IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 3, September 2024, pp. 3426~3436
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i3.pp3426-3436

Enhanced you only look once approach for automatic phytoplankton identification

Ovide Decroly Wisnu Ardhi[1,2,3], Tri Retnaningsih Soeprobowati[4,5,7], Kusworo Adi[4,6], Esa Prakasa[3], Arief Rachman[8]

[1]Department of Information System, School of Postgraduate Studies, Universitas Diponegoro, Semarang, Indonesia
[2]Department of Informatics Engineering, Vocational School, Universitas Sebelas Maret, Surakarta, Indonesia
[3]Research Center for Data and Information Science, National Research and Innovation Agency (BRIN), Bandung, Indonesia
[4]School of Postgraduate Studies, Universitas Diponegoro, Semarang, Indonesia
[5]Department of Biology, Faculty of Science of Mathematics, Universitas Diponegoro, Semarang, Indonesia
[6]Department of Physics, Faculty of Science of Mathematics, Universitas Diponegoro, Semarang, Indonesia
[7]Cluster for Paleolimnology, Universitas Diponegoro, Semarang, Indonesia
[8]Research Center for Oceanography, National Research and Innovation Agency (BRIN), Jakarta, Indonesia

Article Info

Article history:

Received Dec 30, 2023
Revised Feb 29, 2024
Accepted Mar 11, 2024

Keywords:

Asymmetric convolution
Marine biodiversity
Plankton detection
Vision transformer
YOLOv8

ABSTRACT

Conventionally, identifying phytoplankton species is challenging due to human taxonomical knowledge limitations. Advanced technology can overcome this problem. A novel model that accurately enhances phytoplankton detection and identification classification by combining asymmetric convolution and vision transformers (ACVIT) within the YOLOv8m framework is promoted with ACVIT-YOLO. The performance of this model surpasses the original YOLOv8m model, exhibiting a notable 2.4% enhancement in precision, 5.5% improvement in recall, and 1.1% gain in mAP 50 score. The enhanced effectiveness of ACVIT-YOLO compared to the YOLOv8m model, further demonstrated by the decreased giga floating-point operations (GFLOP), decreased parameter count, and compact dimensions, significantly improves the automation of phytoplankton species identification. This suggests that the ACVIT-YOLO model could produce a better prediction system for identifying phytoplankton with similar accuracy to the original YOLOv8m model but with lower computational power and resource usage.

Corresponding Author:

Ovide Decroly Wisnu Ardhi
Department of Information System, School of Postgraduate Studies, Universitas Diponegoro
St. Imam Bardjo SH No.5, RT.002, Pleburan, Semarang, Indonesia
Email: ovide@staff.uns.ac.id

1. INTRODUCTION

Phytoplankton plays a vital role in marine ecosystems as the primary producer, energy transfer link, and maintaining the biodiversity and ecological stability within the ecosystem [1], [2]. Phytoplankton plays an important role in oxygen production [3], influences regional and global temperatures, and regulate the aquatic biogeochemical cycle [4]. Phytoplankton also considered as a good bioindicator in aquatic ecosystems due to its rapid response to the environmental changes and anomalies [5].

Due to the country's unique geographical features, Indonesia is known for its high biodiversity and endemicity level in its marine ecosystems, particularly coral reef ecosystems [6]. Indonesian marine ecosystems should contain highly diverse and possibly endemic marine phytoplankton species. However, no official records of phytoplankton species exist, although several studies have documented between 150–400 species of marine phytoplankton in Indonesia [7], [8]. However, phytoplankton identification in Indonesia is still done conventionally using light microscopy, which is quite time-consuming and prone to error due to limitations in knowledge, the physiological, and psychological state of the human expert that performs the species identification [9]. Morphological similarities between plankton species often lead to misidentification [9].

The increasing anthropogenic activities, combined with global change, many ecological problems due to harmful algal blooms (HABs) tend to rise in frequency, duration, and impact in the last few decades [10] including in Indonesia [11]. Research indicates that at least 200 out of approximately 5000 known phytoplankton species produce toxins harmful to human health [10]. Consequently, rapid, accurate, and efficient identification of phytoplankton species is crucial to quickly identify the causative species of HABs and mitigate their negative effects on marine ecosystems, particularly in Indonesia.

Computer-assisted species identification using machine learning (ML) techniques has been developed recently. In recent years, the convolutional neural networks (CNN) method has performed well in determining plankton [12]. For example, the VGG-16 architecture can achieve excellent accuracy for classification tasks on the SIPPER dataset [13]. Similarly, a combination of inception/residual, VGGNet, and multi-layer perceptrons has achieved high precision [14]. More attractive models such as Detrecton2 Shufflenet V2, DenseNet201, and deep transfer learning (DTL) ResNet18 can demonstrate high performance identify various phytoplankton species [15]–[19]. Another approach using a combination of CNN, polar filter, support vector machine (SVM), and the PlanktonIK framework is promising [20]–[22]. Another significant advance is that a modified AlexNet model can achieve impressive precision [23], while deep learning with K-harmonic and SVM can also show good results [24]. Results from another study using application density-based spatial clustering were equally impressive [25].

A breakthrough in deep learning for real-time, fast, and accurate object recognition is the YOLO architecture [26], which has been demonstrated in a study to identify 80 diatom species precisely [27]. Similarly, the Algae-YOLO model with the efficient channel attention (ECA) attention mechanism achieves substantial mean average precision (mAP) values [28], while the YOLO-x-based approach yields more promising detection metrics [29]. The YOLOv3 model has higher accuracy in algal image identification, for example, in identifying microalgae growth in drinking water, compared to the more conventional approaches [30]. The complex YOLO model also helps classify HABs in custom datasets, and the data augmentation approach improves the model performance [31]. On the other hand, 1D asymmetric convolution is a new development in CNNs that improves CNN efficiency and resilience to rotational distortion in square convolution kernels without additional hyperparameters being needed [32]. At the same time, computer vision has been entirely transformed by the vision transformer [33]. While transformer's initial goal was to enhance English comprehension, it has since outperformed conventional CNNs in object identification and picture categorization, detection, and segmentation [34].

This research focuses on developing a deep-learning model designed to identify phytoplankton accurately and efficiently. By incorporating asymmetric convolution alongside vision transformers (ACVIT), we aim to explore the synergistic effects these technologies may have on improving the performance of YOLOv8m. The successful implementation of this final model is anticipated to play a crucial role in the advancement of an intelligent phytoplankton detection system, enabling precise classification of these organisms.

