International Journal of Advances in Applied Sciences (IJAAS)
Vol. 13, No. 4, December 2024, pp. 1013~1022
ISSN: 2252-8814, DOI: 10.11591/ijaas.v13.i4.pp1013-1022

Enhanced U-Net models with encoder and augmentation for phytoplankton segmentation

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

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

Article Info

Article history:

Received Dec 30, 2023
Revised May 19, 2024
Accepted Sep 9, 2024

Keywords:

Augmentation
Encoder
Phytoplankton recognition
Semantic segmentation
U-Net

ABSTRACT

This study comprehensively analyzes U-Net models for semantic segmentation in phytoplankton image recognition, leveraging encoders such as EfficientNet-B5, MobileNetV2, ResNet50, and ResNeXt50 and employing the Adam optimizer. The research highlights the U-Net MobileNetV2 model with optical distortion, which achieves notable test scores with 93.69% Dice, 88.14% intersection over union (IoU), 99.89% Precision, and 100% Recall, underscoring the efficacy of the applied augmentation strategies, including geometric and distortion transforms, and color and blur techniques. The U-Net ResNet50 model with mix transform consistently demonstrates high accuracy in critical metrics, outperforming others, while EfficientNet-B5 with blur suggests increased model sensitivity with improved recall. These results underscore the crucial role of encoder-augmentation synergy in model performance. Training and testing times across models have remained under 250 seconds, reflecting methodological efficiency. Overall, these results demonstrate the model's excellent performance for the semantic segmentation task.

Corresponding Author:

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

1. INTRODUCTION

Indonesia is one the megadiversity countries and contains part of 'the heart of coral triangle', which is well known for its high diversity level of coral reef and reef fish species, which consists of many endemic species [1]. Similar to coral reefs and reef fishes, it is assumed that phytoplankton communities within Indonesian marine ecosystems exhibit considerably high diversity, with some species potentially endemic. Although no accurate official records exist, various studies have documented between 150 to 400 species of marine phytoplankton in Indonesia, with some species suspected to be endemic, new records, or new species [2]–[6]. As a fundamental component of highly productive tropical marine ecosystems, phytoplankton communities perform vital ecological functions as the primary producers, essential links in energy transfer, and stabilizers of the biodiversity and environmental balance in the ecosystem [7]–[9]. Furthermore, phytoplankton communities also drive the aquatic biogeochemical cycles, which influence the entire marine food web while also producing oxygen and regulating the regional and global ocean temperature. Due to the quick response of phytoplankton communities towards changes in water environmental conditions, they are regarded as excellent bioindicators and often used as proxies for ecological monitoring to detect anomalies or disturbances, such as eutrophication or pollution, within the coastal ecosystem [7], [9]–[11].

In recent decades, aquatic ecological problems, such as harmful algal blooms (HABs), have increased in many tropical countries, including Indonesia [12], driven by increased human activity combined with the effect of climate change. The combined impact of climate change and anthropogenic activities have changed the environmental conditions, provided more competitive advantages, promoted toxin producers, and increased some HAB species' biomass production [13], [14]. From 2005 to 2021, at least 58 different studies have reported cases of HABs in other coastal areas in Indonesia, such as Lampung Bay, Jakarta Bay, and Ambon Bay [14]. Negative impacts, including mass fish mortality and fatal paralytic shellfish poisoning (PSP) disease in humans, have been reported from many HAB-affected areas, particularly in Cirebon [14]. Given that, at least 200 out of the 5,000 known phytoplankton species produce toxins harmful to humans [14] and at least 10 out of 17 most common HABs causative species in Indonesia were toxin producers [14], rapid and accurate genus or species level identification have become essential to identify and mitigate the effects of their blooms on marine ecosystems and coastal communities.

So far, light microscopy continues to be the preferred method for phytoplankton identification in Indonesia despite its labor-intensive nature, susceptibility to errors, and limitations imposed by human physiological and psychological factors [14]. Misidentification is a common issue, often due to the morphological similarities between different plankton species [14]. However, machine learning (ML) methods using convolutional neural networks (CNN) have been used and shown excellent performance in identifying various phytoplankton species in recent years [15]. In phytoplankton research, the substantial variability in image sizes and quality poses a significant challenge for conventional CNN-based classifiers, which typically require uniform-sized input images [16]. This necessity for resizing, a process that either disregards [17], [18] or preserves the image's aspect ratio [8], [19], [20], comes with its set of advantages and disadvantages. While ignoring the aspect ratio can distort images and potentially affect feature extraction and learning outcomes, preserving it, especially when aligning with the image's background color, has produced optimal results on imaging FlowCytobot (IFCB) data [21]. However, resizing inevitably leads to some loss of detail and information, especially concerning phytoplankton size. This issue is somewhat mitigated by including image size information as metadata [22]. Moreover, the variability in image quality within these datasets necessitates employing various pre-processing techniques to enhance classification accuracy [16]. Strategies such as discarding low-quality images [23], segmentation [24], and denoising [25], along with advancements like CNN-based super-resolution techniques, have been proposed to refine image quality, though the effectiveness of these approaches in improving recognition accuracy warrants further investigation [26].

