ICIC Express Letters
Part B: Applications
Volume **14**, Number **6**, June **2023**

ICIC International ©2023 ISSN 2185-2766
pp. 597–604

A STUDY ON MICROORGANISMS DETECTION FROM MICROSCOPIC IMAGES USING DEEP LEARNING

Hiroshi Shiratsuchi[1], Misato Furuta[2], Kenta Furukado[2], Kozo Horiuchi[3] and Takanori Matsuzaki[1]

[1]Graduate School of Humanity-Oriented Science and Engineering
Kindai University
11-6 Kayanomori, Iizuka-shi, Fukuoka 820-8555, Japan
{ sira; takanori }@fuk.kindai.ac.jp

[2]Cowsy Co., Ltd.
488-1 Kayanomori, Iizuka-shi, Fukuoka 820-0011, Japan
envi@cowsy.jp

[3]Gururi Co., Ltd.
2-1-6 Uomachi, Kokurakita-ku, Fukuoka 802-0006, Japan
kozo.horiuchi@gururi.co.jp

Received November 2022; accepted January 2023

Abstract. *This paper aims at the microorganisms image detection system that uses machine learning to estimate the status of aerobic microorganisms in activated sludge, which is necessary for the stable management and operation of a purification facility for factory wastewater. In other words, if the proposed system supports the detection of the type and number of microorganisms in the aeration tank, the condition of the purification facility is easy to estimate. YOLOv4, one of the most accurate single-step object detectors, has been used to detect microbial objects from microscopic images. In this case, the input image size is resized based on annotation information. As a result, a good result was obtained, which means Average Precision, an index of detection accuracy, was about 73.12%. Then, the determination of the anchor size is also calculated based on the bounding box information for a more appropriate rectangular area prediction.*
Keywords: Microorganisms, Wastewater treatment facility, Deep learning, Object detection

1. **Introduction.** There are regulation values for the quality of water discharged from factories defined by the Ministry of the Environment in Japan [1]. For this reason, factories purify the water sufficient for the effluent standards before releasing it into the river. Such wastewater treatment facilities use aerobic microorganisms called activated sludge to decompose and treat the pollutants in the wastewater [2]. A high level of expertise and experience is required for the stable operation of facilities. However, the problem is that the time and cost necessary to inherit these are limited.

For the operation and maintenance of the current wastewater treatment facilities, technical employees visit the installation site of the treatment facilities once a week to once a month to conduct the below investigations and adjustments. First, each facility is fine-tuned based on visual confirmation and data from water quality and various sensor devices (pH, dissolved oxygen, oxidation-reduction potential, activated sludge suspended solids and others). The validity of these operations is confirmed by analyzing microscopic observation of the microorganisms from bringing back the activated sludge. If the on-site response is inadequate, simple manipulation, such as valve adjustment, is requested from the local personnel who do not have expertise in the facility as a remote response. At the

DOI: 10.24507/icicelb.14.06.597

next visitation of technical employees, appropriate adjustments operate. Thus, analyzing microorganisms conditions on-site allows technicians to respond quickly and accurately.

As mentioned above, the types and numbers of microorganisms appearing in activated sludge are correlated with the condition of the wastewater treatment facility. However, the estimation accuracy is not stable due to differences in the microorganisms identification ability of technical employees in actuality. For this reason, several methods have been proposed to automatically detect target microorganisms. These methods can be broadly classified into three categories [4]: threshold based segmentation, region based segmentation, and edge based segmentation, as specific examples of the third segmentation method, a phase matching of microscope images [6] and classification of the processed images by SVM (Support Vector Machine) [7]. However, these methods have not been sufficiently accurate for the microorganisms detection. In machine learning, the deep learning model trains using input and teacher pairs called datasets. In other words, select the best deep learning model to extract the trends and features of the target problem, and this model trains by the training data of the datasets. The model obtained in the above process can recognize, identify, classify, and predict. The authors proposed a method to identify microorganisms from microscopic images using googLeNet, one of the Convolutional Neural Network (CNN) models. Here, CNN is a well-known deep learning architecture inspired by the natural visual recognition mechanisms of living organisms [3]. In our proposed method, twenty kinds of microorganism names could be recognized with recognition rate of about 87%, though the region in which the microorganism exists had to be extracted from the microscope image beforehand [8].

