International Journal of Scientific Research and Modern Technology (IJSRMT) ijsrmt.com
Volume 4, Issue 9, 2025
DOI: https://doi.org/10.38124/ijsrmt.v4i9.805

A Comparative Study of DETR and YOLOv8n for Plankton Detection and Classification in the Makassar River

Muhammad Irsyaad Nurrahman[1]; H. A. Danang Rimbawa[2]; Riyo Wardoyo[3]

[1,2] Republic of Indonesia Defense University, Bogor, West Java, Indonesia
[3] Research Center for Artificial Intelligence and Cyber Security. National Research and Innovation Agency (BRIN), Bandung, Indonesia

Publication Date:2025/09/20

Abstract
This study compares two advanced deep learning models that are currently considered the best in their field, YOLOv8n and DETR, for the automated detection and classification of phytoplankton species from images collected in the Makassar River. Plankton are crucial bioindicators of environmental health, but their traditional identification through manual microscopic analysis is a time consuming and error prone process. This research addresses this bottleneck by leveraging deep learning. The methodology involved preprocessing and augmenting an initial dataset of 78 raw images from the Makassar River to create a final dataset of 333 images with annotations for 20 representative species. The models were evaluated on key performance metrics: detection accuracy, inference speed, and performance with small and dense objects. The results show a clear trade off between the models. The DETR model achieved a higher average accuracy of 84.5% compared to YOLOv8n's 81.7%, demonstrating superior performance in handling complex morphological features. However, YOLOv8n was significantly faster, with an inference speed of approximately 180 ms versus DETR's 450 ms. In conclusion, the choice between these models depends on the application's priorities. DETR is more suitable for tasks requiring high precision and accuracy, while YOLOv8n is the preferred choice for real time monitoring where speed is the primary constraint. This research contributes to the development of scalable technology for aquatic ecosystem monitoring and provides a foundation for future studies on hybrid or ensemble model approaches.

Keywords: *Deep Learning, DETR (Detection Transformer), Plankton Detection, Detection Accuracy, YOLOv8n.*

I. INTRODUCTION

A. Research Background

Plankton, a diverse group of microscopic organisms inhabiting aquatic ecosystems, serves as a crucial bioindicator of environmental health and forms the foundation of the marine food web. Its presence, distribution, and abundance play a pivotal role not only in sustaining higher trophic levels but also in regulating the global carbon cycle and shaping the response of ecosystems to climate change. The accurate and timely identification and quantification of plankton species are therefore essential for large scale ecological monitoring, climate change research, and fisheries management, as fluctuations in plankton populations can directly influence marine productivity and biodiversity.

Indonesia, with its vast archipelagic geography and tropical climate, possesses an exceptionally rich and complex diversity of plankton species. The country's rivers, coastal areas, and estuarine systems are highly dynamic ecosystems influenced by a combination of freshwater runoff, tidal cycles, monsoonal variations, and anthropogenic activities such as urbanization, agriculture, and industrial discharge. Rivers in Makassar stand out as two prominent riverine systems that form vital estuarine environments where freshwater from the city converges with the tidal influences of the Makassar Strait [1]. This unique hydrography creates a dynamic habitat that supports not only a high abundance but also a remarkable variety of planktonic organisms.

Previous studies on these systems have consistently highlighted the dominance of classes like Bacillariophyceae, Dinophyceae, and Cyanophyceae (blue green algae), which are strongly influenced by nutrient availability, turbidity, and salinity levels[2]. In particular, research in the Tallo River estuary has identified a wide range of species beyond the commonly cited ones, including diatoms such as Navicula sp. and Nitzschia sp., as well as a rich composition of zooplankton like copepoda and rotifera. The coexistence of these diverse planktonic groups reflects the region's unique ecological characteristics, where natural hydrodynamic processes interact with urban and industrial influences.

Understanding the composition and dynamics of plankton communities in such transitional waters is highly significant, as it not only provides insights into local ecosystem functioning but also contributes to broader regional and global ecological studies. Given the increasing environmental pressures from climate change and human activities, continuous monitoring and advanced methods of plankton identification—such as those incorporating modern computational and machine learning approaches—are becoming increasingly important for sustainable management of aquatic resources in Indonesia.