The U-Net architecture, which emphasizes encoder-decoder methods, represents notable progress in addressing semantic segmentation difficulties [27]. The objective of this design is to reduce noise at different scales in order to enhance performance by employing a comprehensive refining methodology [28]. A novel methodology has been developed that combines conventional neural differential equations with the structural framework of U-Net to delineate cell regions accurately [29]. The use of U-Net in microscopic segmentation is very effective, so it has been put together in different ways to get better results [30]. The purpose of the environmental microorganism data set fifth version (EMDS-5) dataset is to evaluate and compare the U-Net method for dividing numerous objects in images [31]. As an alternative, the low-cost U-Net (LCU) was created, which builds on the basic U-Net model to make it easier to separate images of microorganisms that live in the environment while also trying to use as little memory as possible [32]. According to previous research, the U-Net++ variation works better than regular patch-level segmentation methods that use visual transformers to get accurate data on microorganisms in the environment at the pixel level [33].

Previous research focused on modifying the U-Net architecture for semantic segmentation but did not explicitly discuss the influence of the encoder on the U-Net architecture and image augmentation techniques. This paper proposes an improvement of the U-Net architecture, several encoder and decoder networks have been integrated with the U-Net. Various image augmentation approaches are implemented, such as geometric, distortion, color, and blur transformation types, to elevate the algorithm's reliability. The main contribution of this study is to assess the integration of the encoder within the U-Net architecture for semantic segmentation of transparent phytoplankton images. Additionally, the research explores different strategies for enhancing the images. The purpose of this study is to showcase the remarkable capability of U-Net in detecting and utilizing features from images that exhibit substantial variations in terms of size, shape, and clarity. Moreover, it does a comparative analysis of the collaboration between the encoder and decoder using various augmentation techniques, thereby facilitating the implementation of ensemble learning methodologies. The comparison above not only underscores the versatility of U-Net in terms of various preprocessing methods but also its ability to enhance the robustness and precision of segmentation outcomes.

2. METHOD

The present study involved the development of a segmentation model for transparent images of phytoplankton. The U-Net architecture was employed inside a research workflow, as depicted in Figure 1. This study has three primary phases: preparation, processing, and model implementation. During the pre-processing stage, the image was annotated with species names, and subsequently, the image was resized to dimensions of 224×224 pixels. Subsequently, several augmentation techniques are employed to generate supplementary photos. The treated images are partitioned into several datasets for training, validation, and testing purposes. This stage is commonly referred to as processing. The utilization of data separation facilitates the process of training, validating, and testing four distinct U-Net architectures. These networks include several encoders and are trained using images obtained from various augmentation techniques. The final stage is assessing and contrasting the efficacy of all the models generated.