This paper aims to develop a microorganisms image detection system that supports the inheritance of expertise in the operation and maintenance of wastewater treatment facilities. In other words, a method for detecting twenty kinds of microorganisms targeted in the previous research without prior cutting work is studied. Specifically, YOLO, one of the one-stage object detection algorithms, is used. The extracted image region of appropriate size that contains the target microorganism is used as the training dataset. Furthermore, for the aspect ratio, which is an internal parameter that is important when detecting the position of YOLO, a value calculated by the k-means method based on the bounding box information of the training dataset is adopted instead of the default value. As described above, the proposed system will expect to use effectively as technical employees' training and support software. As followed, Section 2 represents an overview of the wastewater treatment facility. Section 3 describes the outline of object detection. For detecting microorganisms by YOLOv4, these procedures and the preparations discuss actual learning results. Section 5 summarizes the paper.

2. Overview of Wastewater Treatment Facilities.

Factory wastewater varies greatly depending on the type of operation. This section will use a food manufacturing plant as an example. The wastewater in these factories does not contain toxic substances, but it often contains substances that may cause environmental pollution [9]. Figure 1 shows the diagram of the wastewater treatment facility based on the standard activated sludge method treating wastewater from a food manufacturing plant. In the figure, the aeration tank uses an aerobic microbial treatment system called activated sludge. In addition, due to the nature of food production, the handling of raw materials, processed products manufactured, and production volume change dynamically from season to season, so the quantity and quality of water in factory effluent vary from day to day. Therefore, the operation and maintenance of the wastewater treatment system have a property of the frequent need for fine adjustments [2].

The operation of wastewater treatment facilities in both aeration and sedimentation tanks is critical to pollutant treatment performance. The status of decomposition and treatment of pollutants in factory wastewater depends on the presence of microorganisms