2. METHOD

In this study, we develop a modified model for detecting the phytoplankton using ACVIT-YOLO in the research workflow, as shown in Figure 1. The study consists of three main stages: preprocessing, processing, and model implementation. During preprocessing, image annotation with species names is conducted, followed by resizing the photos to 640×640 pixels. Augmentation techniques are then used to generate additional images. The processed images are divided into training, validation, and testing datasets. This phase is known as processing. Data splitting enables the training, validation, and testing of two distinct YOLOv8m architectures, one original and one modified namely ACVIT-YOLO. The final step involves evaluating and comparing the performance of all the models created.

2.1. Experiment environment

This experimental study was conducted using the Python 3+ language, PyTorch, and the ultralytics framework [35]. The setup for the experiment was implemented on Google Colab Pro+ using GPU specifications, the A100 (12 CPUs, 85.5 GB RAM). The A100 GPU was chosen because to its superior specifications in comparison to the V100 and T4 GPUs offered on Google Colab Pro+.

2.2. Data preprocessing and augmentation

This deep learning model research uses the plankton image database (cPID) at the Oceanographic Research Center, National Research and Innovation Agency (RCO-BRIN) [36]. The database utilizes images captured by an inverted microscope with phase contrast to enhance the visibility of transparent phytoplankton species [37], [38]. The PID comprises a varied plankton image collected 18 Indonesian marine locations, spanning 2011 to 2019. In the study, we annotated 324 images from this database using the LabelMe tool [39], focusing on 20 distinct species.

We utilized various augmentation techniques, including flipping, rotation, shearing, saturation, and exposure adjustments [40], to expand our plankton image set from 324 to 3,471 images (Figure 2). To further diversify the model, we applied standard albumentations [41] like blur and median blur at 1% probability with varying intensities, creating a spectrum of blurred images. In addition, we transformed color photos into grayscale with a 1% likelihood of replicating various real-life scenarios. The local contrast was enhanced by implementing the contrast-limited adaptive histogram equalisation (CLAHE) method, with a clip limit ranging from 1 to 4 and an 8×8 grid size for histogram equalization [42].

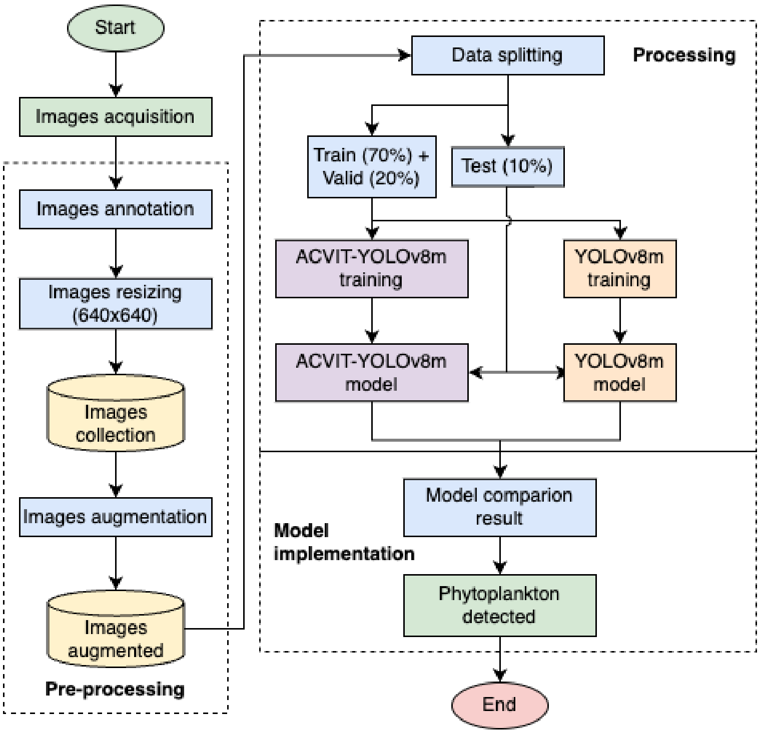


Figure 1. Research workflow of plankton detection

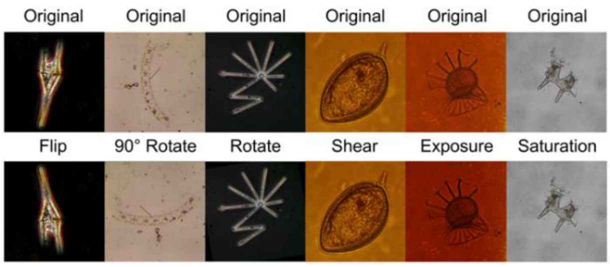


Figure 2. Original images and examples of image augmentations

2.3. Architectural designs of models

In this study, we use the YOLOv8 architecture that was developed by ultralytics in January 2023 [35], [43]. Initially, the YOLOv4 model enhances the efficacy and accuracy of neural networks by integrating the scaling cross stage partial network (CSP) [30]. The YOLOv4 model uses the scaling cross CSP to improve the performance and precision of neural networks by applying scaling algorithms and cross-stage partial connections [30]. A significant step forward in YOLOv5's development is the integration of bottleneck CSP with three convolutions (C3) [44]. An evolutionary step towards more incredible speed and efficiency was reducing the number of convolutions from three to two (C2f) [44]. To determine the loss for the bounding box, the YOLOv8 model integrates the C2f module with the complete intersection over union (CIoU) and distribution focal loss (DFL) loss functions [45]. The importance of YOLOv8's structure in phytoplankton identification is evaluated in this study. While YOLOv8m can simultaneously detect objects of various sizes within the same image, we develop a method to further enhance the detection capabilities of YOLOv8m by modifying the C2f module with asymmetric convolution as shown in Figure 3.