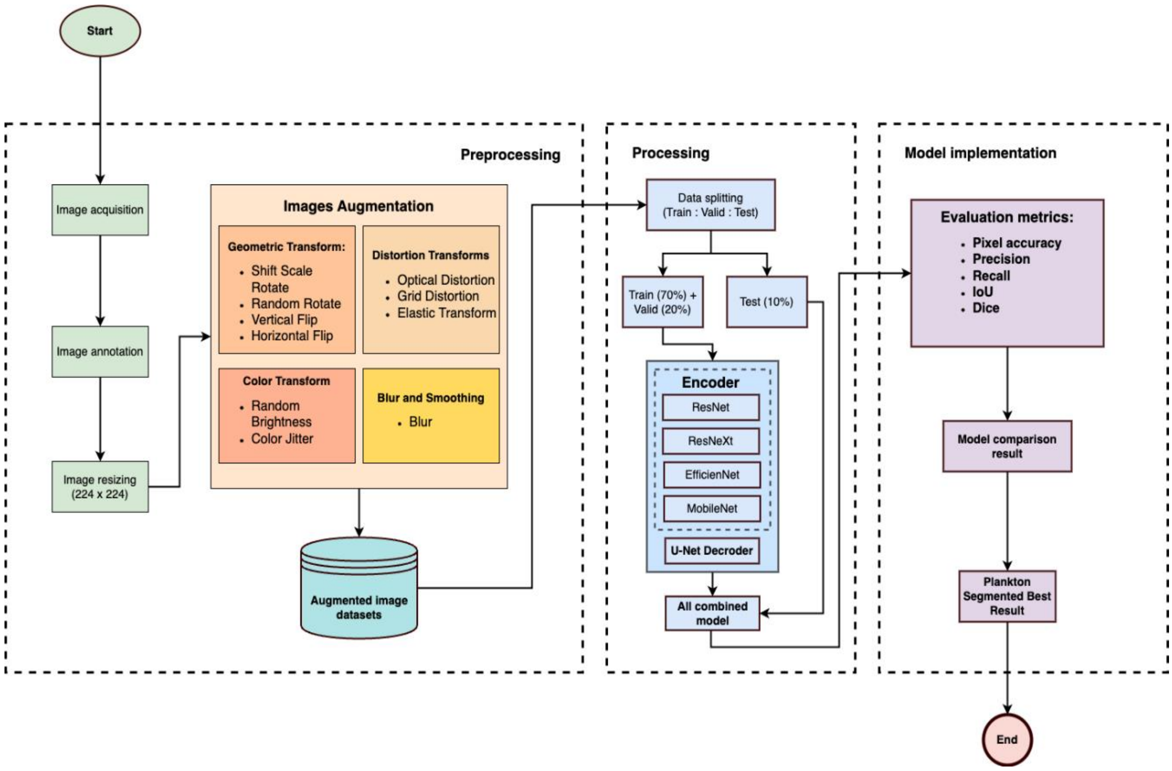

Figure 1. Research workflow of phytoplankton semantic segmentation

2.1. Experiment environment

The model was built and utilized with PyTorch, a widely used machine learning framework and programming language Python 3+. Our experimental setup was carried out on Google Colab Pro+, which offered varying GPU specifications, such as K80, T4, P100, and 52 GB of RAM. This research utilizes high-resolution microscopy HABs data from the phytoplankton image database (cPID), housed at the Oceanographic Research Center, National Research and Innovation Agency (RCO-BRIN) [2]. The data used in this study were collected during a research expedition spanning from 2011 to 2019, encompassing 18 locations in Indonesian waters.

2.2. Data preprocessing and augmentation

For this work, we utilized the LabelMe application [34] to annotate images from the database. This study partitions the data in a 70:20:10% ratio and employs various data augmentation techniques to address data scarcity. The Albumentations library in Pytorch, known for its efficiency in image transformation, supports this process [35]. Table 1 lists the augmentation types used to enrich the image dataset [36].

Augmentation techniques are typically categorized into four groups: geometric transformation, distortion transformation, color transformation, and blur and smoothing. Geometric transformations, such as shift scale rotation, are utilized to change images through actions such as movement, zooming, and spinning. On the other hand, random rotation is employed to rotate images at intervals of 90°. The images are flipped using probabilities in the vertical flip and horizontal flip functions. Different distortion transformations, such as optical distortion, grid distortion, and elastic transform, are employed to introduce a range of optical and wave-like distortions. Random brightness and color jitter are utilized to randomly modify the brightness and various other attributes of colors. The application of a blurring effect to images facilitates the process of augmentation.

Table 1. Augmentation techniques

Geometric transforms	Distortion transforms	Color transforms	Blur and smoothing
Shift scale rotate	Optical distortion	Random brightness	Blur
Random rotate	Grid distortion	Color jitter	
Vertical flip	Elastic transform	Hue saturation	
Horizontal flip			

2.3. Model architectures

To improve segmentation performance, this study employs transfer learning with encoders that have been pre-trained, a method proven to enhance model performance [37]. This experiment uses encoders from the ResNet50, ResNeXt, MobileNetV2, and EfficientNet V2 B5 architectures. This study utilizes advanced deep-learning models for image analysis. ResNet, a significant advancement over AlexNet and VGGNet, uses a residual layer and global average pooling to process 224×224-pixel images efficiently. ResNeXt, extending ResNet, introduces "cardinality" with multiple transformation paths, balancing learning ability and computational efficiency [38]. MobileNetV2, an architecture designed specifically for mobile and embedded applications, incorporates depthwise separable convolutions across its 54 layers [39]. EfficientNet scales model dimensions using a compound variable and mobile inverted bottleneck convolution (MBConv) blocks, achieving optimal performance on the ImageNet dataset with its compact and fast-converging architecture [40].

2.4. Training and evaluation

Our research used a 70:20:10% data split ratio to develop and extensively analyze the optimal CNN model. Throughout the training phase, images were grouped into 'mini batches' of 16 due to memory constraints. We employed an Adam optimizer for optimization, adjusting CNN parameters with a 0.9 momentum [41] and a learning rate of 0.0001 [42]. L2 regularization with a factor of 0.0001 was applied to mitigate overfitting [43]. This study employed the Adam optimizer, chosen for its demonstrated effectiveness in practical scenarios as documented in prior research [44]. Following each training epoch, the model's advancement was evaluated using the validation set. The training process would terminate if no improvements were observed over the last five epochs or upon reaching 50 total epochs. The training set underwent shuffling at the beginning of each epoch to inject diversity. To evaluate the performance of different models, we use a range of metrics to examine the segmentation results in our study. Our assessment incorporates the Dice coefficient [45], Jaccard index (also known as intersection over union or IoU), pixel accuracy, precision, and recall [46], represented by (1)-(5).