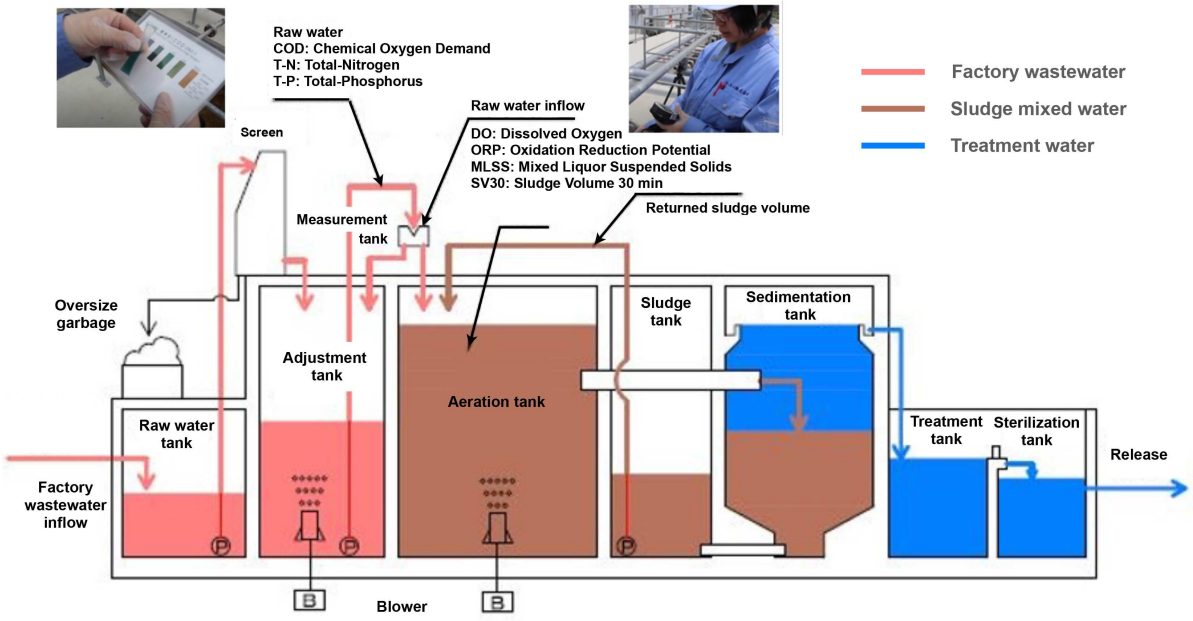


FIGURE 1. (color online) Overview of wastewater treatment facilities

in the treatment equipment. There is a relationship between the condition of the processing equipment and the type and quantity of microorganisms present, which is well known. Therefore, these indicators are significant factors in determining the operation and maintenance operations of the wastewater treatment facility. However, it is difficult to observe them at the site due to the equipment, and the activated sludge is usually taken back and observed under an electron microscope. The time required to determine the water quality of the aeration tank may range from 15-30 minutes for skilled technicians to more than 2 hours for an inexperienced technician. Furthermore, even among them, differences in identification ability can lead to differences in the inferred results of the state of the processing equipment.

The activated sludge was collected from 38 wastewater treatment facilities in food manufacturing plants. A total of 5,736 images were extracted from collected samples by the microscope. These image data contain a wide variety of microorganisms. Twenty microorganisms shown in Table 1 were selected for the necessary to estimate water quality. In the table, microorganisms such as Arcella and Centropyxis appear in large numbers when nitrification in the tank progresses and the pH of the treated water decreases. In addition, Lecane and Lepadella have the characteristic of appearing when the influent water concentration is low and the load is extremely low. Moreover, Epistylis and Euglypha appear in large quantities when the facility is under low load and sludge demolition is in progress. Other microorganisms also vary in the types and populations that appear depending on the load of the aeration tank or wastewater treatment facility. Therefore, if determining the number of each microorganism per unit area, the condition of the wastewater treatment facilities will be possible to estimate.

3. **Object Detection by Deep Learning.** In our previous research, CNN was used as a classification problem to discriminate the images cutting out an including region of the microorganisms. This section describes object detection, which estimates the location and class of microorganisms. Table 2 shows the difference between the classification problem and object detection. This table shows that object detection is effective for microorganisms scattered in microscopic images.

TABLE 1. List of target microorganisms

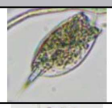
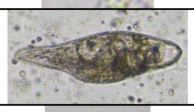
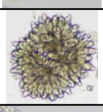
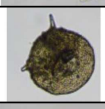
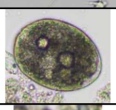
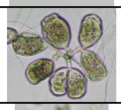
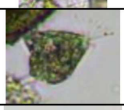
No.	Name	Image	No.	Name	Image
1	Arcella		11	Lecane	
2	Aspidisca		12	Lepadella	
3	Blepharisma		13	Opercularia	
4	Carchesium		14	Paramecium	
5	Centropyxis		15	Peranema	
6	Chaetonotus		16	Philodinidae	
7	Chaetospira		17	Prorodon	
8	Epistylis		18	Pyxidicula	
9	Euglypha		19	Tokophrya	
10	Euplotes		20	Vorticella	

TABLE 2. Difference between classification and object detection

Property in an image	**Object detection**	**Classification**
Number of targets	plurality	one
Size of objects in an image	small to large	large
Position estimation	yes	no

There are three implementation methods for object detection: (i) sliding window detection, (ii) two-stage detection, and (iii) one-stage detection. First, using a sliding window, an image is cut out by various preset window sizes and slide widths. Each image is detected when the CNN classified output exceeds a threshold. In this method, the window size and the slide width must be set appropriately, and there is a problem that an enormous amount of calculation is required [10]. Second, the two-step detection includes (a) extraction of candidate region proposals and (b) processing such as determination of candidate regions and rectangle correction. In R-CNN (Region-based Convolutional Neural Network), which is known as above representative algorithm, it is determined for a candidate region obtained by combining similar pixels called "selective search". Then if there is an overlapping rectangular region, a rectangular adjustment called non-max suppression is performed [11]. Third, one of the features of one-stage detection is that it operates at high speed because it performs area detection and identification in parallel. To achieve one-stage object detection, YOLO divides the image into S×S grids and computes rectangle detection and class prediction for each grid in a coordinated manner [12]. In object detection, the accuracy of the predicted rectangular area is evaluated by the overlap rate

as IoU (Intersection over Union), and the overall prediction is evaluated by an index based on correct/false information called AP (Average Precision). Here, IoU is the ratio of overlap area and union area, and AP evaluates each rectangle by IoU as an integral value of the precision-recall curve.

