Cleaning Robot Vision System Based on RGBD Camera and Deep Learning YOLO-based Object Detection Algorithm

Shimaa-Amin Bergies, Phuc Thanh-Thien Nguyen, and Chung-Hsien Kuo*, *Member, RST*

***Abstract*— Autonomous cleaning for a wide environment is a big challenge due to energy and time consumption. This paper introduces a novel vision strategy for the auto-guided cleaning vehicle to wash dirty areas in the indoor environment. A modified YOLOv3-based deep learning algorithm is utilized to detect trash objects and the location of soiled regions. Furthermore, a dataset with various uncleaned floors is collected to deal with solid, liquid, or reflected trash under the supervised fashion. The experimental results confirmed that the effectiveness of the proposed autonomous cleaning system could handle the detected uncleaned areas with low energy and time consumption compared with other cleaning systems.**

***Index Terms*— Cleaning Robot, YOLOv3, RGBD camera, Point Cloud.**

I. Introduction

CLEANING floor is an important factor to avoid accidents and diseases in daily life. However, it is a repeating, boring task and consumes long working hours. In some cases, it is serious work, especially during epidemics or in hazardous environments, such as hospitals, factories, and laboratories. Therefore, the presence of autonomous cleaning robots has become an urgent necessity. Recently, various mobile robots have been devoted to different cleaning purposes, as facade cleaning [2], staircase cleaning [3], pavement cleaning [4], garden cleaning [6], and floor cleaning [7]. However, it is not smart enough to perform the cleaning process autonomously [7] because they had inefficient performance while consuming more energy during the cleaning process, especially in large, open areas. Moreover, one of the most important challenges of cleaning robots is missing uncleaned places during the cleaning process.

Various learning methods have recently been introduced for a robotic system to empower smart cleaning. A robot manipulator [1] is utilized for trash collecting in grass fields based on deep neural networks SegNet and ResNet. However, the robot's speed was slow, and it only detected large garbage in the outdoor scenario. An ultrasonic-guided robot [7] is designed based on an IoT strategy for path planning of indoor cleaning. The proposed method detects the trash area and concentrates on path navigation and obstacle avoidance. Ramalingam *et al.* [5] proposed a debris detection approach using Deep Convolutional Neural Network (DCNN) to collect solid trash on the floor. The proposed method can detect two trash classes without determining the distance between the trash and robot positions. Salimi *et al.* [11] used Support Vector Machines (SVM) to classify trash features into organic waste, nonorganic waste, and non-waste for bin robots. The proposed method utilizes the Haar-Cascade approach to recognize any objects on the floor. Then, a Histogram of Oriented Gradient (HOG) [12] combined with Gray-Level Co-Occurrence Matrix (GLCM) [13] is devoted to extracting the trash features.

Furthermore, Bansal *et al.* [10] proposed a Convolutional Neural Network (CNN) using the back-propagation algorithm [9] to recognize the large-backed garbage for recycling by a robotic arm. Jost *et al.* [14] introduced a CNN for water streak detection for scrubber dryers. Meanwhile, Carolis *et al.* [15] proposed a real-time YOLOv3 object detection algorithm [22] to detect outdoor trash. The approach utilized video streams around cities to detect the abandoned garbage and illegal dumpsters. Similarly, Wang *et al.* [16] proposed a region-based CNN Faster RCNN. It concentrates less on a single piece of waste and focuses more on analyzing a big collection of outdoor urban garbage taken by single photo. The disadvantage of this method is that ease-to-meet overfitting problem during training progress. On the other hand, Golovinskiy *et al.* [17] introduced a scene-level deep network to segment outdoor waste objects by utilizing a few potential object-containing regions

Although most proposed cleaning robots could work in specific outdoor scenarios, they are usually utilized to pick up solid trash and do not consider the reflected wet, dirty areas, and the small size trash. Also, the mentioned approaches did not address obtaining the distance between the trash and the cleaning machine. Therefore, this paper proposes a vision learning-based approach that can clean the wet regions and small-size trash to deal with existing problems. The proposed method is limited to focus on cleaning indoor small trash, with the distance of trash being near two meters far from the robot.

The main contribution of this paper includes three main parts, as follows:

1) The proposed novel RGB-D vision system is utilized on a real cleaning machine that can efficiently detect indoor trash by a modified version of YOLOv3, called YOLOv3-trash, and estimate the position of the dirty area in the real world using the detected object's point cloud.
2) A trash dataset is collected with specific indoor types of trash, including tiny solid trash, dust region, and wet region. The collected dataset is challenging when considering some difficult-for-detection characteristics, such as to reflect wet trash, far small trash, or dust.
3) The experiment results on different YOLOv3-trash

This paper was submitted on February 25, 2022

Corresponding author: Chung-Hsien Kuo, he is currently a Professor with the Department of Mechanical Engineering at National Taiwan University, Roosevelt Road, Taipei, 106319 Taiwan (e-mail: chunghsien@ntu.edu.tw)

variants, named A, B and C, demonstrate that the proposed vision cleaning system could work efficiently and effectively in the open-space indoor environment.