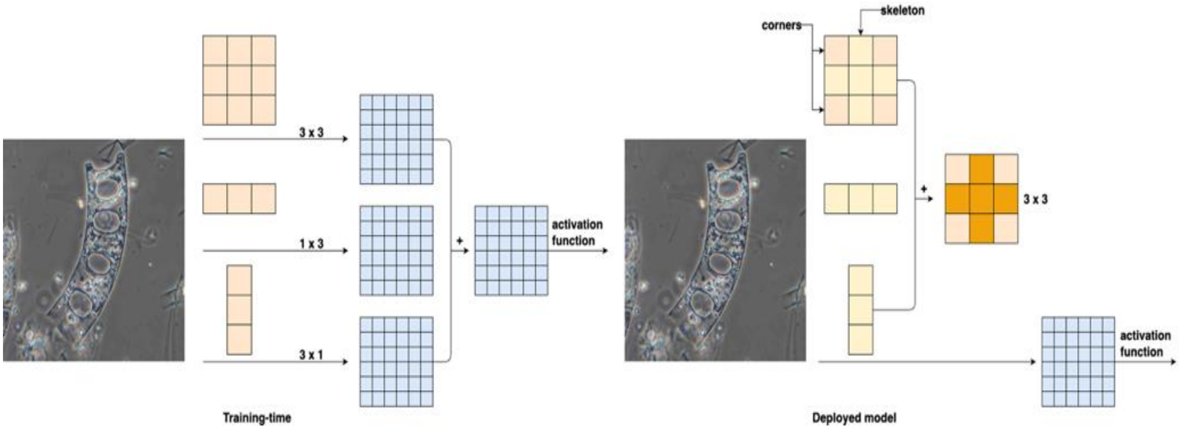


Figure 3. Conceptual of asymmetric convolution with the aims to increase the efficiency of convolution kernels and strengthen square convolution

On the other hand, we replace other C2f blocks using C3TR blocks to switch the resistance in the network. Multi-head attention in transformers aims to understand the overall context within an image. In (1) represents the transformation of the scaled dot-product attention from the vision transformer [46], replacing the C2f block. Where d_k is the dimension key used for scaling. Multi-head attention [46] is represented in (2).

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

$$MultiHead(Q,K,V) = Concat(head_1, \ldots., head_h)W^O \quad (2)$$

where $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$.

The enhanced YOLOv8m architecture in this study is depicted in Figure 4(a), showcasing the use of asymmetric convolution and vision transformers. Figure 4(a) illustrates the overall architecture of the improved ACVIT-YOLO derived from YOLOv8m. We propose the ACVIT-YOLO architecture in Figure 4(b) by modifying C2f to C2fx after two initial convolutions. Next, we replace the C2f block before SPPF with the C3TR module. Figures 4(c) and 4(d) show the bottleneck changes in C2fx with the 3×3 kernel changed to 1×3 and 3×1 kernels. Figure 4(e) shows the change in block C2f to C3TR.

2.4. Training and evaluation

In this study, we utilized the A100 GPU training environment and divided our dataset into training, validation, and testing (with a ratio of 70%: 20%: 10%). We employed stochastic gradient descent (SGD) to enhance and optimize the model efficiently [47]. A learning rate of 0.01 was applied to manage the pace of updating network weights, setting a balance between learning speed and accuracy [48]. A momentum set at 0.9 was used to accelerate SGD's movement in appropriate directions and minimize oscillations [49]. The training extended over 300 epochs, allowing the model ample time for iterative learning from the data and progressive performance improvement [50]. We analyzed the detection results using precision, recall, and mAP metrics [51], represented in (3)–(5).

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$mAP = \sum_{q=1}^{Q} \frac{AveP(q)}{q} \tag{5}$$

Additionally, a confusion matrix is utilized for a more nuanced assessment of our model, aligning actual and predicted categories and effectively visualizing sample distribution, thereby ensuring an in-depth understanding of the model's accuracy and efficacy across different scenarios [52].

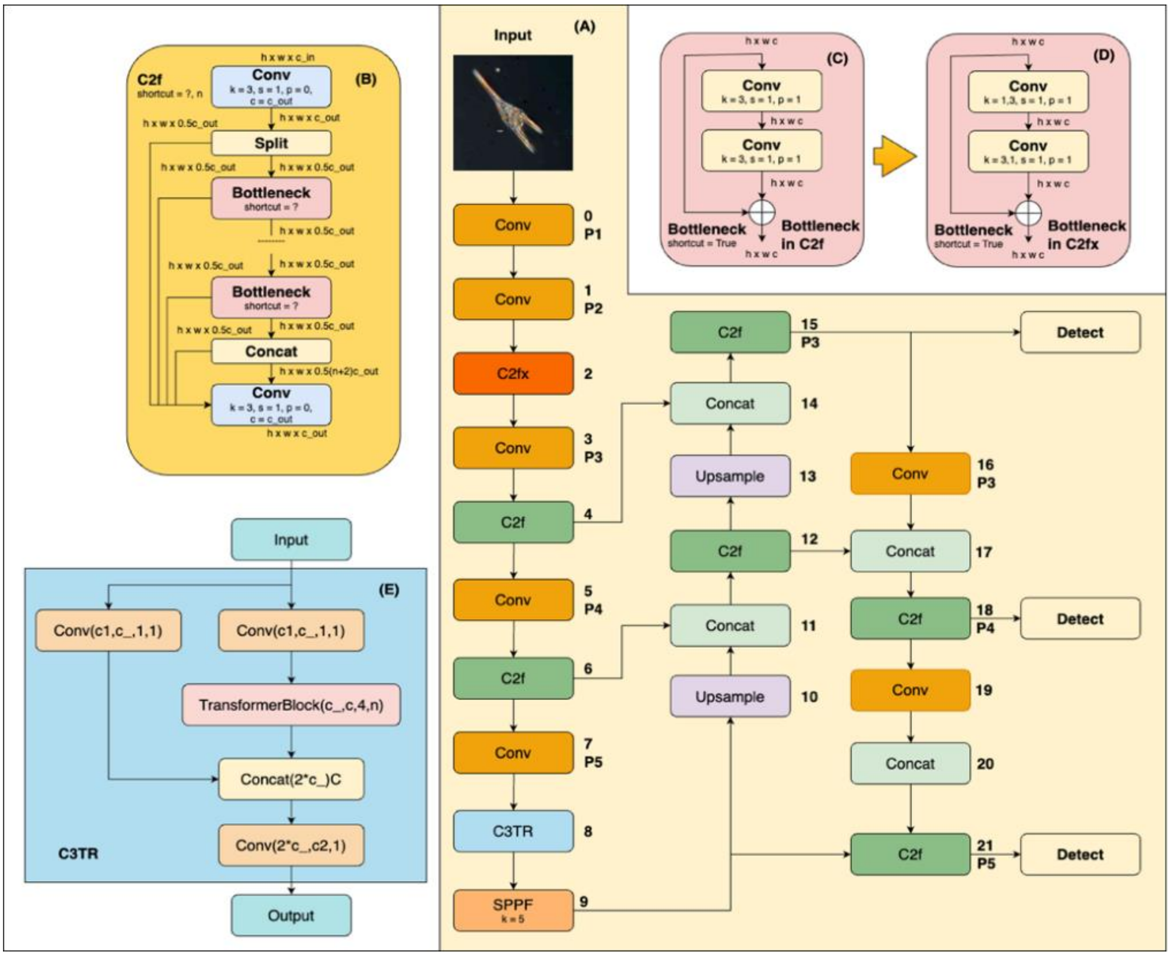


Figure 4. ACVIT-YOLOv8m architecture: (a) full architecture, (b) C2f, (c) Bottlenect in C2f, (d) bottleneck in C2fx, and (e) C3TR