In the following, microorganism detection in microscopic images is discussed by using YOLO. YOLO(v1) [12] was one of the first models to realize object detection in one step, which previously consisted of two-stage: detection and identification. YOLO divides the input image into S×S (S = 7 in the paper) grid cells and learns bounding box estimation and object prediction in a coordination. In the bounding box estimation, B rectangular regions are prepared in advance, and the possibility that an object exists in the rectangular of the grid is predicted as a box confidence score. On the other hand, in the object prediction, a conditional class probability P is estimated in which each grid cell is determined to be C classes to be determined. The bounding box region is estimated from the scores of the obtained conditional class probability P and the box confidence score. This region is fine-tuned by a technique called NMS (Non-Maximum Suppression) using the IoU value. The model consists of 24 CNN layers, four pooling layers for feature extraction, and two fully coupled layers for predicting box confidence scores and conditional class probabilities. Compared to Faster R-CNN [13], an improved version of R-CNN [11] released around the same time, YOLOv1's advantages are its simple structure and fast detection. On the other hand, its detection accuracy is not as good as two-step object detection, and it has the disadvantage that it is not good at detecting small objects because it uses grid cells. Therefore, an improved model YOLOv2 with an anchor box, batch normalization, and higher resolution was proposed. YOLOv3 was proposed to improve the accuracy of recall and localization. In other words, anchor boxes are calculated based on the k-means method from MS COCO [15], one of the large datasets, and predicted using three-sized scale images created by Feature Pyramid Networks (FPNs). This model consists of 53 layers of CNN, Residual Net, and FPN structure.

YOLOv4 classifies the components of an object detector as follows: backbone (feature extraction of images), head (calculation of features from the backbone feature map), neck (class classification and location estimation), freebie (accuracy improvement method that does not increase inference cost), and a special (significantly increases detection accuracy with some increase in inference cost) [16]. The architecture with the highest performance is selected through a round-robin trial of various methods developed for each element.

4. Microorganism Detection in Microscopic Images by YOLOv4.

As described in the previous section, YOLOv4 is composed of the optimal combination. The YOLOv4 system can also be configured in detail. For example, many augmentations such as blur, mosaic, partial disappearance, scaling, and rotation are also implemented. Each time an input image is presented as training data, the processing is performed based on specified probability and range. On the other hand, since object detection is based on area estimation from anchor boxes arranged in a grid, objects of small size relative to the input image size are hard to detect.