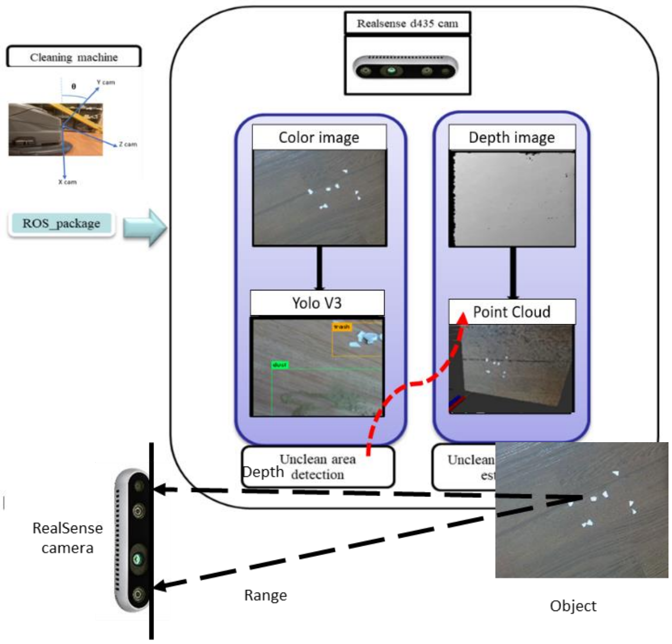


Fig. 2. RealSense depth camera setup.

II. Cleaning Robot System

A. Hardware Setup

The proposed autonomous cleaning robot vision system includes a RealSense-D435 Depth Camera (2×1 megapixel (OV9282) with a baseline of 50mm) attached in a KARCHER cleaning machine, as shown in Fig. 1.

B. Software Installation

RealSense Point Cloud ROS Node is used to calculate the distance between the autonomous cleaning machine and the dirty area based on the camera depth and bounding box of the detected trash. Each pixel of output images of the RealSense camera is represented with the RGB values and distance value, as shown in Fig. 2. Later, the vision system utilizes the Darknet YOLO Object Detection ROS Package to detect three different dirty classes, including dust, water, and small paper trash types.

III. Trash Detection System

A. YOLOv3 Object Detection

Redmon *et al.* [18], [20] has suggested a single-stage object detection network, named You Only Look Once (YOLO), that combines the classification stage and the region proposal network (RPN) in one network shortly and effectively, leading extremely fast computational speed and satisfactory efficiency as the ideal method for real-time object detection. Furthermore, YOLO models can explicitly predict objects by specifying the classes predicted through their bounding box and category. Among different variants of the YOLO-based model, YOLOv3 [22] is one of the most suitable models for real-time object detection. It utilizes three different scales to predict objects and the logistic regression used for the score of each bounding box analogous to feature pyramid network (FPN) [23]. The main architecture of YOLOv3 is Darknet-53. The Darknet-53 network combines multiple blocks that use two convolution layers in between with the same stride to reduce dimension. Every block consists of a bottleneck structure of 1×1 convolution layers accompanied by 3×3 layers with the skipping connections. Although Darknet-53 and ResNet [31] can possess billion floating-point operations (BFLOP), Darknet-53 is twice faster than RestNet with the same classification performance. Therefore, YOLOv3 can detect tiny objects with higher speed and accuracy. To obtain better features and representation for detected objects, Huang *et al.* [24] proposed a density–related architecture for YOLOv3, as illustrated in Fig. 3. It led to more compact and accurate models for determining the object.

B. YOLOv3-Trash Model

In the proposed vision system, a modified YOLOv3 model, known as YOLOv3-trash, is used to detect the tiny dirty and wet dirty in the floor environment, as shown in Fig. 3. In detail, the proposed YOLOv3-trash model changes the Residual Block in the Transition Layer from the original YOLOv3 to