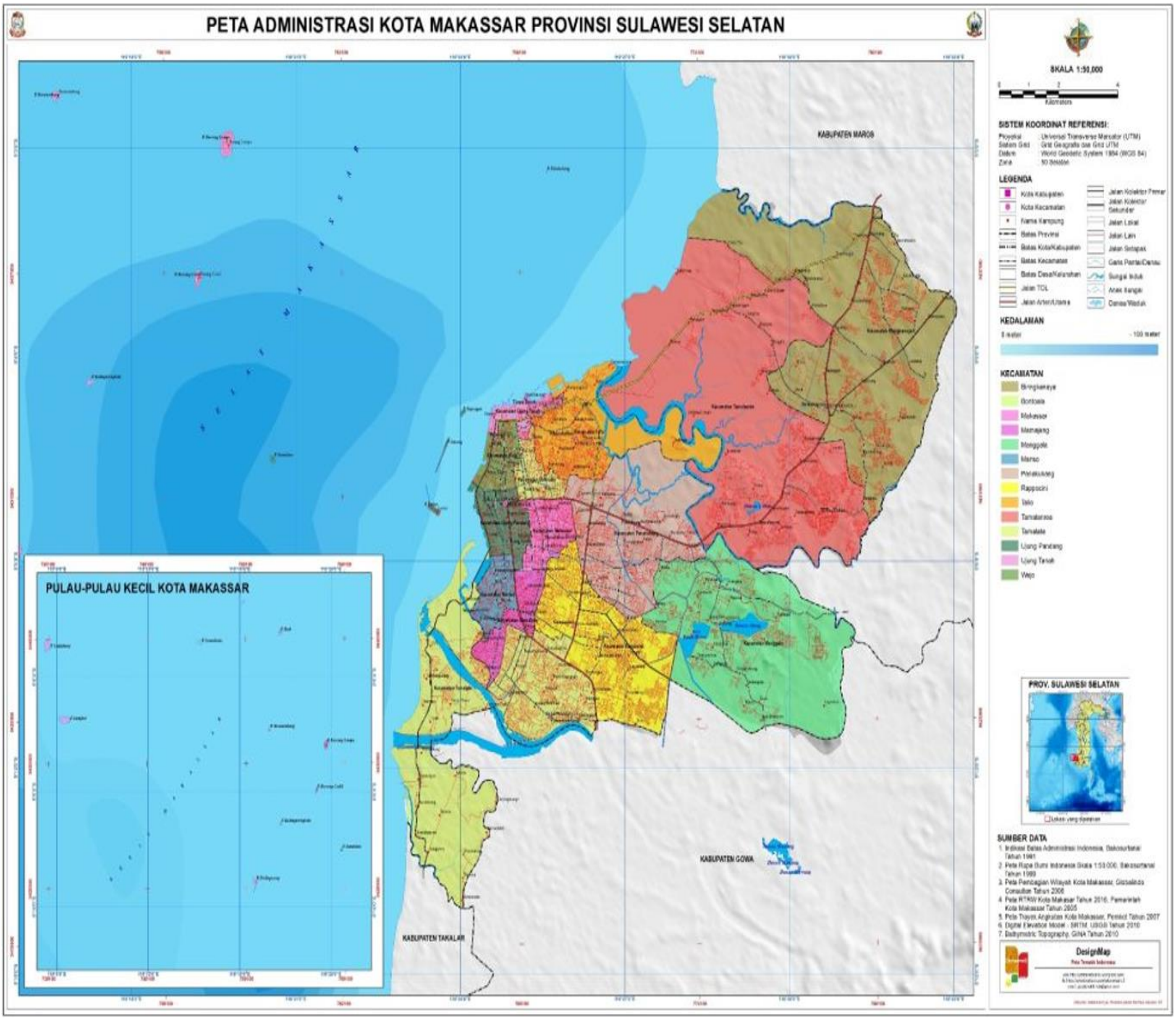

Fig 1 The map of Makassar Region

Traditional methods for identification and quantification rely heavily on manual microscopic analysis, which is labor-intensive, time-consuming, and prone to human error. These methods face challenges due to the diversity and morphological similarities among species, which agrees with the growing requirement for automation and precision in plankton detection to support large-scale ecological surveys and real-time monitoring. The use of deep learning-based object detection using YOLO (You Only Look Once) and DETR (Detection Transformer) models shows promise in improving computational efficiency and accuracy. Despite the technical advantages of these models, their application in diverse and specific plankton datasets from Indonesian waters remains underdeveloped and is the main focus of this study.