3. RESULTS AND DISCUSSION

From this study, a detailed comparison of YOLOv8m and ACVIT-YOLO, focusing on performance and efficiency was shown in Table 1. ACVIT-YOLO demonstrates superior object detection capabilities with high precision (0.924) and recall (0.805) (Table 1). It surpasses YOLOv8m in mAP50 with a score of 0.911, although YOLOv8m leads in mAP50-95 with 0.851 (Table 1). The ACVIT-YOLO model has the lowest giga floating-point operations (GFLOP) value (74.9), the smallest number of parameters (24.07 million), and the smallest model size (48.5 MB). Thus, indicating lower computational power and resource use compared to the original YOLOv8m model. However, we do note that in the fastest training and inference time, YOLOv8m was faster than ACVIT-YOLO by reaching 2.58 hours and 3.1 ms. Figure 5 displays Radar graph of balance performance (precision, recall, mAP50, and mAP50-95) and efficiency of the original YOLOv8m (in yellow) and our ACVIT-YOLO (in purple). Figure 5 illustrates the performance and efficiency, with the original YOLOv8m model represented in yellow, while the ACVIT-YOLO model is depicted in purple.

The precision (x-axis) and recall (y-axis) metrics, along with mAP50 curves for both the YOLOv8m and ACVIT-YOLO was shown in Figure 6. In this case, we categorize species-based performance metrics into excellent (above 90%), good (80%-89%), moderate (70%-79%), and poor (below 40%). YOLOv8m model achieved excellent mAP50 scores for thirteen species, and five displayed good performance. However, each model showed poor performance for Odontella mobiliensis, with precision-recall less than 0.55 (Figure 6). Even so, the precision-recall data show that the performance of our ACVIT-YOLO can compete with the original YOLOv8m (Figure 6). Figure 6(a) displays the precision and recall curves of YOLOv8m for all phytoplankton species. Meanwhile, Figure 6(b) is for the ACVIT-YOLO model.

Table 1. Comparison of model performance (YOLOv8m and ACVIT-YOLO)

Performance comparison	Model	
	YOLOv8m	ACVIT-YOLO
Precision	0.924	0.924
Recall	0.798	0.805
mAP50	0.906	0.911
mAP50-95	0.851	0.849
Parameter (M)	25.85	24.07
FLOPs (G)	78.7	74.9
Size (MB)	52.1	48.5
Time (h)	2.0	2.1
Inference (ms)	3.6	3.9

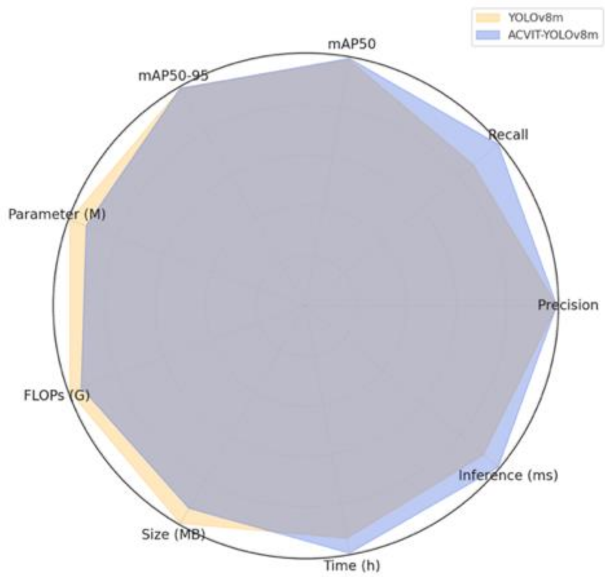
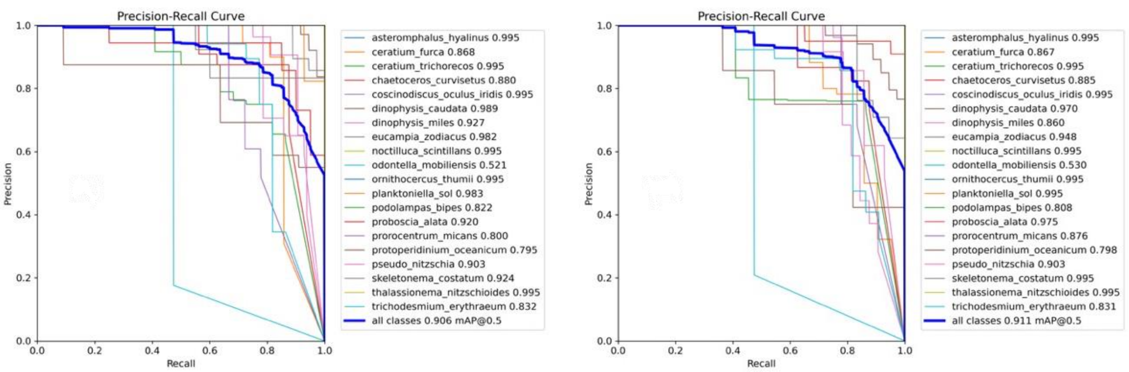