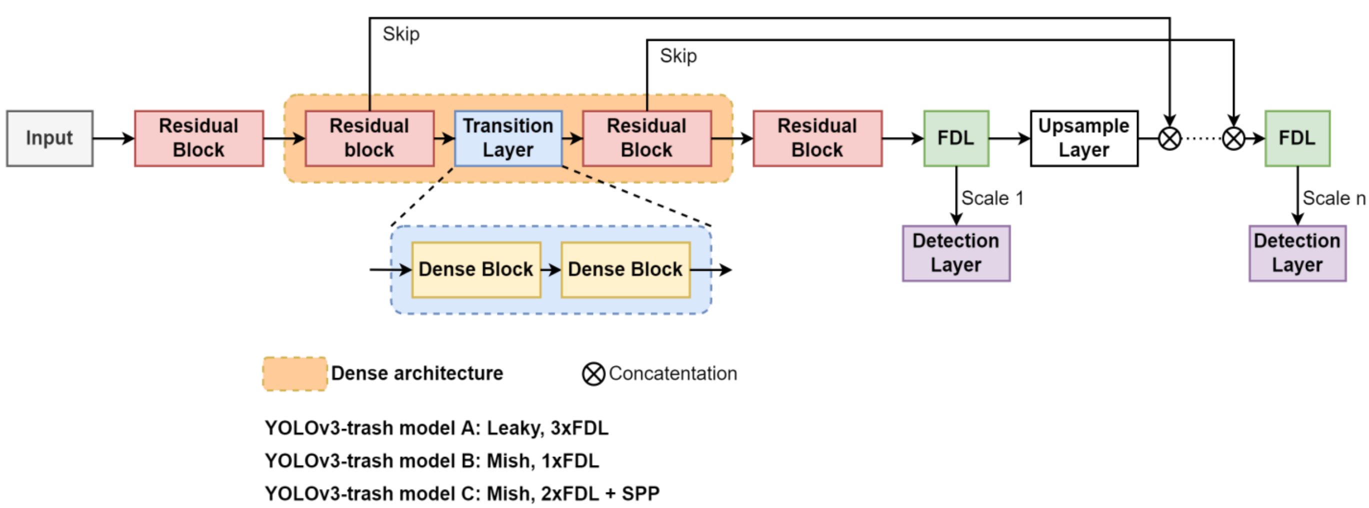


Fig. 3. The architecture of the YOLOv3-trash model with Dense architecture.

Dense Blocks [24], [30]. Moreover, a 3×3 Convolutional Layer and a 1×1 Bottleneck Layer are added in every Dense Block. Furthermore, to make the YOLO-trash model more accurate and compact, the Transitional Layer is placed between two Dense Blocks. The main objective of these modifications is to obtain detection on multiple feature maps through different layers of the network. Therefore, it will help obtain higher accuracy of tiny object detection from different environments. The YOLOv3-trash model also increases sizes of concatenated features from $26 \times 26 \times 768$ to $26 \times 26 \times 2816$ and from $13 \times 13 \times 384$ to $13 \times 13 \times 1408$ features in the FPN to preserve more fine-grained in detecting tiny trash. To observe the effectiveness of the proposed modification, YOLOv3-trash models has been divided into the YOLOv3-trash model A, B, and C. In-depth, Model A includes three FDL modules and uses the Leaky ReLU [25] as the activation function. Otherwise, model B adds only one FDL layer and uses Mish [26] activation function. Besides including two FDL modules, model C also adds an SPP module [27], [28] for multi-scale feature excavation, as described in Fig. 4, and Mish [26] activation function. Plus, the last Residual Block 4×1024 of the YOLO-trash model C is introduced before the SPP module, as shown in Fig. 4. When the convolutional layers go deeper, the receptive field of a single layer increases gradually, leading to improving the ability to extract the feature process of the YOLOv3-trash model. The Mish activation function also plays a key role in improving the performance of all deep neural networks when introducing a smooth non-linearity activation function:

$$\mathrm{Mish}(x) = \tanh(\ln(1 + e^x)) \tag{1}$$

where $\ln(1 + e^x)$ as the soft plus activation function [26].

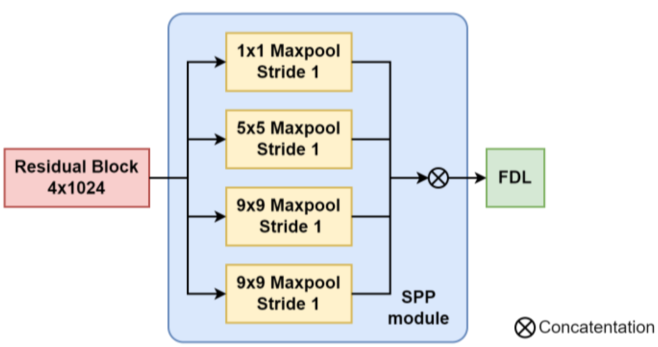


Fig. 4. SPP module in YOLOv3-trash model C.

C. *Real-World Coordinate Estimation*

The distance between the camera fixed on cleaning machine and the front trash is computed as following formula:

$$d = h \times \tan(\theta_1 + \theta_2) \tag{2}$$

where d as estimated distance of the trash from camera, h as the height of the camera on the machine, θ_1 as the half of camera view angle and θ_2 as the camera tile angle.