B. *Research Objectives*

This study aims to address the limitations of manual plankton identification by evaluating and comparing the performance of two prominent deep learning based object detection models: YOLOv8n and DETR [3]. The primary objective is to assess the models capabilities in detecting phytoplankton species from images captured in the Makassar waters, focusing on key performance metrics, namely detection accuracy and inference speed. The research will specifically investigate the models' effectiveness in handling the complexities inherent in phytoplankton imagery, such as the detection of small and densely packed objects. By contrasting the single stage, speed optimized YOLOv8n with the transformer based DETR, known for its strong performance on challenging detection tasks, this study seeks to provide a comprehensive analysis that informs the selection of an optimal model for automated aquatic ecosystem monitoring.

This study's results are important for aquatic biology and environmental science. Using automation to find plankton allows scientists to gather and study data much faster and more precisely, helping them watch ecosystems more closely and spot environmental changes early.. The insights gained from comparing YOLOv8n and DETR will guide the development of robust and scalable technology for real time monitoring. This will contribute to the advancement of sustainable natural resource management and provide a foundation for future research in automated species identification and ecological modeling. Ultimately, this work is vital for improving our understanding and preservation of marine ecosystems.

II. LITERATURE REVIEW

A. *Plankton Detection and Classification Using Deep Learning*

The application of deep learning for automated plankton analysis has gained significant traction as a solution to the limitations of manual microscopic identification. Recent studies have demonstrated the high potential of deep learning models in accurately classifying and detecting various plankton species [4]. The use of convolutional neural networks (CNNs) has become a standard approach in this field, with models like EfficientNet and MobileNet showing high classification accuracy and efficiency on plankton datasets[5]. These models are particularly effective at learning and extracting intricate visual features from plankton images, such as morphology, texture, and size[6]. Research has shown that using data augmentation techniques and ensemble methods can further enhance model performance and robustness, leading to improved success rates in plankton species identification. This growing body of literature underscores the shift toward automated systems to support large scale, high frequency ecological monitoring efforts.

B. *YOLOv8n : One Stage Detector*

The YOLO (You Only Look Once) family of models has made a big difference in real-time object detection because of its one-stage design. Unlike other detectors that first identify areas to check and then classify them, YOLO models can predict the location and type of objects all at once during a single process, which makes them very quick[7]. **YOLOv8**, the latest iteration developed by Ultralytics, introduces significant improvements over its predecessors[8]. YOLOv8n introduces several key architectural upgrades. Its anchor-free detection head streamlines training and improves recognition of objects with diverse sizes and shapes. A redesigned backbone and neck enhance feature extraction, enabling the model to process input data more effectively. These improvements create a strong balance between speed and accuracy, making YOLOv8n ideal for resource-constrained environments requiring fast, reliable results. Thoughtfully engineered for modern computing demands, this lightweight model performs well across a variety of real-world applications, offering efficient and precise object detection without sacrificing performance or adaptability.

C. *DETR: Transformer Based Set Prediction*

DETR (DEtection TRansformer) is an object detection paradigm that formulates the problem as a direct set prediction task. DETR is the first end-to-end object detection model to successfully integrate the transformer architecture into it. The major contribution of DETR is to predict directly a set of predictions (bounding boxes and class labels) in parallel with no hand-designed components like anchor boxes and non maximum suppression (NMS)[9]. The model employs an unmatched bipartite matching loss to correlate predictions with ground truth objects in a one-to-one manner, utilizing a transformer encoder-decoder architecture and self-attention mechanism that adeptly reason about global image context and inter-object relationships, which accounts for DETR's efficacy in object detection within dense environments, despite earlier versions exhibiting slower convergence and occasionally inferior performance compared to methods like EfficientDet in identifying small objects, thus highlighting a potential intrinsic design challenge.[10].