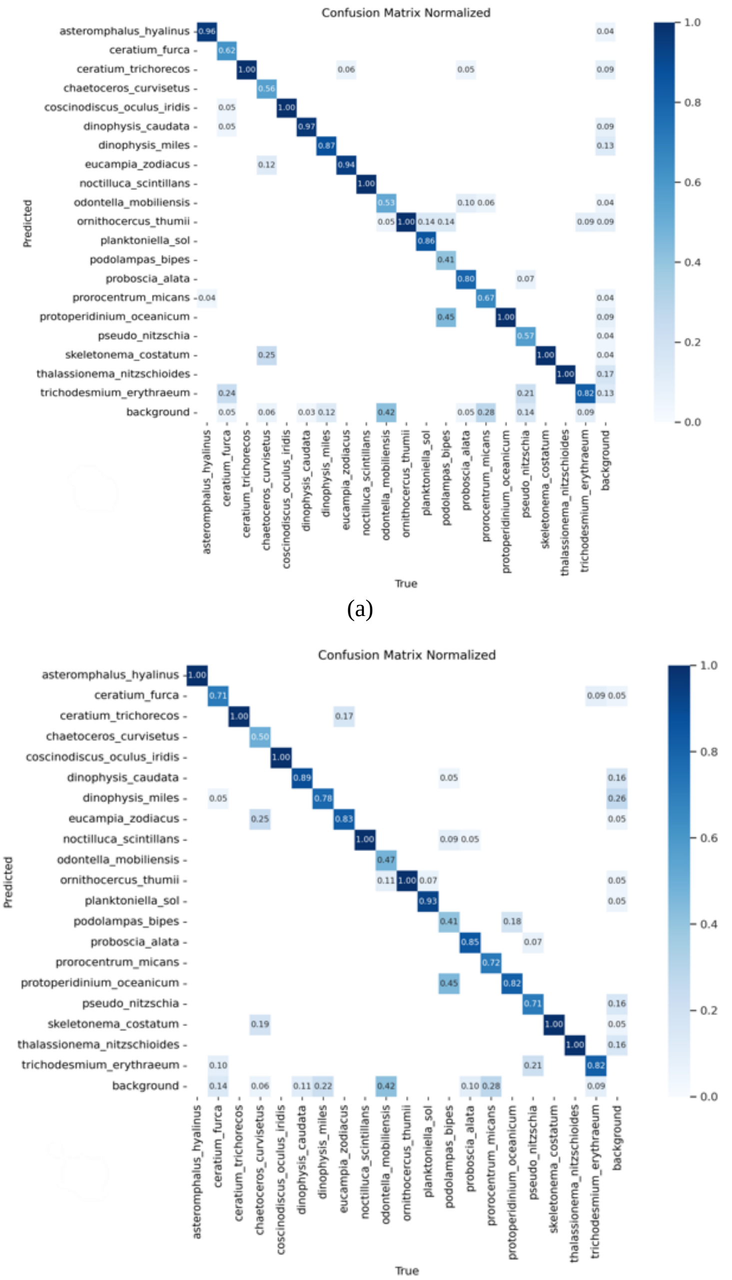
Figure 5. Radar graph of balance performance (precision, recall, mAP50, and mAP50-95) and efficiency of the original YOLOv8m and our ACVIT-YOLO



(a) (b)

Figure 6. Precision and recall curves: (a) YOLOv8m and (b) ACVIT-YOLO

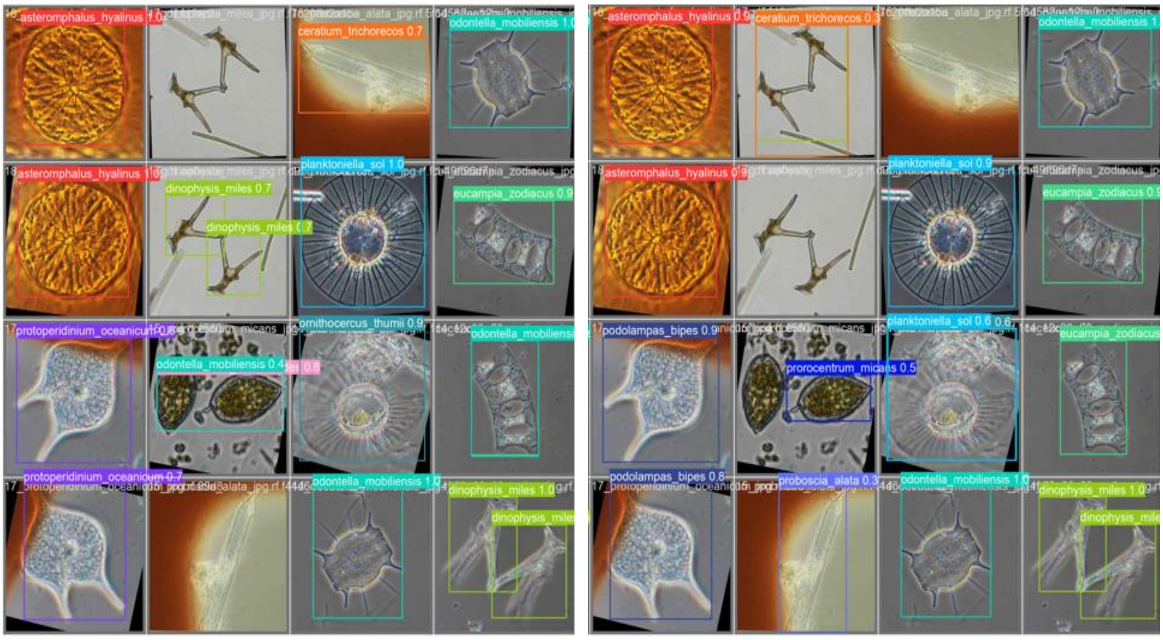
Confusion matrices evaluate and help provide insights into their accuracy and guide design optimizations. These two-dimensional matrices display classification accuracy by correlating actual and predicted categories, with cell data indicating sample proportions in each category. The results indicate a 100% accuracy for seven phytoplankton species across all models (Figure 7). Figure 7(a) indicates that YOLOv8m reached over 90% accuracy for three species, while Figure 7(b) shows that ACVIT-YOLO achieved this level of accuracy for one species. Additionally, YOLOv8m accurately identified four species with more than 80% precision, as depicted in Figure 7(a), and ACVIT-YOLO did the same for five species, as illustrated in Figure 7(b). However, the accuracy fell below 80% for the other species, which is consistently shown in both the YOLOv8m and ACVIT-YOLO confusion matrix (Figure 7). Consistent with the outcomes of precision-recall curves, as depicted in Figure 6, both YOLOv8m and ACVIT-YOLO show difficulties in identifying Odontella mobiliensis species (Figure 7). The reason for these identification problems or errors was unclear. However, based on our previous studies, such errors could caused by low image quality, incomplete cell components, overlapping cells (objects), or insufficient image data for training the model [9].



(b)

Figure 7. Comparison of confusion matrix (a) YOLOv8m and (b) ACVIT-YOLO

Even so, our proposed ACVIT-YOLO model indicates its proficiency by accurately recognizing diverse phytoplankton species, as shown in the detailed visualization of the prediction outcomes for different phytoplankton species (Figure 8). The ACVIT-YOLO model in our study could accurately detect the presence of phytoplankton cells and correctly predict the species name (Figure 8). Specifically, Figure 8(a) illustrates the detection results for phytoplankton species using the YOLOv8m model, whereas Figure 8(b) showcases the performance of the ACVIT-YOLO model.