After extracting the boundary of detected trash objects in a 2D image by proposed YOLOv3-trash models, the real-world coordinate (x, y, z) of the center of the detected boundary is estimated by Point Cloud Library (PCL). Based on the image coordinate (u, v) of the color image, the real-world coordinates (x, y, z) can be estimated from:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = H^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \tag{3}$$

where $(x, y, z)^T$ as the estimated coordinate, $(u, v, 1)^T$ as the image coordinate in the homogeneous form, and H as the homographic matrix. The homographic matrix is automatically obtained and maintained by on-chip self-calibration method of Realsense API.

IV. Experimental Results And Evaluation

A. *Data Construction*

The main purpose of the proposed vision system is to deploy an autonomous cleaning machine that is reliable, efficient, energy-consumption saving during the cleaning process. In addition, the performance of the learning-based vision system depends on training dataset construction. Therefore, the trash objects are defined as trash in indoor scenarios, collected from the lab environment to give a general verification for the indoor floor scenario. The dataset consists of 3000 images of three categories of popular trash types in an indoor environment, including dust, water, small paper trash, as shown in Fig. 5. The proposed dataset also takes into account the tiny dust trash, water-drop area, and paper trash of different sizes.

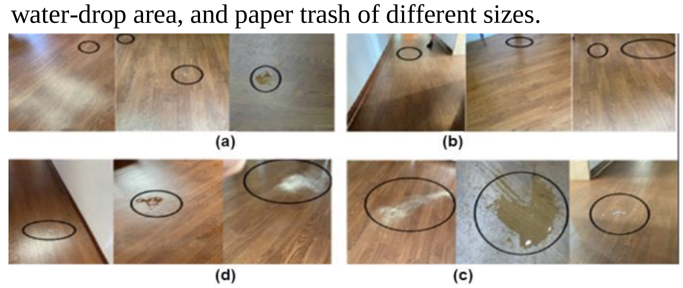


Fig. 5. Several examples of the collected dataset (a) Tiny trash, (b) Faraway trash, (c) Trash with light reflection, (d) Big trash.

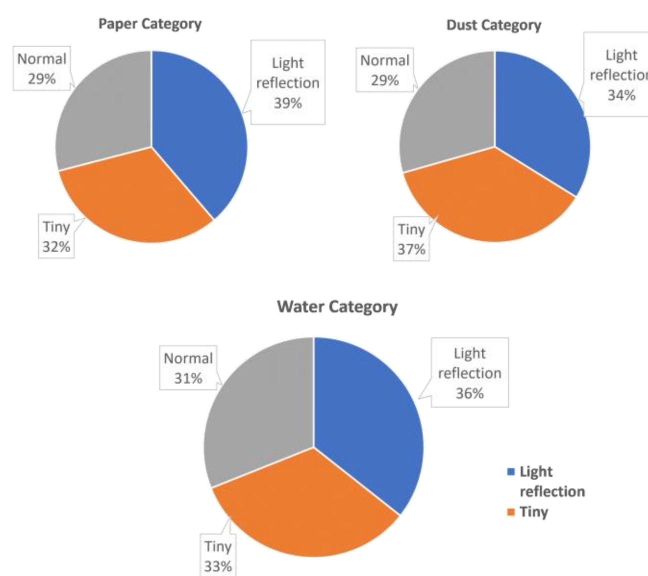


Fig. 6. The proportions of effected aspects in three dirty categories (paper trash, dust, and water).

