Computer Engineering and Applications Vol. 12, No. 2, February 2023

Video Based Fish Species Detection Using Faster Regional Convolution Neural Network

*Muhammad Naufal Rachmatullah, Akhiar Wista Arum

Departement of Informatics, Faculty of Computer Science, Universitas Sriwijaya
Intelligent System Research Group, Sriwijaya University, Indonesia

**naufalrachmatullah@unsri.ac.id*

ABSTRACT

Fish recognition and classification represent significant challenges in marine biology and agriculture, promising fields for advancing research. Despite advancements in real-time data collection, underwater fish recognition and classification still require improvement due to challenges such as variations in fish size and shape, image quality issues, and environmental changes. Feature learning approaches, particularly utilizing convolutional neural networks (CNNs), have shown promise in addressing these challenges. This study focuses on video-based fish species classification, employing a feature learning-based extraction method through CNNs. The process involves two main stages: detection and classification. To address the detection and classification in video a Faster Region Convolutional Neural Network (RCNN) with transfer learning techniques are applied, achieving a mean average precision of 84% for detection and classification tasks. These techniques offer promising avenues for enhancing fish recognition and classification in diverse environments.

Keywords: Faster RCNN, Convolutional Neural Network, Fish Detection, transfer learning.

1. INTRODUCTION

Image processing and analysis techniques for underwater cameras have attracted rapid attention. This is because the use of underwater cameras is the safest observation method in fisheries research [1]. For example, the Cam-Trawl method, a combination of multiple cameras used in fish farms, has been successfully applied to capture images and videos of fish [2]. The camera sampling approach has not only been successful in assisting oceanographers in fish conservation but also provides an effective approach to sampling data from the vast diversity of marine animals, including fish [3]. However, this approach results in the availability of an abundance of data. Therefore, image processing for automatic fish identification is necessary to enable more in-depth analysis.

The use of image processing techniques in fish classification has its own constraints. This constraint is caused by the special characteristics of underwater video. According to Qin et al. (2016)[4], underwater video is of poor quality due to the extreme conditions in the open ocean. The imaging devices used are designed to record low-resolution video with the aim of capturing as much data as possible. In addition, the lack of lighting causes limited vision and blurred shapes of objects.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Furthermore, fish move freely in the ocean causing the movement of the fish in the video to vary greatly.

In the research domain of video-based fish species classification, there are two important processes. The first process is to perform object detection in the video. This process extracts the object to be analyzed, namely the fish. Furthermore, the second process is to perform the fish species classification process based on the object detection results in the first process [5]. The resulting features must be discriminative so as to provide good accuracy.

Feature extraction with a learning approach is an approach that allows the system to learn informative features directly from the data (image) [6]. The advantage of the feature learning-based feature extraction method is that it does not require feature specifications or prior knowledge, allowing this method to produce more complex features. One of the feature learning methods that performs quite well in performing classification tasks with large data is deep learning (DL).

Deep learning has found success across numerous research fields, yielding satisfactory outcomes [7], [8]. Within the realm of image analysis and computer vision, one of the prominent architectures is the convolutional neural network (CNN). CNNs offer the advantage of directly learning features from input images. These features are encapsulated in a feature map (fmap), preserving significant and distinctive data information [9]. Moreover, CNN architecture requires minimal preprocessing, further enhancing its appeal.

This research will try to address the problem of video-based fish species classification. The proposed approach to overcome these problems is to apply a feature learning approach at the feature extraction stage using deep learning methods. The deep learning architecture used is convolutional neural network. It is expected that later with this research the results of the feature extraction given can describe data features that have a high level of discrimination so as to increase the accuracy of the classification process.

2. MATERIAL AND METHODS

2.1 DATASET

The dataset used in this study was taken from the study [10]. The dataset consists of 77 videos. In addition to videos, datasets in the form of images are also used in this study. The image dataset is used as training data to create a model for classifying fish species. The fish image dataset is taken from the same source. The images used totaled 22,443 images spread across 15 species. The fish images used are RGB with sizes varying from 22 x 35 pixels to 408 x 171 pixels. Examples of fish images used are shown in Figure 1. Information about the species and distribution of the image data used is shown in Figure 2.