(a) (b)

Figure 8. Box prediction comparison: (a) YOLOv8m and (b) ACVIT-YOLO

The results suggested that our architectural modifications incorporating asymmetric convolution and vision transformer in the YOLOv8m model could significantly enhance detection capabilities for phytoplankton species. The model's high detection capability was shown by the mAP50 value of 0.911 in our ACVIT-YOLO model, which is higher than the original YOLOv8m model. Integration of advanced techniques was known to demonstrate varied improvements in detection accuracies for various phytoplankton species. For example, using YOLOv5 with deep convolutional generative adversarial network (DC-GAN), yielding a map of 90.1%, underscores the benefit of generative adversarial networks in enhancing feature extraction and model performance [31]. Conversely, the YOLOx-s-FDA approach, with its mAP50 of 0.654 and precision of 0.959, highlights the model's precision in detections [29]. Furthermore, the adaptation of YOLOv5 with ShuffleV2 and the ECA attention module, achieving a mAP50 of 0.981, illustrates the efficacy of attention mechanisms in focusing on relevant features for improved detection [28]. These results suggest that future work could productively explore further YOLO modifications, mainly focusing on data augmentation with generative adversarial network (GAN) and including contemporary attention mechanisms to advance detection performance.

4. CONCLUSION

Our comprehensive study demonstrates the advanced object detection capabilities of ACVIT-YOLO, outperforming YOLOv8m in several key metrics. ACVIT-YOLO's high precision (0.924) and recall (0.805) alongside its superior mAP50 score (0.911) highlight its effectiveness, despite YOLOv8m's lead in mAP50-95 (0.851). Our evaluation, encompassing various species, shows both models excel in most categories, with ACVIT-YOLO having an edge in efficiency parameters like lower GFLOPs, fewer parameters, and smaller size. Despite YOLOv8m's faster training and inference, ACVIT-YOLO's balanced performance and efficiency offer significant advancements in object detection. This method may apply in biomonitoring activity or determine dominant species in HABs problems to develop suitable management.

ACKNOWLEDGEMENTS

This research, part of a degree by Research at the National Research and Innovation Agency (BRIN), received funding from the Research and Innovation for Advanced Indonesia (RIIM) program-Indonesia Endowment Fund for Education Agency (LPDP) at phase 3 and the Doctoral Dissertation Research (PDD) at Universitas Sebelas Maret, with grant numbers B-838/II.7.5/FR.06/5/2023 and B-4905/III.6/KS.00/5/2023 for (RIIM and LPDP), and 228/UN27.22/PT.01.03/2023 for PDD-UNS. The funders did not influence the experimental design, data collection and analysis, decision to publish, or manuscript preparation.

REFERENCES

[1] D. A. Hemraj, M. A. Hossain, Q. Ye, J. G. Qin, and S. C. Leterme, "Plankton bioindicators of environmental conditions in coastal lagoons," *Estuarine, Coastal and Shelf Science*, vol. 184, pp. 102–114, Jan. 2017, doi: 10.1016/j.ecss.2016.10.045.

[2] P. González, A. Castaño, E. E. Peacock, J. Díez, J. J. Del Coz, and H. M. Sosik, "Automatic plankton quantification using deep features," *Journal of Plankton Research*, vol. 41, no. 4, pp. 449–463, 2019, doi: 10.1093/plankt/fbz023.

[3] A. M. Redden, T. Kobayashi, I. Suthers, L. Bowling, D. Rissik, and G. Newton, "Plankton processes and the environment," in *Plankton: A guide to their ecology and monitoring for water quality*, Csiro Publishing, 2019, pp. 21–35.

[4] M. J. Behrenfeld *et al.*, "Biospheric primary production during an ENSO transition," *Science*, vol. 291, no. 5513, pp. 2594–2597, Mar. 2001, doi: 10.1126/science.1055071.

[5] T. R. Soeprobowati and J. Jumari, "The water quality index and phytoplankton communities of Kokoh Putih River, Sembalun, East Lombok, Indonesia," *Aquaculture, Aquarium, Conservation & Legislation*, vol. 15, no. 4, pp. 2025-2040, 2022.

[6] G. R. Allen, "Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes," *Aquatic Conservation*, vol. 18, no. 5, pp. 541–556, Jul. 2008, doi: 10.1002/aqc.880.

[7] A. Rachman, "Checklist and estimation of total number of phytoplankton species in Pari, Tidung, and Payung Islands, Indonesia," *Biodiversitas*, vol. 21, no. 6, 2020, doi: 10.13057/biodiv/d210616.

[8] E. Prakasa, A. Rachman, D. R. Noerdjito, and R. Wardoyo, "Development of segmentation algorithm for determining planktonic objects from microscopic images," *IOP Conference Series: Earth and Environmental Science*, vol. 944, no. 1, Dec. 2021, doi: 10.1088/1755-1315/944/1/012025.

[9] P. Culverhouse *et al.*, "AS WE SEE IT* Automatic image analysis of plankton: future perspectives," *Marine Ecology Progress Series*, vol. 312, pp. 297–309, Apr. 2006, doi: 10.3354/meps312297.

[10] G. M. Hallegraeff *et al.*, "Perceived global increase in algal blooms is attributable to intensified monitoring and emerging bloom impacts," *Communications Earth & Environment*, vol. 2, no. 1, Jun. 2021, doi: 10.1038/s43247-021-00178-8.

[11] S. H. Samudra *et al.*, "The phenomenon of harmful algae blooms (HABs) based on literature study in Indonesia sea waters from 2005-2021," *IOP Conference Series: Earth and Environmental Science*, vol. 1251, no. 1, Oct. 2023, doi: 10.1088/1755-1315/1251/1/012044.

[12] T. Eerola *et al.*, "Survey of automatic plankton image recognition: challenges, existing solutions and future perspectives," *Artificial Intelligence Review*, vol. 57, no. 5, Apr. 2024, doi: 10.1007/s10462-024-10745-y.

[13] H. Al-Barazanchi, A. Verma, and S. X. Wang, "Intelligent plankton image classification with deep learning," *International Journal of Computational Vision and Robotics*, vol. 8, no. 6, pp. 561–571, 2018, doi: 10.1504/IJCVR.2018.095584.