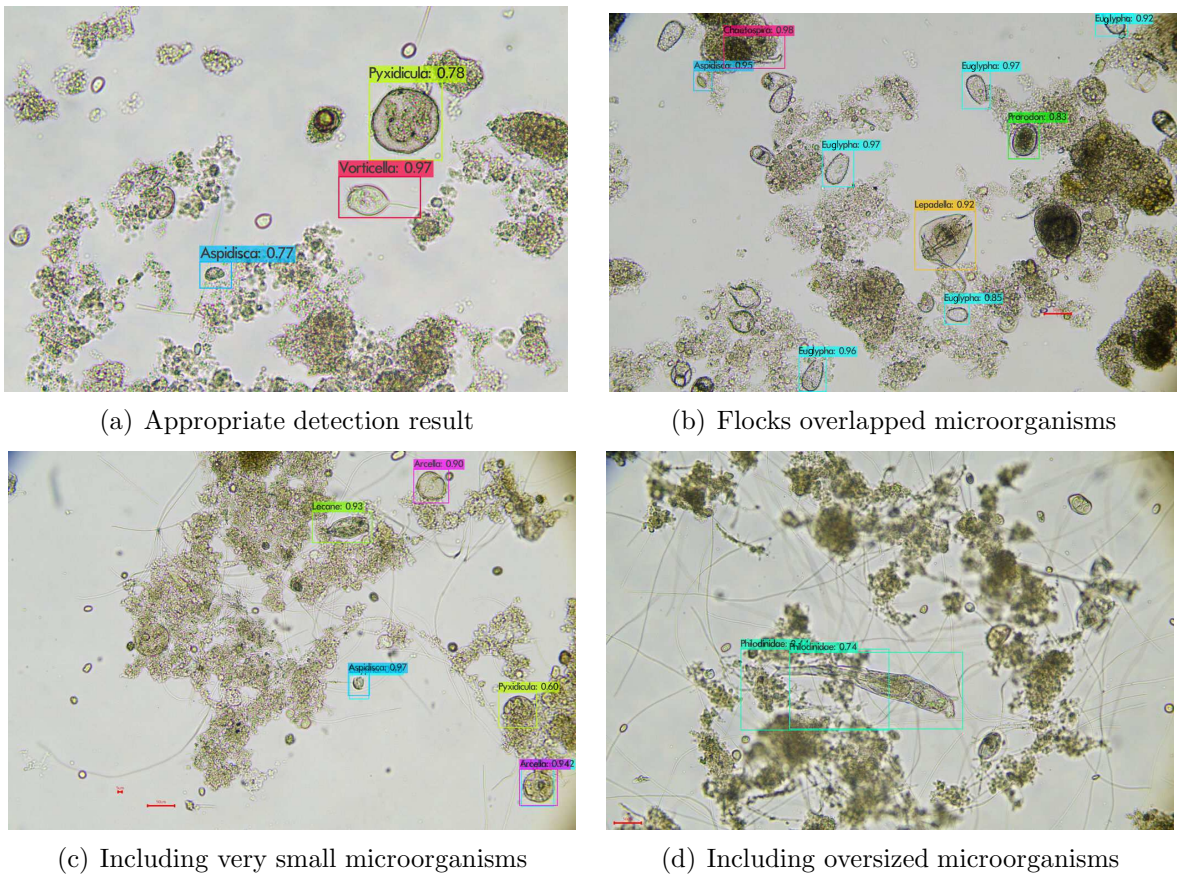
The image size of the training dataset presented in the YOLOv4 learning is generally 640 × 640 pixels. However, the image size produced as a microscope image is large (3000 × 2000). In addition, the image size of the target microorganisms varies from 30 × 30 to as large as 1024 × 1024. Image size reduction by a simple affine transformation may result in loss of image information and reduced detection accuracy, especially for microorganisms of small size. Therefore, in this section, the extraction of training data based on annotation information is examined to optimize the input image of YOLO, i.e., a method for cutting the microorganism image size into 640 × 640 to 1024 × 1024 is studied. The size of the anchor box is also set to a value based on the annotation information of the microorganism to improve detection accuracy.

TABLE 3. Calculation result of Average Precision (AP)

Name	AP [%]	TP	FP
Arcella	64.87	76	93
Aspidisca	75.63	326	276
Blepharisma	76.92	10	1
Carchesium	63.20	29	10
Centropyxis	82.12	184	215
Chaetonotus	76.61	49	8
Chaetospira	70.87	111	42
Epistylis	60.11	92	95
Euglypha	80.50	119	393
Euplotes	78.07	36	7
Lecane	82.09	197	125
Lepadella	78.12	130	30
Opercularia	35.78	12	29
Paramecium	73.49	54	8
Peranema	90.74	109	25
Philodinidae	80.89	148	71
Prorodon	78.12	118	123
Pyxidicula	67.15	97	113
Tokophrya	78.82	23	13
Vorticella	68.26	140	34

The computational simulations were carried out using the model of YOLOv4-P6, one of the YOLOv4-large implementations [17]. There are 1804 training data and 782 images for verification. Table 3 shows the number of AP (Average Precision), TP (True Positive), and FP (False Positive) for each target microorganism. From the table, AP for each microorganism ranged from 60.11% to 90.74%, and mAP (mean AP) was 73.12%. The increase in FP is due to a plurality of rectangular regions detected for one object. When the proposed data augmentation was not performed, the mAP of overall and AP of each microorganism under similar experimental conditions ranged from 27.56% to 70.03% and 43.44%, respectively. Figure 2 shows an example of detection when 3000 $\times$ 2000 microscope images are presented as test data to the model after learning. In the figure, the microorganism area denotes by a frame, and the microorganism name and score are shown in the upper part. The images of the detection results show that good microorganism detection is performed. However, there are some cases where the target microorganisms are present but not detected in each image. For example, in Figure 2(b), there is some undetected Euglypha. These are not detected properly because two or more microorganisms overlap or create overlapping areas with clumps of microorganisms called flocks. Also, very small Pyxidicula and Aspidisca are not detected shown in Figure 2(c). In addition, too large microorganisms such as Philodinidae in Figure 2(d) are counted as multiple microorganisms. One possible reason for the above problems is that the default anchor size values used in YOLOv4 candidate region prediction are calculated from the MS COCO data set. Simulation results when the anchor size for the YOLO layer is changed to a value calculated from the training dataset will be provided at the time of publication, as computer simulations are currently underway.