$$Dice\ Coefficient = \frac{2\times TP}{(2\times TP)+FP+FN} \tag{1}$$

$$AP = \sum_{q=1}^{Q} \frac{AveP(q)}{q} \tag{2}$$

$$Pixel\ accuracy = \frac{Total\ Correctly\ Classified\ Pixels}{Total\ Number\ of\ Pixels} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

3. RESULTS AND DISCUSSION

This study presents an analysis of the performance of various models in recognizing phytoplankton images, with a focus on semantic segmenting phytoplankton. The research incorporates the U-Net model combined with four distinct encoders: EfficientNet-B5, MobileNetV2, ResNet50, and ResNeXt50. The evaluation employs several metrics, such as Dice, Jaccard, precision average (PA), precision (P), and recall (R), to gauge the models' effectiveness on both training and test datasets.

We found that the U-Net MobileNetV2 model, applying optical distortion transformation, achieved the highest scores on the test set for each evaluation metric: a Dice score of 93.69%, a Jaccard (IoU) score of 88.14%, precision at 99.89%, and recall at 100%. This performance marks a significant advancement compared to the training set scores. Other models, including U-Net+ResNet50 and U-Net + ResNeXt50, also demonstrated commendable performances (Table 2). However, the U-Net MobileNetV2 model notably excelled in all assessed metrics (Table 2). The complete top five models' performance metrics for test Dice, IoU, precision, and recall are displayed in Table 2.

Our findings show that U-Net MobileNetV2 with optical distortion excels in phytoplankton segmentation. Figure 2 shows the performance metrics of precision, recall, Dice value, IoU, pixel accuracy, and time. As shown by the effectiveness of precision and recall (Figure 2(a) and (b)). In contrast, the U-Net ResNet50 model with mix transform consistently demonstrated high Dice scores, outperforming MobileNetV2 models, which showed more significant score variability, indicating overfitting risks (Figure 2(b)). Figure 2(c) highlights the U-Net ResNet50 model's consistently high IoU performance, contrasting with the variability seen in models using elastic transform and grid distortion. Figure 2(d) reveals ResNet50's consistent pixel accuracy compared to MobileNetV2's fluctuating results. Figures 2(e) and (f) show the U-Net MobileNetV2 model maintaining training and testing times under 250 seconds, unaffected by different augmentations.

This study investigates a comprehensive set of U-Net models using multiple encoders and augmentations. However, additional research may be required to confirm the reliability of these models. In-depth studies are particularly needed regarding their performance on different types of phytoplankton images and in various environmental conditions.

Our research shows that the U-Net MobileNetV2 with optical distortion is more resilient and effective in segmenting phytoplankton than other models. Future research may look into more diverse augmentation strategies and practical methods for further enhancing segmentation accuracy. Exploring the use of attention mechanisms and vision transformers (ViT) with generative adversarial networks (GAN) to augment limited datasets could provide additional insights. Moreover, experimenting with the YOLOv8 backbone to create a new U-Net architecture is an exciting avenue for future investigation.

A tripartite arrangement linked to the semantic segmentation of the phytoplankton recognition procedure is depicted in Figures 3-5. Figure 3 showcases a collection of original phytoplankton input images of different sizes and shapes. The images' binary representations of the relevant phytoplankton regions are displayed in Figure 4 as input masks. Figure 5 shows the projected masks of the deep learning network, illustrating its attempt to isolate and highlight the phytoplankton patches. The model's ability to accurately recognize and separate phytoplankton from the image background is showcased by comparing the input and predicted masks.

Table 2. Top 5 models of test Dice, IoU, precision, and recall

No	Model Name	Performance Metrics			
		Dice (%)	IoU (%)	Precision (%)	Recall (%)
1	U-Net MobileNetV2 (Optical Distortion)	93.69	88.14	99.89	100.00
2	U-Net ResNet50 (Vertical Flip)	92.68	86.38	99.76	99.99
3	U-Net MobileNetV2 (Vertical Flip)	92.03	85.27	99.68	99.99
4	U-Net MobileNetV2 (Blur)	91.75	84.78	99.19	99.99
5	U-Net ResNet50 (Hue Saturation)	90.69	82.99	99.10	99.96

The results show that our architecture, which uses many encoders and augmentation methods in the U-Net model, can improve phytoplankton species segmentation. The U-Net MobileNetV2 (optical distortion) model has higher Dice (93.69%), IoU (88.14%), precision (99.89%), and recall (100%) values than previous models. Many microorganisms' species' segmentation performance improves with sophisticated methods. For example, the LCU-Net architectural model achieved a Dice score performance of 87.13%, a Jaccard index performance of 79.74%, a precision performance of 90.14%, and a recall performance of 87.12% with low processing resources [32]. The new dataset EMDS-5 uses U-Net for multi-object photo segmentation, showing Dice, Jaccard, and recall metrics of 85.24%, 77.41%, and 82.28% respectively [31]. These findings suggest we should examine U-Net design changes to increase segmentation performance. More research into pre-processing and augmentation strategies is needed. Using attention mechanisms [47] and ViT with GANs to augment limited datasets [32] and experimenting with the YOLOv8 backbone [48] to create a new U-Net architecture are promising directions for future research.