[14] T. Kerr, J. R. Clark, E. S. Fileman, and N. Pugeault, "Collaborative deep learning models to handle class imbalance in FlowCam plankton imagery," *IEEE Access*, vol. 8, pp. 170013-170032, 2020, doi: 10.1109/ACCESS.2020.3022242.

[15] G. Wacquet and A. Lefebvre, "EcoTransLearn: an R-package to easily use transfer learning for ecological studies—a plankton case study," *Bioinformatics*, vol. 38, no. 24, pp. 5469–5471, Dec. 2022, doi: 10.1093/bioinformatics/btac703.

[16] Y. Zhang, Y. Lu, H. Wang, P. Chen, and R. Liang, "Automatic classification of marine plankton with digital holography using convolutional neural network," *Optics & Laser Technology*, vol. 139, 2021, doi: 10.1016/j.optlastec.2021.106979.

[17] K. Cheng, X. Cheng, Y. Wang, H. Bi, and M. C. Benfield, "Enhanced convolutional neural network for plankton identification and enumeration," *PLoS ONE*, vol. 14, no. 7, Jul. 2019, doi: 10.1371/journal.pone.0219570.

[18] B. Guo *et al.*, "Automated plankton classification from holographic imagery with deep convolutional neural networks," *Limnology and Oceanography: Methods*, vol. 19, no. 1, pp. 21–36, 2021, doi: 10.1002/lom3.10402.

[19] S. Kakehi *et al.*, "Identification and counting of pacific oyster crassostrea gigas larvae by object detection using deep learning," *Aquacultural Engineering*, vol. 95, Nov. 2021, doi: 10.1016/j.aquaeng.2021.102197.

[20] Z. Yang, J. Li, T. Chen, Y. Pu, and Z. Feng, "Contrastive learning-based image retrieval for automatic recognition of *in situ* marine plankton images," *ICES Journal of Marine Science*, vol. 79, no. 10, pp. 2643–2655, Dec. 2022, doi: 10.1093/icesjms/fsac198.

[21] X. Cheng, Y. Ren, K. Cheng, J. Cao, and Q. Hao, "Method for training convolutional neural networks for in situ plankton image recognition and classification based on the mechanisms of the human eye," *Sensors*, vol. 20, no. 9, pp. 1-16, 2020, doi: 10.3390/s20092592.

[22] R.-M. Plonus, J. Conradt, A. Harmer, S. Janßen, and J. Floeter, "Automatic plankton image classification—Can capsules and filters help cope with data set shift?," *Limnology and Oceanography: Methods*, vol. 19, no. 3, pp. 176–195, 2021, doi: 10.1002/lom3.10413.

[23] S. Hong *et al.*, "Classification of freshwater zooplankton by pre-trained convolutional neural network in underwater microscopy," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, pp. 252-258, 2020, doi: 10.14569/IJACSA.2020.0110733.

[24] J. Y. Luo *et al.*, "Automated plankton image analysis using convolutional neural networks," *Limnology and Oceanography: Methods*, vol. 16, no. 12, pp. 814–827, 2018, doi: 10.1002/lom3.10285.

[25] L. MacNeil, S. Missan, J. Luo, T. Trappenberg, and J. LaRoche, "Plankton classification with high-throughput submersible holographic microscopy and transfer learning," *BMC Ecology and Evolution*, vol. 21, no. 1, pp. 1-11, Dec. 2021, doi: 10.1186/s12862-021-01839-0.

[26] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: scaling cross stage partial network," in *IEEE/cvf conference on computer vision and pattern recognition*, pp. 13024-13033, doi: 10.1109/CVPR46437.2021.01283.

[27] J. Salido, C. Sánchez, J. R. -Santaquiteria, G. Cristóbal, S. Blanco, and G. Bueno, "A low-cost automated digital microscopy platform for automatic identification of diatoms," *Applied Sciences*, vol. 10, no. 17, Aug. 2020, doi: 10.3390/app10176033.