5. **Conclusion.** This paper proposes a microorganisms image detection system that supports the inheritance of expertise in the operation and maintenance of wastewater treatment facilities. The proposed microorganisms detection system was able to identify 20



(a) Appropriate detection result

(b) Flocks overlapped microorganisms

(c) Including very small microorganisms

(d) Including oversized microorganisms

FIGURE 2. Result of microorganisms detection by YOLOv4

microorganisms necessary for the determination of water quality with high accuracy. Microorganisms detection is possible regardless of the skill level by using our system. In addition, although not presented in this paper, this system can detect microorganisms in the video as well as still images. Therefore, this allows on-site microbiological testing without bringing activated sludge back to the site. However, since the detection process requires a large GPU load, actions such as transferring video to the server are required.

Flocks removal methods were considered by a chemical such as Gram stain and alcohol cleaning, and digital noise reduction for microscopic images. The problem with these methods is that the microorganisms are removed along with the flocks. However, the flock was successfully disassembled by the method using the crushing vibration device, and we plan to collect training data after adjusting the operating conditions. These methods will be a subject for future study.

REFERENCES

[1] *General Wastewater Standards, Ministry of the Environment*, https://www.env.go.jp/water/impure/haisui.html, Accessed in Nov. 2022 (in Japanese).

[2] *Industrial Wastewater Management in Japan*, https://www.env.go.jp/en/focus/docs/files/20120801-51.pdf, Accessed in Nov. 2022.

[3] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, Gradient-based learning applied to document recognition, *Proc. of IEEE*, vol.86, no.11, pp.2278-2324, 1998.

[4] F. Kulwa et al., A state-of-the-art survey for microorganism image segmentation methods and future potential, *IEEE Access*, vol.7, pp.100243-100269, 2019.

[5] F. Zhou and J. Liu, Microbial contour extraction based on edge detection, *2016 8th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp.1-5, 2016.

[6] M. B. Khan, H. Nisar and C. A. Ng, Image processing and analysis of phase-contrast microscopic images of activated sludge to monitor the wastewater treatment plants, *IEEE Access*, vol.6, pp.1778-1791, 2018.

[7] S. Selvan, J. F. Joseph, K. T. D. Raj and B. J. Selvan, Analysis and classification of active sludge in wastewater, *Journal of Physics: Conference Series*, vol.1916, no.1, 012016, 2021.
[8] H. Shiratsuchi, M. Furuta, Y. Wakamori, K. Horiuchi and T. Matsuzaki, A study on the microorganisms image recognition assist system based on deep learning, *ICIC Express Letters*, vol.16, no.6, pp.621-628, 2022.
[9] C. Rachid, *Chemicals in the Food Industry: Toxicological Concerns and Safe Use*, Springer International, 2020.
[10] C. Farabet, C. Couprie, L. Najman and Y. LeCun, Learning hierarchical features for scene labeling, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.35, no.8, pp.1915-1929, 2013.
[11] R. Girshick, J. Donahue, T. Darrell and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, *arXiv.org*, arXiv: 1311.2524, 2013.
[12] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, You only look once: Unified, real-time object detection, *arXiv.org*, arXiv: 1506.02640, 2015.
[13] S. Ren, K. He, R. Girshick and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *arXiv.org*, arXiv: 1506.01497, 2015.
[14] J. Redmon and A. Farhadi, YOLOv3: An incremental improvement, *arXiv.org*, arXiv: 1804.02767, 2018.
[15] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick and P. Dollár, Microsoft COCO: Common objects in context, *arXiv.org*, arXiv: 1405.0312, 2014.
[16] A. Bochkovskiy, C. Wang and H. M. Liao, YOLOv4: Optimal speed and accuracy of object detection, *arXiv.org*, arXiv: 2004.10934, 2020.
[17] C. Wang, A. Bochkovskiy and H. M. Liao, Scaled-YOLOv4: Scaling cross stage partial network, *arXiv.org*, arXiv: 2011.08036, 2020.