Singh, " Region-based Object Detection and Classification using Faster R-CNN," in Computational Intelligence and Communication Technology (CICT 2018) , Ghaziabad, India, 2018.
18. Y. Liu, "An Improved Faster R-CNN for Object Detection," in 11th International Symposium on Computational Intelligence and Design, Hangzhou, China, 2018.
19. A. Raorane, S. Patil and L. Kurup, "Logo Recognition Using Deep Learning and Storing Screen Time in MongoDB Database," International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE), vol. 8, no. 5, pp. 2535 - 2539, 2019.
20. B. Liu, W. Zhao and Q. Sun, "Study Of Object Detection Based On Faster R-CNN," in Chinese Automation Congress (CAC), Jinan, China, 2017.