Observations of the data were made to see the characteristics and challenges faced when conducting research. Based on the observations from the video and image datasets, some of the problems related to the fish detection process encountered are:

- **Lighting changes.** During the recording process, the captured light changes gradually (due to changes in sunlight intensity) or suddenly (due to camera lighting going out).

- **Changing video background**. The changing background of the video complicates the detection process. Therefore, the algorithm must be able to distinguish between moving objects that are detected as fish and objects that are classified as background.
- **Occlusion**. The object under study cannot be fully detected in some frames because the object is covered by other objects.
- **Clutter**. The irregular position of the object inhibits the background modeling process and complicates the segmentation process.
- **Camouflage**. The target to be detected has colors and patterns that are almost similar to the background, making the object difficult to detect even by humans.
- **Noise**. The recorded video has a lot of noise that affects the detection results.
- **Environmental Issues**. The quality of video footage is also affected by water quality, such as debris, the presence of marine plants, the amount of garbage under the sea, etc.

FIGURE 1. Example of Dataset.

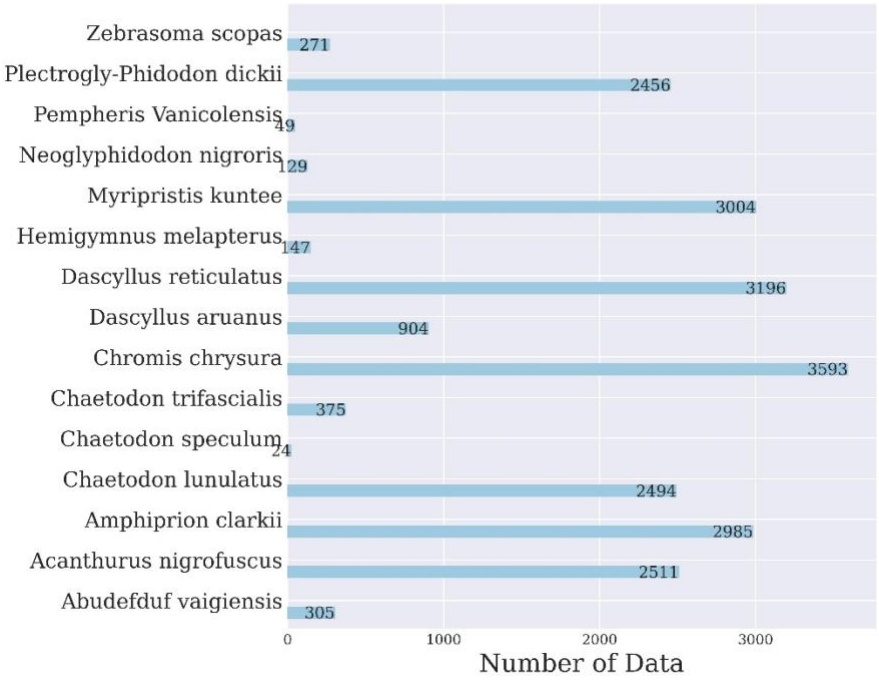

FIGURE 2. Fish species dataset Distribution

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

2.2 TRANSFER LEARNING

Transfer learning is a machine learning method in which a model that has been developed for a domain is reused as a starting point for modeling a different domain [11], [12]. According to [13] are three general conditions under which transfer learning can improve the learning process [14], [15]. First, the initial performance achieved through the transfer learning stage is higher than the performance of the model without the transfer learning process (higher start). Second, the ability of the model to learn the data during training increases (higher slope). Third, the final performance level that can be achieved by the transfer learning method is better than the model without transfer learning (higher asymptote) you. An illustration of the performance improvement through transfer learning is shown in Figure 3.