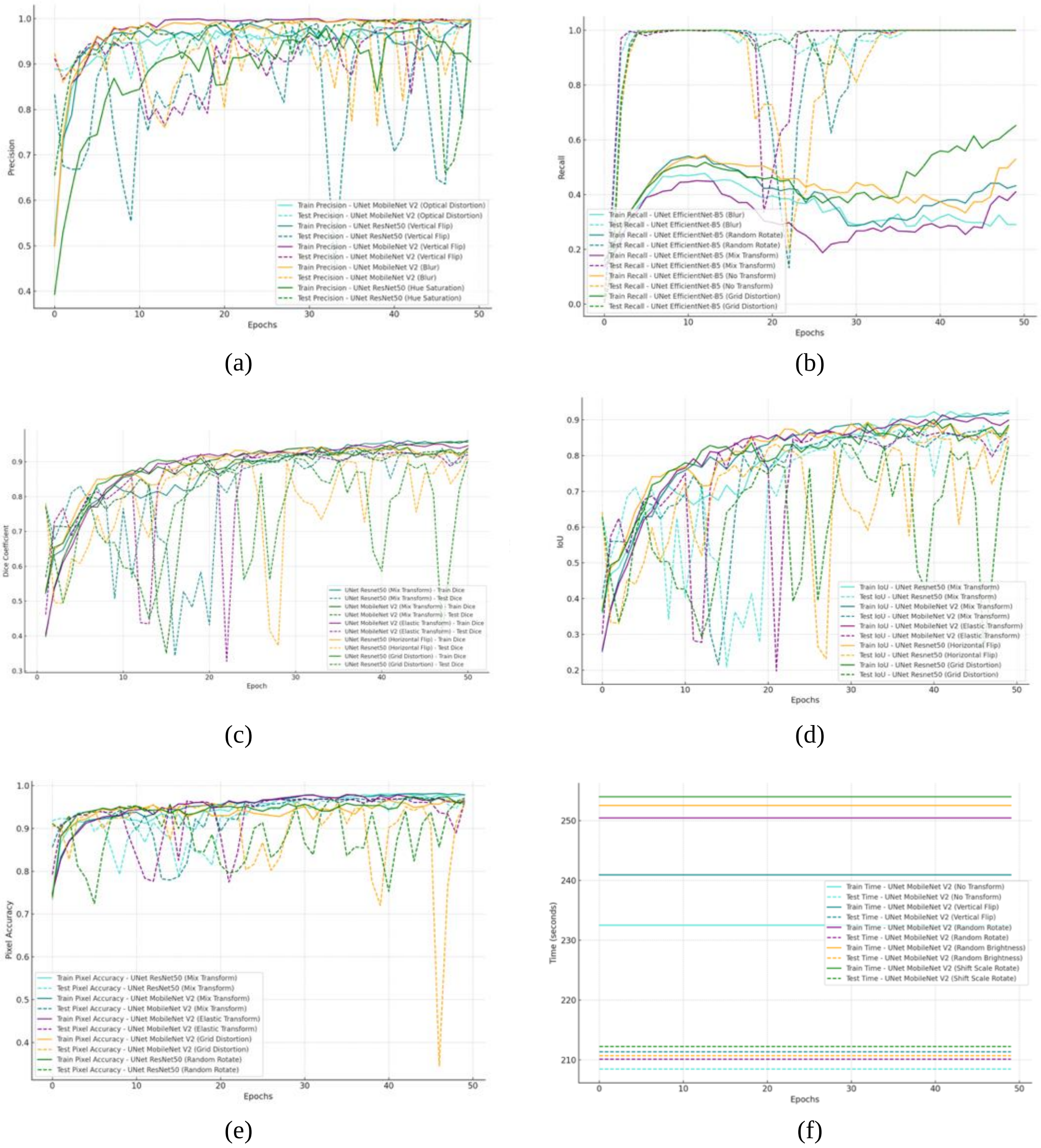

Figure 2. Metrics performance of (a) precision, (b) recall, (c) Dice score, (d) IoU, (e) pixel accuracy, and (f) time-consuming