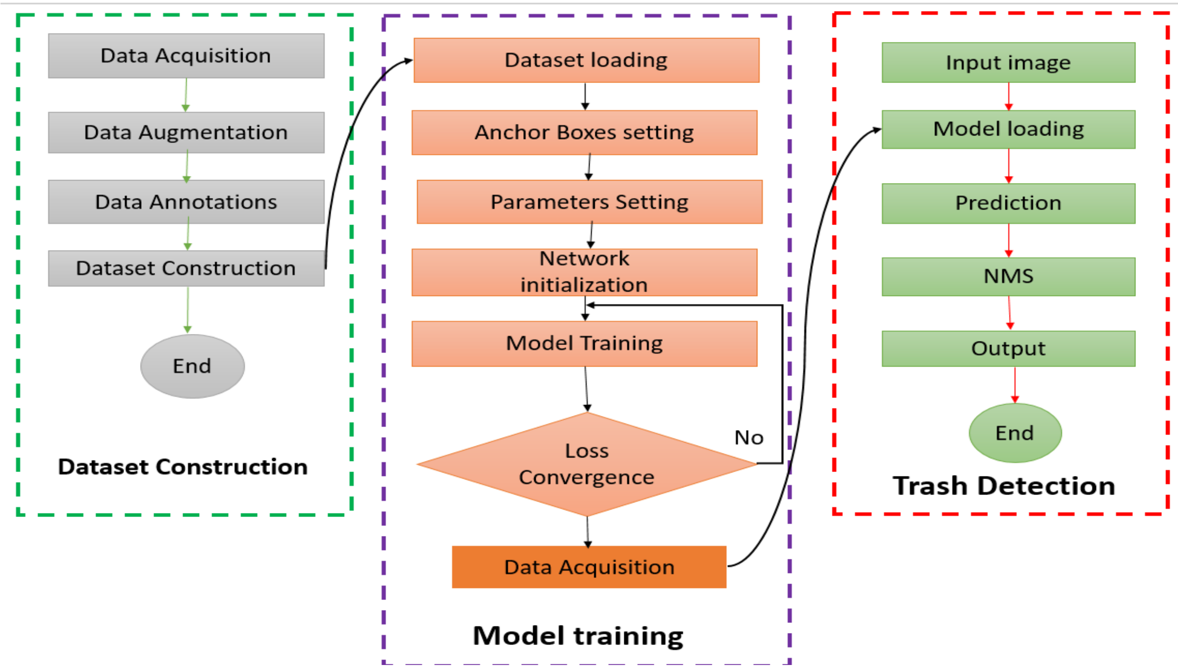


Fig. 7. The pipeline of dataset construction, training, and trash detection processes.

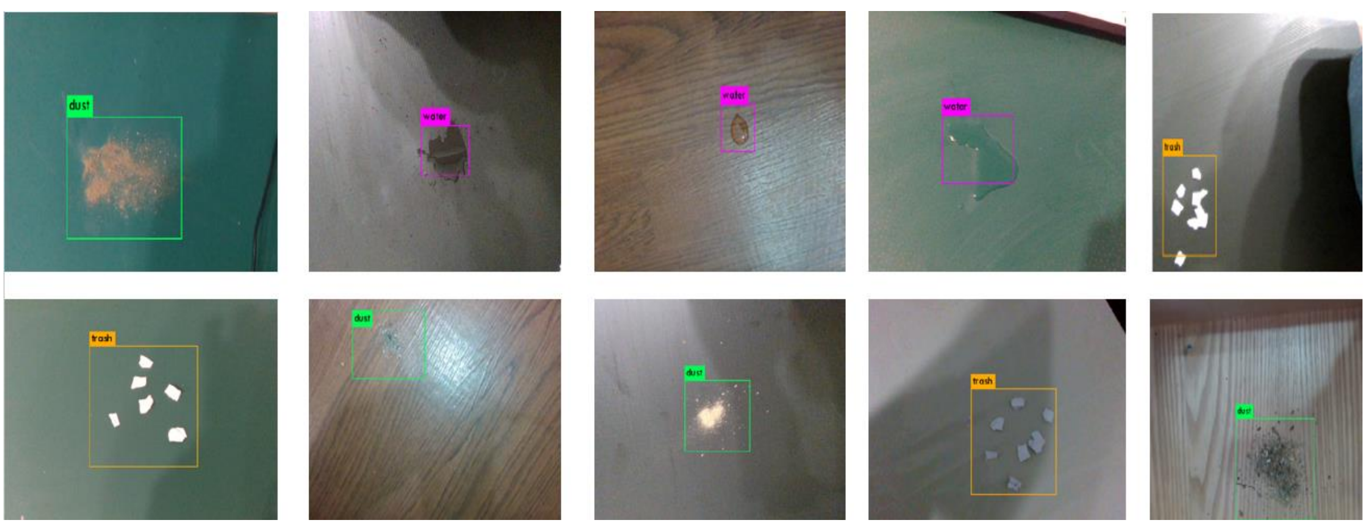


Fig. 8. The performance of YOLOv3-trash model C with different size, shape, and lighting reflection aspects of defined trash objects in practice.

In addition, the collected dataset also considers the distance of trash and the light reflection problem on wet dirty. As shown in Fig. 6, the collected dataset satisfies the amount and variant in each main category (trash, dust, water), helping to obtain high validation accuracy. In particular, the amounts of each main category are almost same (about 1000, 900, and 1000 images for paper trash, dust, and water, respectively), with proportions of affected aspects on each category, such as light reflection, tiny trash, are similar. The size of trash objects is various with 5cm in the minimum diameter and 50cm in the maximum diameter.