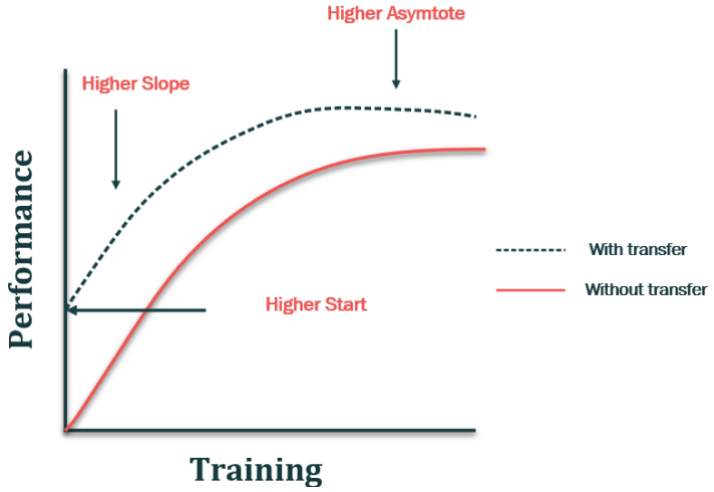

FIGURE 3. Improved Performance Using Transfer Learning.

In this research, the CNN architecture used for the transfer learning process is Faster RCNN [16], [17].

2.3 FASTER RCNN ARCHITECTURE

Faster RCNN is a CNN architecture used to perform object detection and classification processes [18], [19], [20]. This architecture is based on regional-based CNN. In general, the Faster RCNN process is as follows:

1. The input image is processed in the convolution layer which then produces a feature map of the image.
2. Then, the Region proposal network is applied to the feature maps. This process produces a proposal object along with its objetness score.
3. Region of Interest pooling layer is applied to each proposal object to equalize the size of all proposal objects.
4. Finally, the proposal object is passed to the fully connected layer which has a softmax layer and a linear regression layer, to perform object classification as well as regressing the bounding boxes of the object. An illustration of the Faster RCNN architecture is shown in Figure 4. Faster CNN architecture, is

used because it can perform classification and detection quickly and accurately when compared to its two predecessors Reginal Proposal Network [21] and Fast RCNN [22], [23].

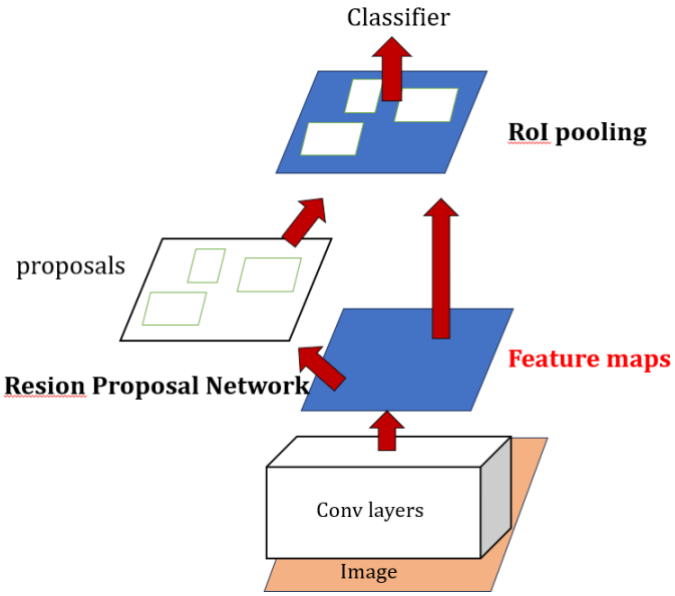

FIGURE 4. Faster RCNN Architecture.

2.4 EVALUATION METRIC

In this research, performance measurement uses the average precision and mean average precision methods. A The testing parameter for the detection and classification process is the mean average precision (mAP). Mean average precision is the average of the maximum precision values at different recall values. In calculating the average precision (AP), the results of all model predictions are collected and sorted by predicted confidence level (from highest to lowest confidence level). In addition, the prediction results are also evaluated. In the case of object detection, the prediction accuracy is calculated by comparing the bounding box of prediction and ground-truth (if, the IoU value of both ≥ 0.5 then the detection is considered correct).