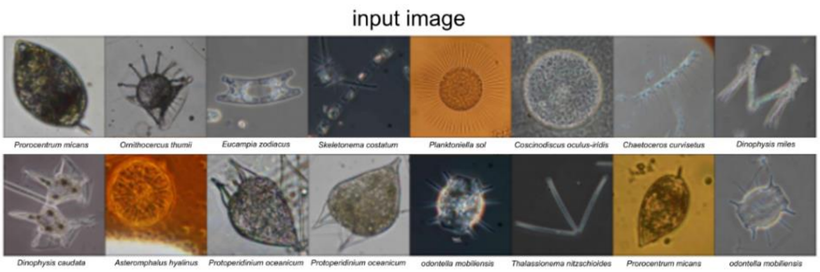

Figure 3. Result of phytoplankton semantic segmentation of the input image

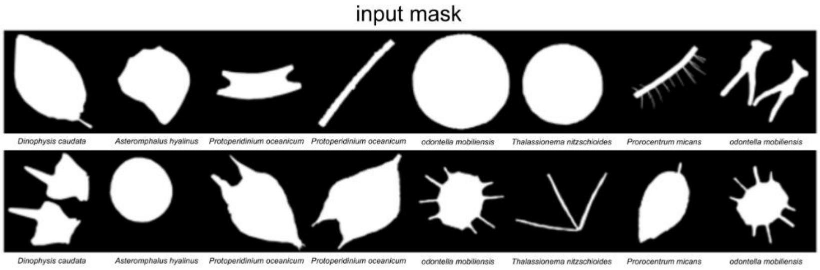

Figure 4. Result of phytoplankton semantic segmentation of the input mask

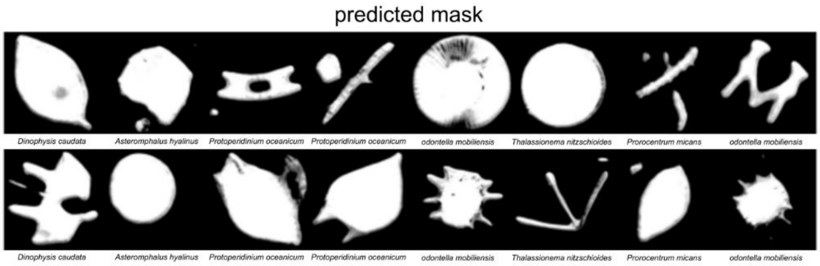

Figure 5. Result of phytoplankton semantic segmentation of the predicted mask

4. CONCLUSION

This study proposes enhancing the performance of semantic segmentation models by integrating multiple encoders and Adam optimizers into the U-Net architecture. Recent observations indicate that the pixel accuracy of the U-Net ResNet50 model with integrated transformations is consistently high. Our findings offer definitive proof that integrating MobileNetV2 with optical distortion attains a high level of precision, rather than being caused by increased quantities of training data. The outcomes of evaluating diverse U-Net models with several encoders and augmentations demonstrate exceptional accuracy in semantic segmentation performance. Properly selecting an encoder and augmentation strategy significantly impacts the performance of a model. Using EfficientNet-B5 with Blur augmentation demonstrates enhanced recall, thus augmenting the model's sensitivity. The experiments conclusively illustrate the model's exceptional efficiency and performance.

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, publication decision, or manuscript preparation.

REFERENCES

[1] G. R. Allen, “Conservation hotspots of biodiversity and endemism for indo-pacific coral reef fishes,” *Aquat. Conserv. Mar. Freshw. Ecosyst.*, vol. 18, no. 5, pp. 541–556, Jul. 2008, doi: 10.1002/aqc.880.

[2] A. Rachman, A. S. Suwarno, and S. Nurdjaman, “Application of deep (machine) learning for phytoplankton identification using microscopy images,” in *Proceedings of the 7th International Conference on Biological Science (ICBS 2021)*, Atlantis Press International B.V., 2022. doi: 10.2991/absr.k.220406.032.

[3] E. Prakasa, A. Rachman, D. R. Noerdjito, and R. Wardoyo, “Development of segmentation algorithm for determining planktonic objects from microscopic images,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 944, no. 1, p. 012025, Dec. 2021, doi: 10.1088/1755-1315/944/1/012025.

[4] A. Rachman, “Checklist and estimation of total number of phytoplankton species in Pari, Tidung, and Payung Islands, Indonesia,” *Biodiversitas J. Biol. Divers.*, vol. 21, no. 6, May 2020, doi: 10.13057/biodiv/d210616.

[5] S. Tang, A. Rachman, N. Fitria, H. Thoha, and B. Chen, “Phytoplankton changes during se monsoonal period in the Lembeh Strait of North Sulawesi, Indonesia, from 2012 to 2015,” *Acta Oceanol. Sin.*, vol. 37, no. 12, pp. 9–17, Dec. 2018, doi: 10.1007/s13131-018-1283-4.

[6] A. Rachman, A. Purwandana, and N. Fitriya, “Phytoplankton community structure of the Makassar Strait, Indonesia,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 789, no. 1, p. 012006, Jun. 2021, doi: 10.1088/1755-1315/789/1/012006.

[7] L. Tarafdar *et al.*, “Responses of phytoplankton community structure and association to variability in environmental drivers in a tropical coastal lagoon,” *Sci. Total Environ.*, vol. 783, p. 146873, Aug. 2021, doi: 10.1016/j.scitotenv.2021.146873.