B. Experimental setup

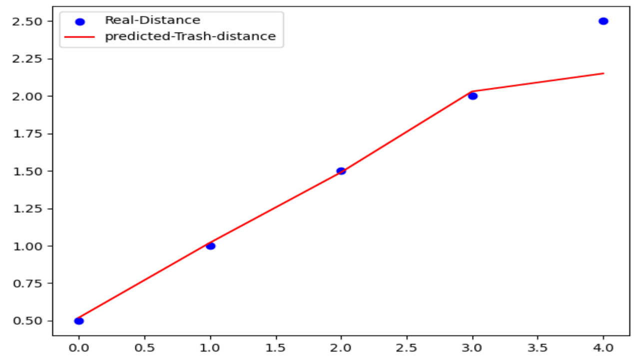
The proposed models were trained in the Google Colab, with CUDA Toolkit 10.0 and NVIDIA driver 384. The trained models then were deployed in Ubuntu 16.4 and ROS kinetic with python 2.7 and OpenCV library on the real cleaning machine. The detailed steps of the dataset construction, trash detection training, and trash detection processes of the YOLOv3-trash model are described in Fig. 7. The dataset is split into training and validation set with ratio of 80/20.

C. Experimental Results on YOLOv3-Trash Models

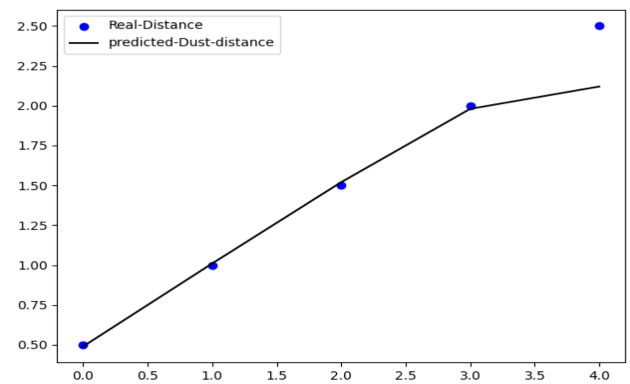
The proposed trash detection model is evaluated in RGB images with a resolution of 416×416 pixels. As shown in Table. 1, the plain YOLOv3 model achieves about 93% accuracy after 3000 iterations. In contrast, the modification in YOLOv3-trash model B and C by changing Leaky ReLU to Mish activation function can boost training and validation accuracy. Furthermore, Mish activation and SPP module utilization in YOLOv3-trash model C gives the best performance, with an average accuracy reaching 95.74%. The real-world performance of the proposed YOLOv3-trash models is also illustrated in Fig. 8. The visual results indicate that the proposed detector could detect well tiny, reflected trash.

TABLE I
ACCURACY COMPARISON BETWEEN YOLOV3, YOLOV3-TRASH MODELS A, B, AND C UNDER RESOLUTION OF 416×416

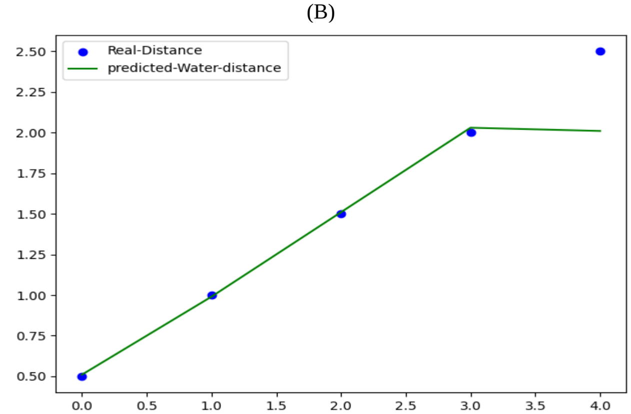
				Accuracy per class (%)		
Methods	Activation	FDL	SPP	Paper	Dust	Wet
YOLOv3 [22]	LeakyReLU	×3	no	93.05	93.08	92.09
Model A	LeakyReLU	×3	no	93.03	93.07	93.40
Model B	Mish	×1	no	95.06	94.05	95.08
Model C	Mish	×2	yes	**96.07**	**95.08**	**96.07**



(A)



(B)



(C)

Fig. 9. The average MAE values between actual and predicted distances for the indoor dirty classes, (A) - Paper, (B) - Dust, (C) - Water. The x- and y-axis presents horizontal and vertical distance (meter in unit)

D. *Trash Localization*

From the evaluation experiments, Table. II shows the maximum and minimum distance range for detecting collected trash categories. In general, the trash can be detected from 0.1m to 1.5m of range. The detecting distance range is not too far but is accepted because the dirty regions are small and not rigid.