The avarage precision (AP) value is calculated based on the area under the precision-recall graph [24]. However, to calculate the area, it is necessary to use an approximation equation to overcome the zigzag pattern in the graph. Each recall value in the graph (0, 0.1, ..., 1.0) is replaced by its precision value using equation (1). The smoothing result on the precision-recall graph is shown in Figure 4.

$$p_{interpolasi\ (r)} = \max_{\tilde{r} \geq r} p(\tilde{r}) \quad (1)$$

The mean average precision (mAP) is calculated by finding Average Precision (AP) for each class and then average over a number of classes [25].

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (2)$$

The mAP incorporates the trade-off between precision and recall and considers both false positives (FP) and false negatives (FN). This property makes mAP a suitable metric for most detection applications .

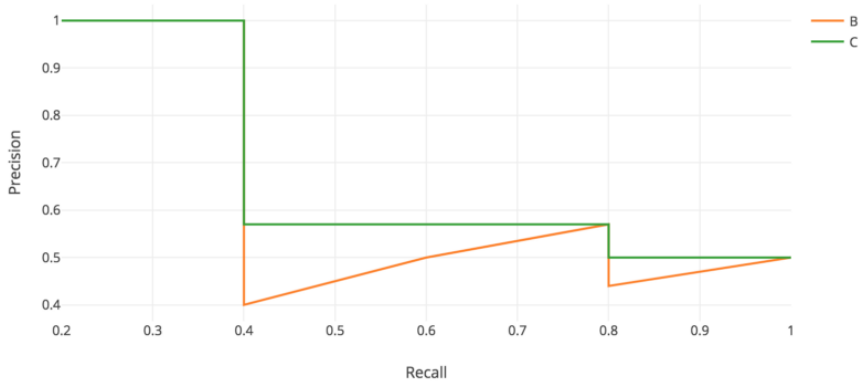

FIGURE 4. Smoothing Result on Precision-Recall Chart.

3. RESULT AND DISCUSSION

The transfer learning model experiment using the Faster RCNN architecture combines the detection and classification processes. The data used in the training process of the transfer learning model is a 20 piece video dataset. However, the video data is only extracted into frames and bounding box marking, and no object cutting process is performed. An example of training data in the training process using transfer learning is shown in Figure 5. The process of testing the transfer learning model is different from testing the previous models. This is because the Faster RCNN architecture handles two tasks at once, namely detection and classification. Model testing uses the mean average precision (mAP) method. The test results using the transfer learning model are shown in Figure 6.