[28] D. Liu, P. Wang, Y. Cheng, and H. Bi, “An improved algae-YOLO model based on deep learning for object detection of ocean microalgae considering aquacultural lightweight deployment,” *Frontiers in Marine Science*, vol. 9, Nov. 2022, doi: 10.3389/fmars.2022.1070638.
[29] H. Yan *et al.*, “YOLOx model-based object detection for microalgal bioprocess,” *Algal Research*, vol. 74, Jul. 2023, doi: 10.1016/j.algal.2023.103178.
[30] J. Park, J. Baek, K. You, S. W. Nam, and J. Kim, “Microalgae detection using a deep learning object detection algorithm, YOLOv3,” *Journal of Korean Society on Water Environment*, vol. 37, no. 4, pp. 275–285, Jul. 2021, doi: 10.15681/KSWE.2021.37.4.275.
[31] Abdullah, S. Ali, Z. Khan, A. Hussain, A. Athar, and H.-C. Kim, “Computer vision based deep learning approach for the detection and classification of algae species using microscopic images,” *Water*, vol. 14, no. 14, Jul. 2022, doi: 10.3390/w14142219.
[32] X. Ding, Y. Guo, G. Ding, and J. Han, “ACNet: strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 1911–1920. doi: 10.1109/ICCV.2019.00200.
[33] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-end object detection with transformers,” in *Computer Vision – ECCV 2020*, vol. 12346, Cham: Springer International Publishing, 2020, pp. 213–229. doi: 10.1007/978-3-030-58452-8_13.
[34] A. Dosovitskiy *et al.*, “An image is worth 16x16 words: transformers for image recognition at scale,” *arXiv-Computer Science*, pp. 1-22, 2023.
[35] G. Jocher, A. Chaurasia, and J. Qiu, “YOLO by Ultralytics,” *GitHub*. Accessed: Sep. 09, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
[36] A. Rachman, A. S. Suwarno, and S. Nurdjaman, “Application of deep (machine) learning for phytoplankton identification using microscopy images,” *7th International Conference on Biological Science (ICBS 2021)*, Yogyakarta, Indonesia, pp. 213-224, 2022, doi: 10.2991/absr.k.220406.032.
[37] F. Zernike, “Phase contrast, a new method for the microscopic observation of transparent objects part II,” *Physica*, vol. 9, no. 10, pp. 974–986, Dec. 1942, doi: 10.1016/S0031-8914(42)80079-8.
[38] K. Thorn, “A quick guide to light microscopy in cell biology,” *MBoC*, vol. 27, no. 2, pp. 219–222, Jan. 2016, doi: 10.1091/mbc.e15-02-0088.
[39] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, “LabelMe: a database and web-based tool for image annotation,” *International Journal of Computer Vision*, vol. 77, no. 1, pp. 157–173, May 2008, doi: 10.1007/s11263-007-0090-8.
[40] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0197-0.
[41] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, “Albumentations: fast and flexible image augmentations,” *Information*, vol. 11, no. 2, Feb. 2020, doi: 10.3390/info11020125.
[42] M. Opoku, B. A. Weyori, A. F. Adekoya, and K. Adu, “CLAHE-CapsNet: Efficient retina optical coherence tomography classification using capsule networks with contrast limited adaptive histogram equalization,” *PLoS ONE*, vol. 18, no. 11, Nov. 2023, doi: 10.1371/journal.pone.0288663.
[43] M. Krishnakumar, “A gentle introduction to YOLOv8,” *Weights and Bias*, 2023. Accessed: Sep. 10, 2023. [Online]. Available: https://wandb.ai/mukilan/wildlife-yolov8/reports/A-Gentle-Introduction-to-YOLOv8--Vmlldzo0MDU5NDA2
[44] “Dive into YOLOv8: How does this state-of-the-art model work?” *openMMLab*. [Online]. Available: https://openmmlab.medium.com/dive-into-yolov8-how-does-this-state-of-the-art-model-work-10f18f74bab1
[45] J. Terven and D. C. -Esparza, “A comprehensive review of YOLO: from YOLOv1 and beyond,” *arXiv-Computer Science*, pp. 1-36, 2023.
[46] Vaswani A. *et al.*, “Attention is all you need,” *Arxiv-Computer Science*, pp. 1–15, 2017.
[47] J. Yang, Y. Wang, Z. Wang, J. Xu, X. Zhang, and S. Tang, “Using stochastic gradient descent and deep learning to defect detection for medicinal hollow capsule,” in *2021 China Automation Congress (CAC)*, 2021, pp. 4926–4933. doi: 10.1109/CAC53003.2021.9728658.
[48] H. Zhang *et al.*, “Recognition of the rhizome of red ginseng based on spectral-image dual-scale digital information combined with intelligent algorithms,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 297, Sep. 2023, doi: 10.1016/j.saa.2023.122742.
[49] M. O. Lawal, “Tomato detection based on modified YOLOv3 framework,” *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-81216-5.
[50] E. Casas, L. Ramos, E. Bendek, and F. Rivas-Echeverría, “Assessing the effectiveness of YOLO architectures for smoke and wildfire detection,” *IEEE Access*, vol. 11, pp. 96554–96583, 2023, doi: 10.1109/ACCESS.2023.3312217.
[51] D. Müller, I. S. -Rey, and F. Kramer, “Towards a guideline for evaluation metrics in medical image segmentation.,” *BMC Research Notes*, vol. 15, no. 1, Jun. 2022, doi: 10.1186/s13104-022-06096-y.
[52] I. Düntsch and G. Gediga, “Confusion matrices and rough set data analysis,” *Journal of Physics: Conference Series*, vol. 1229, no. 1, pp. 1–6, 2019, doi: 10.1088/1742-6596/1229/1/012055.

BIOGRAPHIES OF AUTHORS

Ovide Decroly Wisnu Ardhi obtained a Bachelor of Engineering (B.Eng.) in Electrical Engineering from the Universitas Islam Indonesia in 2008. He then pursued a Master of Engineering (M.Eng.) with a specialization in Information Technology from the Department of Electrical Engineering at Universitas Gadjah Mada, which he completed in 2012. He is pursuing a Ph.D. in Information Systems at the School of Postgraduate Studies, Universitas Diponegoro. Aside from pursuing his doctoral studies, he is a lecturer in the Department of Informatics Engineering, Vocational School at Universitas Sebelas Maret, Surakarta, Indonesia. He is primarily interested in information systems, artificial intelligence, machine learning, deep learning, and bioinformatics. He can be contacted at email: ovide@staff.uns.ac.id or ovide@students.undip.ac.id.

Prof. Dr. Tri Retnaningsih Soeprobowati, MAppSc. a professor in the Department of Biology, Faculty of Science and Mathematics, and the School of Postgraduate Studies at Universitas Diponegoro, Semarang, Indonesia, obtained her bachelor's degree in biology from Universitas Gadjah Mada in 1987. She furthered her education with a Master of Applied Science in Environmental Science from the University of Western Sydney-Hawkesbury, NSW, Australia, in 1996. She completed her Ph.D. in Environmental Science at Universitas Gadjah Mada in 2010. Her expertise primarily lies in environmental sciences and ecology, engineering, and science and technology. She can be contacted at email: trsoeprobowati@live.undip.ac.id.

Prof. Dr. Kusworo Adi, S.T., M.T. a professor in the field of Instrumental Physics at the Department of Physics, Faculty of Science and Mathematics, and the School of Postgraduate Studies at Universitas Diponegoro, Semarang, Indonesia, earned his bachelor's degree in physics from Universitas Diponegoro in 1996. He obtained his Master's in Image Processing from the Institut Teknologi Bandung in 2006. He completed his Ph.D. in Imaging and Image Processing at the Institut Teknologi Bandung in 2010. His research interests are focused on artificial intelligence and computer vision. He can be contacted at email: kusworoadi@lecturer.undip.ac.id.

Dr. Esa Prakasa, S.T., M.T. obtained his Bachelor of Engineering (B.Eng.) in 1998 after completing his undergraduate studies in Nuclear Engineering at Universitas Gadjah Mada, Indonesia. Obtaining a Master of Engineering (M.Eng.) in Electrical Engineering in 2001, he continued his academic pursuits at the identical institution. The Malaysian institution Universiti Teknologi PETRONAS awarded him a Doctor of Philosophy (Ph.D.) in Electrical and Electronic Engineering in 2014. He is a senior researcher at Research Center for Data and Information Sciences of the National Research and Innovation Agency (BRIN). His research interests are image and visual data analysis, pattern recognition, and computer vision. He can be contacted at email: esap001@brin.go.id.

Arief Rachman, M.Biol.Sc. works as a researcher at the Plankton Laboratory, a part of the Research Center for Oceanography under the National Research and Innovation Agency (BRIN). He holds a Master of Biological Science (M.Biol.Sc.). He specializes in marine phytoplankton, particularly studying harmful algae's dynamic and ecological properties. Furthermore, he is accountable for researching these fields, assessing samples and data, and providing direction and instructional supervision to students involved in laboratory internships, final projects, theses, or dissertations. He can be contacted at email: ariefrachman1987@yahoo.com.