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

[9] D. A. Hemraj, M. A. Hossain, Q. Ye, J. G. Qin, and S. C. Leterme, “Plankton bioindicators of environmental conditions in coastal lagoons,” *Estuar. Coast. Shelf Sci.*, vol. 184, pp. 102–114, Jan. 2017, doi: 10.1016/j.ecss.2016.10.045.

[10] T. R. Soeprobowati, N. L. Addadiyah, R. Hariyati, and J. Jumari, “Physico-chemical and biological water quality of warna and pengilon lakes, dieng, central java,” *J. Water L. Dev.*, vol. 51, no. X–XII, pp. 38–49, Jan. 2022, doi: 10.24425/jwld.2021.139013.

[11] Fakhrurrozi, T. R. Soeprobowati, and J. Jumari, “The water quality index and phytoplankton communities of Kokoh Putih River, Sembalun, East Lombok, Indonesia,” *AACL Bioflux*, vol. 15, no. 4, 2022, [Online]. Available: http://bioflux.com.ro/docs/2022.2025-2040.pdf

[12] G. M. Hallegraeff *et al.*, “Perceived global increase in algal blooms is attributable to intensified monitoring and emerging bloom impacts,” *Commun. Earth Environ.*, vol. 2, no. 1, p. 117, Jun. 2021, doi: 10.1038/s43247-021-00178-8.

[13] M. L. Wells *et al.*, “Future hab science: directions and challenges in a changing climate,” *Harmful Algae*, vol. 91, p. 101632, Jan. 2020, doi: 10.1016/j.hal.2019.101632.

[14] S. H. Samudra *et al.*, “The phenomenon of harmful algae blooms (habs) based on literature study in Indonesia sea waters from 2005-2021,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1251, no. 1, p. 012044, Oct. 2023, doi: 10.1088/1755-1315/1251/1/012044.

[15] O. D. W. Ardhi, T. R. Soeprobowati, K. Adi, E. Prakasa, and A. Rachman, “Deep learning methods for plankton identification: a bibliometric analysis and general review,” in *2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering (APICS)*, IEEE, Aug. 2022, pp. 96–101. doi: 10.1109/APICS56469.2022.9918707.

[16] T. Eerola *et al.*, “Survey of automatic plankton image recognition: challenges, existing solutions and future perspectives,” *Artif. Intell. Rev.*, vol. 57, no. 5, 2024, doi: 10.1007/s10462-024-10745-y.

[17] C. Sánchez, N. Vállez, G. Bueno, and G. Cristóbal, “Diatom classification including morphological adaptations using CNNs,” in *Lecture Notes in Computer Science*, Springer, Cham, 2019, pp. 317–328. doi: 10.1007/978-3-030-31332-6_28.

[18] H. A. Al-Barazanchi, A. Verma, and S. Wang, “Performance evaluation of hybrid CNN for sipper plankton image calssification,” in *2015 Third International Conference on Image Information Processing (ICIIP)*, IEEE, Dec. 2015, pp. 551–556. doi: 10.1109/ICIIP.2015.7460262.

[19] J. Dai, R. Wang, H. Zheng, G. Ji, and X. Qiao, “ZooplanktoNet: deep convolutional network for zooplankton classification,” in *OCEANS 2016 - Shanghai*, IEEE, Apr. 2016, pp. 1–6. doi: 10.1109/OCEANSAP.2016.7485680.

[20] I. Correa, P. Drews, S. Botelho, M. S. de Souza, and V. M. Tavano, “Deep learning for microalgae classification,” in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, Dec. 2017, pp. 20–25. doi: 10.1109/ICMLA.2017.0-183.

[21] K. Kraft *et al.*, “Towards operational phytoplankton recognition with automated high-throughput imaging, near-real-time data processing, and convolutional neural networks,” *Front. Mar. Sci.*, vol. 9, Sep. 2022, doi: 10.3389/fmars.2022.867695.

[22] J. S. Ellen, C. A. Graff, and M. D. Ohman, “Improving plankton image classification using context metadata,” *Limnol. Oceanogr. Methods*, vol. 17, no. 8, pp. 439–461, Aug. 2019, doi: 10.1002/lom3.10324.

[23] J. Raitoharju *et al.*, “Data enrichment in fine-grained classification of aquatic macroinvertebrates,” in *2016 ICPR 2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI)*, IEEE, Dec. 2016, pp. 43–48. doi: 10.1109/CVAUI.2016.020.

[24] A. S. Keçeli, A. Kaya, and S. U. Keçeli, “Classification of radiolarian images with hand-crafted and deep features,” *Comput. Geosci.*, vol. 109, pp. 67–74, Dec. 2017, doi: 10.1016/j.cageo.2017.08.011.

[25] 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, p. e0219570, Jul. 2019, doi: 10.1371/journal.pone.0219570.