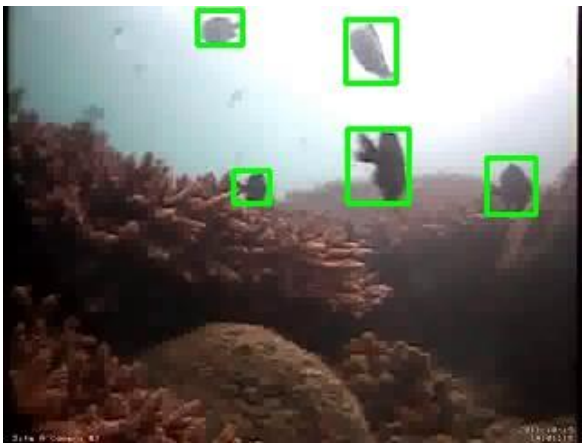

FIGURE 5. Contoh Data *Training* Model Transfer *Learning.*

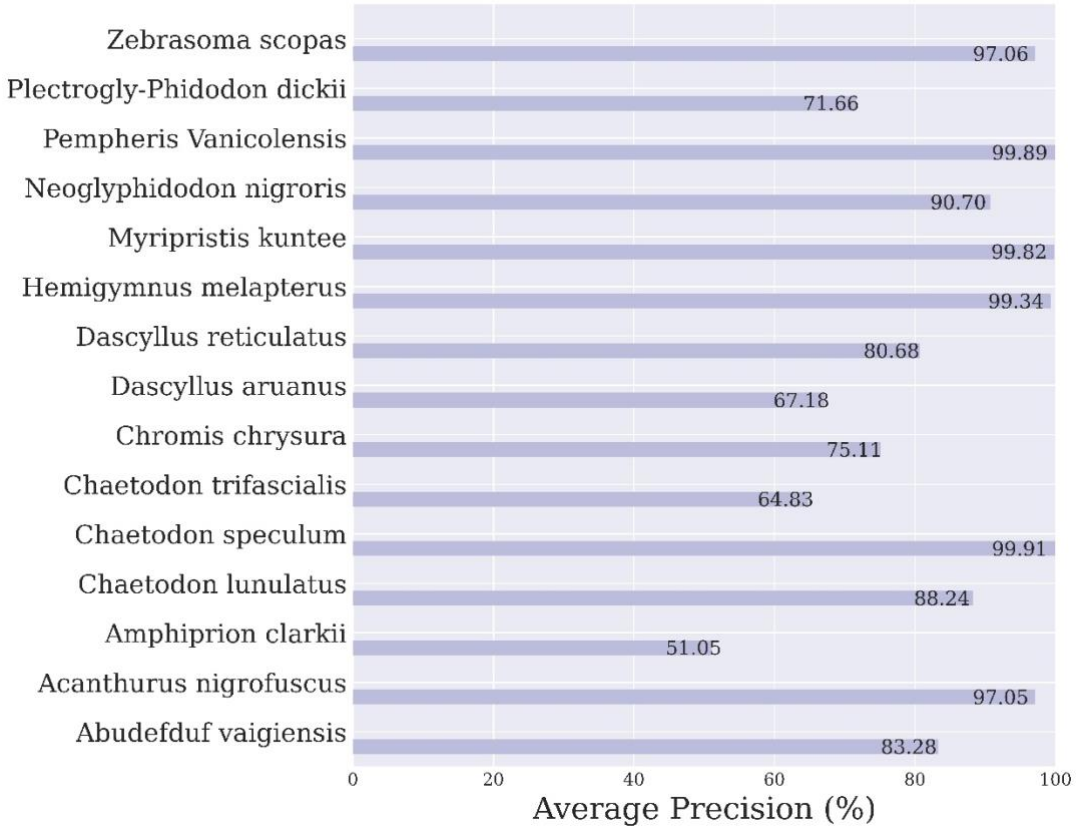

FIGURE 6. Faster RCNN Architecture AP Result in Percent (%)

Based on the test results, it is known that models using transfer learning techniques can improve classification accuracy. The mean average precision (mAP) produced by models built using transfer learning techniques is 84.39%. This increase in mAP is due to the fact that the model is built using an architecture that has been pre-trained using other more complex datasets. The amount of data used to build a pre-trained model is more than one million data for 1000 classes.

4. CONCLUSION

The Faster RCNN model for detecting fish species through video has been successfully developed and demonstrates a sufficiently high level of generalization. This architecture achieves a detection and classification accuracy with a mean average precision (mAP) of 84% on the video-based dataset. The suggestion that can be done for further research is to analyze the CNN architecture that has been robust so that it can determine the parameters that play a role in the process of improving classification accuracy.

REFERENCES

[1] M.-C. Chuang, J.-N. Hwang, and K. Williams, "A feature learning and object recognition framework for underwater fish images," *IEEE Trans. Image Process.*, pp. 1–1, 2016, doi: 10.1109/tip.2016.2535342.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

[2] K. Williams, N. Lauffenburger, M.-C. Chuang, J.-N. Hwang, and R. Towler, "Automated measurements of fish within a trawl using stereo images from a Camera-Trawl device (CamTrawl)," 2016.

[3] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. B. Fisher, "Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos," *Image Processing*, pp. 514–519, 2008.

[4] H. Qin, X. Li, J. Liang, Y. Peng, and C. Zhang, "DeepFish: Accurate underwater live fish recognition with a deep architecture," *Neurocomputing*, vol. 187, pp. 49–58, 2016, doi: 10.1016/j.neucom.2015.10.122.

[5] X. Li, M. Shang, J. Hao, and Z. Yang, "Accelerating fish detection and recognition by sharing CNNs with objectness learning," in *OCEANS 2016 - Shanghai*, 2016, pp. 1–5. doi: 10.1109/OCEANSAP.2016.7485476.