Moreover, the average error of the actual distance and the estimated distance the camera and the detected trash can be obtained using the Mean Absolute Error (MAE):

$$\mathrm{MAE} = \frac{1}{n}\sum_{n}^{i}\left|y_i - Y_p\right| \qquad (2)$$

where y_i and y_p as predicted and actual distance value, n as the number observation.

The experimental real and predicted distances of different indoor trash types and the MAE values between them are shown in Table. III and Fig. 9. As shown in Table. III, the further distance of trash increases the larger MAE values are, but still achieving considerable error values in the range of 0.5m to 2m.

Similarly, Fig. 9. also illustrates MAE values between the actual and predicted distance for each trash class. The predicted distance for each class is close to the actual distance which improves the successfulness of the developed vision system. The experimental results have proved that the proposed vision-based detection system can work efficiently and effectively with under 2m in front of the cleaning machine. The issue may be due to the proposed vision-based trash detection system utilize only RGB image, which is blurred and ambiguous in faraway small trash. This issue may be solved when utilized the multi-sensor information, e.g., sonar, depth, lidar, to vision-based trash detection system.

TABLE II
THE MAXIMUM AND MINIMUM DISTANCE THAT CAN BE DETECTED FOR EACH TRASH CLASS

	Paper	Dust	Wet
Min distance (m) 0.15		0.11 0.10 Max distance (m)	
1.35		1.55 1.23	

TABLE III
THE AVERAGE MAE VALUE BETWEEN ACTUAL AND PREDICTED DISTANCE OF EACH TRASH CLASS

	Paper		Dust		Wet	
Actual (m)	Predicted (m)	MAE	Predicted (m)	MAE	Predicted (m)	MAE
0.5	0.52	**0.02**	0.49	**0.01**	0.51	**0.01**
1	1.02	**0.02**	1.01	**0.01**	0.99	**0.01**
1.5	1.49	**0.01**	1.52	**0.02**	1.51	**0.01**
2	2.03	**0.03**	1.98	**0.02**	2.03	**0.03**
2.5	2.15	**0.35**	2.12	**0.38**	2.01	**0.49**

*Actual means the actual distance of trash in front of the clean machine, Predicted means the estimated distance of the proposed method, and MAE means the MAE value between the actual and predicted distance.

V. CONCLUSION

This work introduces a novel real-time learning-based vision detection system for the auto-guided cleaning vehicle to clean the indoor trash and wet regions. The proposed vision system is created based on the advanced deep learning YOLOv3 algorithm for trash detection. A RealSense camera with the point cloud node in the ROS is utilized to calculate the distance between the detected trash object and the cleaning machine. After detecting the trash and its location, the auto-guided cleaning vehicle can go automatically to clean the trash area only that can save the cleaning time and decrease the consumed time compared with traditional cleaning machines. A challenging dataset is collected with difficult cases related to aspects that can affect the detector performance. The proposed vision system is tested in laboratory scenario to detect different kinds of indoor trash, including wet region, dust region and tiny paper trash. The experimental results emphasize the effectiveness of the proposed vision system to deal with different size of trash and the light reflection problem on the trash, with more time and power computational consumption. The reasonable classification results encourage to extend this research to detection more kind of trash in both indoor and outdoor scenarios. Furthermore, the system can be developed with more lightweight but powerful modern deep learning and be deployed various cleaning machines with ROS.

REFERENCES

[1] J. Bai, S. Lian, Z. Liu, K. Wang and D. Liu, "Deep Learning Based Robot for Automatically Picking Up Garbage on the Grass," in *IEEE Transactions on Consumer Electronics*, vol. 64, no. 3, pp. 382-389, Aug. 2018.

[2] H. Chae et al., "Façade Cleaning Robot With Manipulating and Sensing Devices Equipped on a Gondola," in *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 4, pp. 1719-1727, Aug. 2021.

[3] R. K. Megalingam, A. Prem, A. H. Nair, A. J. Pillai and B. S. Nair, "Staircase cleaning robot: Design considerations and a case study," *2016 International Conference on Communication and Signal Processing (ICCSP)*, 2016, pp. 0760-0764.

[4] A. A. Hayat, R. Parween, M. R. Elara, K. Parsuraman and P. S. Kandasamy, "Panthera: Design of a Reconfigurable Pavement Sweeping Robot," *2019 International Conference on Robotics and Automation (ICRA)*, 2019.

[5] B. Ramalingam, A. Lakshmanan, M. Ilyas, A. Le, and M. Elara, "Cascaded Machine-Learning Technique for Debris Classification in Floor-Cleaning Robot Application," *Applied Sciences*, vol. 8, no. 12, p. 2649, Dec. 2018.

[6] K. N. Baluprithviraj, M. M. Madhan, T. K. Devi, K. R. Bharathi, S. Chendhuran and P. Lokeshwaran, "Design and Development of Automatic Gardening System using IoT," *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, 2021, pp. 750-754.