[26] W. Ma *et al.*, “Super-resolution for in situ plankton images,” in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, IEEE, Oct. 2021, pp. 3676–3685. doi: 10.1109/ICCVW54120.2021.00411.

[27] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: convolutional networks for biomedical image segmentation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[28] S. Lee, M. Negishi, H. Urakubo, H. Kasai, and S. Ishii, “MU-Net: multi-scale U-Net for two-photon microscopy image denoising and restoration,” *Neural Networks*, vol. 125, pp. 92–103, May 2020, doi: 10.1016/j.neunet.2020.01.026.

[29] D. Li *et al.*, “Robust blood cell image segmentation method based on neural ordinary differential equations,” *Comput. Math. Methods Med.*, vol. 2021, pp. 1–11, Aug. 2021, doi: 10.1155/2021/5590180.

[30] R. Azad *et al.*, “Medical image segmentation review: the success of U-Net,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–20, 2024, doi: 10.1109/TPAMI.2024.3435571.

[31] Z. Li *et al.*, “EMDS-5: environmental microorganism image dataset fifth version for multiple image analysis tasks,” *PLoS One*, vol. 16, no. 5, p. e0250631, May 2021, doi: 10.1371/journal.pone.0250631.

[32] J. Zhang *et al.*, “LCU-net: a novel low-cost U-Net for environmental microorganism image segmentation,” *Pattern Recognit.*, vol. 115, p. 107885, Jul. 2021, doi: 10.1016/j.patcog.2021.107885.

[33] H. Yang *et al.*, “Comparative study for patch-level and pixel-level segmentation of deep learning methods on transparent images of environmental microorganisms: from convolutional neural networks to visual transformers,” *Appl. Sci.*, vol. 12, no. 18, p. 9321, Sep. 2022, doi: 10.3390/app12189321.

[34] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, “LabelMe: a database and web-based tool for image annotation,” *Int. J. Comput. Vis.*, vol. 77, no. 1–3, pp. 157–173, May 2008, doi: 10.1007/s11263-007-0090-8.

[35] 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, p. 125, Feb. 2020, doi: 10.3390/info11020125.

[36] E. Andersson and R. Berglund, “Evaluation of data augmentation of mr images for deep learning,” 2018. [Online]. Available: https://lup.lub.lu.se/luur/download?fileOId=8952748&func=downloadFile&recordOId=8952747

[37] A. Lumini and L. Nanni, “Deep learning and transfer learning features for plankton classification,” *Ecol. Inform.*, vol. 51, pp. 33–43, May 2019, doi: 10.1016/j.ecoinf.2019.02.007.

[38] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 5987–5995. doi: 10.1109/CVPR.2017.634.

[39] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: inverted residuals and linear bottlenecks,” in

Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.
[40] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," *arXiv*, May 2019, [Online]. Available: http://arxiv.org/abs/1905.11946
[41] M. O. Lawal, "Tomato detection based on modified YOLOv3 framework," *Sci. Rep.*, vol. 11, no. 1, p. 1447, Jan. 2021, doi: 10.1038/s41598-021-81216-5.
[42] Y. Li, C. Wei, and T. Ma, "Towards explaining the regularization effect of initial large learning rate in training neural networks," *arXiv*, Jul. 2019, doi: 10.48550/arXiv.1907.04595.
[43] Z. Hao, Y. Jiang, H. Yu, and H.-D. Chiang, "Adaptive learning rate and momentum for training deep neural networks," in *Lecture Notes in Computer Science*, Springer, Cham, 2021, pp. 381–396. doi: 10.1007/978-3-030-86523-8_23.
[44] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *arXiv*, 2017, doi: 10.48550/arXiv.1412.6980.
[45] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945, doi: 10.2307/1932409.
[46] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Res. Notes*, vol. 15, no. 1, p. 210, Dec. 2022, doi: 10.1186/s13104-022-06096-y.
[47] A. Kumthekar and G. R. Reddy, "Redesigning u-net with dense connection and attention module for satellite based cloud detection," *IAES Int. J. Artif. Intell.*, vol. 11, no. 2, p. 699, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp699-708.
[48] J. Terven and D. Cordova-Esparza, "A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-nas," *arXiv*, 2024, doi: 10.48550/arXiv.2304.00501.

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, Universitas Gadjah Mada, which he completed in 2012. He is pursuing a PhD 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. Contact him via email at 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, Science and Technology. For correspondence, Contact her via email at 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 reached for correspondence via email at 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. After obtaining a Master of Engineering (M. Eng) in Electrical Engineering in 2001, he continued his academic pursuits at the same institution. The Malaysian institution Universiti Teknologi PETRONAS awarded him a Doctor of Philosophy (PhD) in Electrical and Electronic Engineering in 2014. He is a senior researcher at the 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. Please email him at esap001@brin.go.id with any inquiries or correspondence.

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, pa Convolutional Networks for Biomedical Image Segmentationrticularly 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. Kindly address all inquiries and comments to ariefrachman1987@yahoo.com.