[6] M. C. Chuang, J. N. Hwang, and K. Williams, "Supervised and unsupervised feature extraction methods for underwater fish species recognition," in *Proceedings - 2014 ICPR Workshop on Computer Vision for Analysis of Underwater Imagery, CVAUI 2014*, Institute of Electrical and Electronics Engineers Inc., Nov. 2014, pp. 33–40. doi: 10.1109/CVAUI.2014.10.

[7] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A Survey of Deep Neural Network Architectures and Their Applications."

[8] A. Jalal, A. Salman, A. Mian, M. Shortis, and F. Shafait, "Fish detection and species classification in underwater environments using deep learning with temporal information," *Ecol Inform*, vol. 57, May 2020, doi: 10.1016/j.ecoinf.2020.101088.

[9] V. Sze, Y.-H. Chen, T.-J. Yang, and J. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," pp. 1–32, 2017.

[10] A. Joly, *LifeCLEF 2015 : Multimedia Life Species Identification Challenges To cite this version : HAL Id : hal-01182782 LifeCLEF 2015 : Multimedia Life Species Identification Challenges*. 2015.

[11] Y. Zhou, X. Zhang, Y. Wang, and B. Zhang, "Transfer learning and its application research," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, May 2021. doi: 10.1088/1742-6596/1920/1/012058.

[12] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, "A survey of transfer learning," *J Big Data*, vol. 3, no. 1, Dec. 2016, doi: 10.1186/s40537-016-0043-6.

[13] L. Torrey and J. Shavlik, "Transfer Learning," in *Handbook of Research on Machine Learning Applications and Trends*, IGI Global, 2010, pp. 242–264. doi: 10.4018/978-1-60566-766-9.ch011.

[14] W. Zhu, B. Braun, L. H. Chiang, and J. A. Romagnoli, "Investigation of transfer learning for image classification and impact on training sample size," *Chemometrics and Intelligent Laboratory Systems*, vol. 211, Apr. 2021, doi: 10.1016/j.chemolab.2021.104269.

[15] J. Praveen Gujjar, H. R. Prasanna Kumar, and N. N. Chiplunkar, "Image classification and prediction using transfer learning in colab notebook," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 382–385, Nov. 2021, doi: 10.1016/j.gltp.2021.08.068.

[16] W. Li, "Analysis of Object Detection Performance Based on Faster R-CNN," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2021. doi: 10.1088/1742-6596/1827/1/012085.

[17] N. Youssouf, "Traffic sign classification using CNN and detection using faster-RCNN and YOLOV4," *Heliyon*, vol. 8, no. 12, Dec. 2022, doi: 10.1016/j.heliyon.2022.e11792.

[18] F. Charli, H. Syaputra, M. Akbar[3], S. Sauda, and F. Panjaitan, "Implementasi Metode Faster Region Convolutional Neural Network (Faster R-CNN) Untuk Pengenalan Jenis Burung Lovebird," 2020. [Online]. Available: https://journal-computing.org/index.php/journal-ita/index

[19] Y. Zhu *et al.*, "Faster-RCNN based intelligent detection and localization of dental caries," *Displays*, vol. 74, Sep. 2022, doi: 10.1016/j.displa.2022.102201.

[20] Z. Li *et al.*, "A high-precision detection method of hydroponic lettuce seedlings status based on improved Faster RCNN," *Comput Electron Agric*, vol. 182, Mar. 2021, doi: 10.1016/j.compag.2021.106054.

[21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, 2016, doi: 10.1109/tpami.2015.2437384.

[22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., Curran Associates, Inc., 2015. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf

[23] Q. Zhang, Q. Yang, X. Zhang, Q. Bao, J. Su, and X. Liu, "Waste image classification based on transfer learning and convolutional neural network," *Waste Management*, vol. 135, pp. 150–157, Nov. 2021, doi: 10.1016/j.wasman.2021.08.038.

[24] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. Da Silva, "A comparative analysis of object detection metrics with a companion open-source toolkit," *Electronics (Switzerland)*, vol. 10, no. 3, pp. 1–28, Feb. 2021, doi: 10.3390/electronics10030279.

[25] P. Henderson and V. Ferrari, "End-to-end training of object class detectors for mean average precision," Jul. 2016, [Online]. Available: http://arxiv.org/abs/1607.03476

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)