[7] A. P. Murdan and P. K. Ramkissoon, "A smart autonomous floor cleaner with an Android-based controller," *2020 3rd International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ELECOM)*, 2020, pp. 235-239.

[8] S. Hossain, B. Debnath, A. Anika, M. Junaed-Al-Hossain, S. Biswas and C. Shahnaz, "Autonomous Trash Collector Based on Object Detection Using Deep Neural Network," *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, 2019, pp. 1406-1410.

[9] D. Rumelhart, G. Hinton and R. Williams, "Learning representations by back-propagating errors," *Nature*, 323, 533–536, 1986.

[10] S. Bansal, S. Patel, I. Shah, A. Patel, J. Makwana and R. Thakker, "AGDC: Automatic Garbage Detection and Collection," *ArXiv*, abs/1908.05849, 2019.

[11] Akanksha, M. Gupta and M. M. Tripathi, "Smart Robot for Collection and Segregation of Garbage," *2021 International Conference on Innovative Practices in Technology and Management (ICIPTM)*, 2021, pp. 169-173.

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 886-893 vol. 1.

[13] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features for Image Classification," in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973.

[14] J. Uriel and R. Bormann, "Water Streak Detection with Convolutional Neural Networks for Scrubber Dryers," *ICVS*, 2019.

[15] B. D. Carolis, F. Ladogana and N. Macchiarulo, "YOLO TrashNet: Garbage Detection in Video Streams," *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, 2020, pp. 1-7.

[16] Y. Wang, and X. Zhang, "Autonomous garbage detection for intelligent urban management," in Proceedings of the 2018 2nd International Conference on Electronic Information Technology and Computer Engineering (EITCE 2018), Shanghai, China, 12 October 2018.

[17] A. Golovinskiy and T. Funkhouser, "Min-cut based segmentation of point clouds," *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, 2009, pp. 39-46.

[18] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788.

[19] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017.

[20] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6517-6525.

[21] G. Liu, J. C. Nouaze, P. L. Touko Mbouembe, and J. H. Kim, "YOLO-Tomato: A Robust Algorithm for Tomato Detection Based on YOLOv3," *Sensors*, vol. 20, no. 7, p. 2145, Apr. 2020.

[22] J. Redmon and A. Farhadim, "YOLOv3: An Incremental Improvement," *ArXiv*, abs/1804.02767, 2018.

[23] T. -Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 936-944.

[24] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261-2269.

[25] V. Nair and G. E. Hilton, "Rectified linear units improve restricted boltzmann machines," in *International Conference on Machine Learning (ICML)*, 807−814, 2010.

[26] D Mia, "Mish: A self-regularized nonmonotonic neural activation function,", *2021 The British Machine Vision Conference (BCMVC)*, 2020.

[27] K. He, X. Zhang, S. Ren and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904-1916, 1 Sept. 2015.

[28] Y.-Q. Huang, J.-C. Zheng, S.-D. Sun, C.-F. Yang, and J. Liu, "Optimized YOLOv3 Algorithm and Its Application in Traffic Flow Detections," *Applied Sciences*, vol. 10, no. 9, p. 3079, Apr. 2020.

[29] M. O. Lawal, "Tomato detection based on modified YOLOv3 framework," *Scientific Reports*, 11, 2021.

[30] L. Donati, T. Fontanini, F. Tagliaferri, and A. Prati, "An Energy Saving Road Sweeper Using Deep Vision for Garbage Detection," *Applied Sciences*, vol. 10, no. 22, p. 8146, Nov. 2020.

[31] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.

Shimaa-Amin Bergies received the M.S. degree in Department of Electrical Engineering from National Taiwan University of Science and Technology, Taiwan, in 2022. Her current research interests and publications are in the areas of Autonomous Driving, Deep Learning.

Phuc Thanh-Thien Nguyen received his received his MS. Degree in Computer Science, National Taiwan University of Science and Technology, Taiwan in 2020. He is currently Ph.D. student in Department of Electrical Engineering, National Taiwan University of Science and Technology. His research interest is machine learning, deep-learning, and computer vision.

Chung-Hsien Kuo received his BS. Degree in Mechanical Engineering from National Chung Hsing University in 1990 and received the master and Ph.D. degrees in Mechanical Engineering from National Taiwan University, in 1995 and 1999, respectively. Since August 2021, he has been with the National Taiwan University, where he is currently a Professor in the Department of Mechanical Engineering. His research interests include autonomous systems, novel soft robot design, advanced control systems, hybrid robot design combining rigid and soft machines, computer vision and visual servoing, and autonomous manipulation with AI approaches.