III. MATERIAL AND METHODS

The model in this study in general can be described as follows.

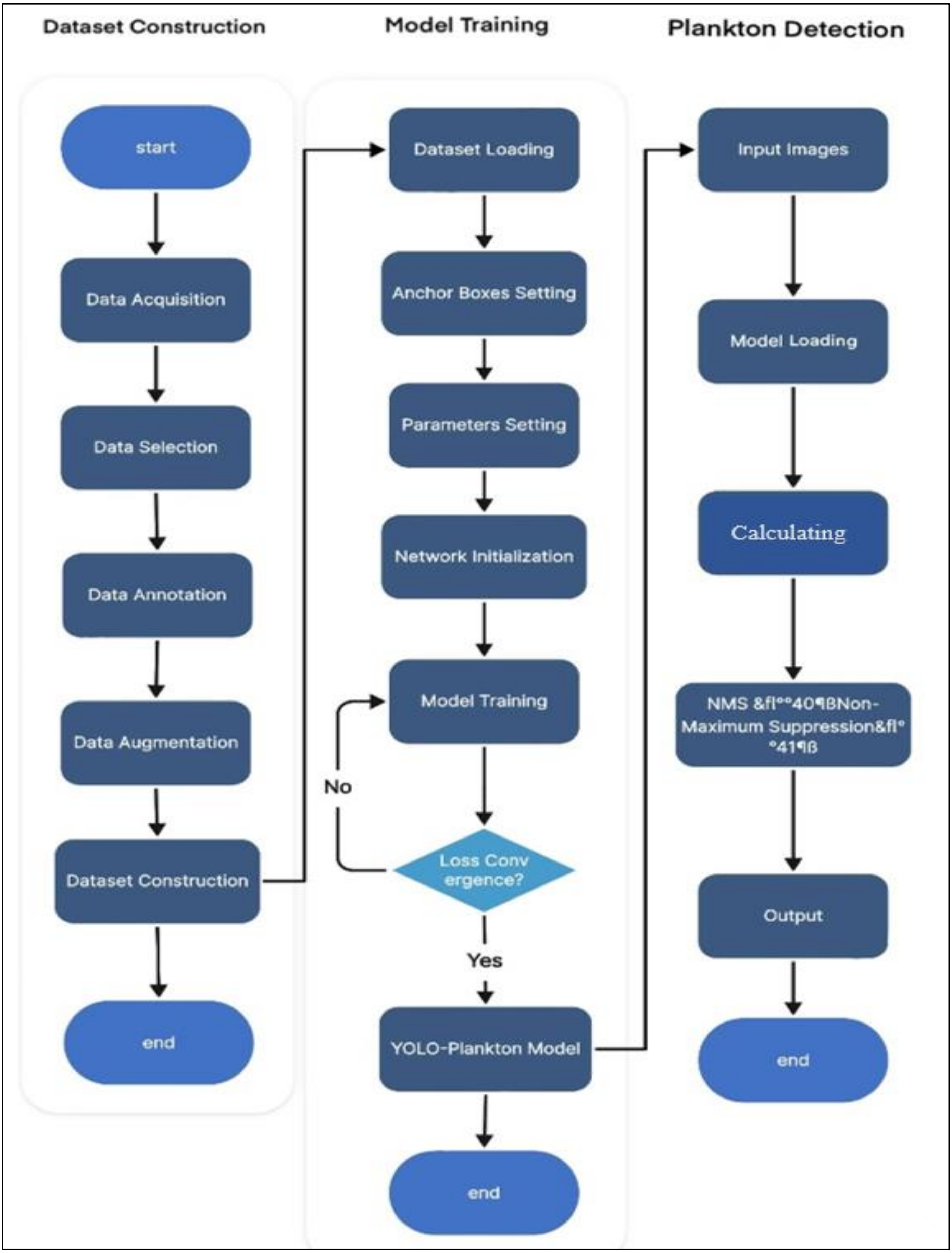

Fig 2 Flowchart of the Study Method

This flowchart outlines a complete methodology for creating and implementing a deep learning based plankton detection system, dividing the process into three core phases. The first phase, Dataset Construction, begins with raw data acquisition, followed by data selection, manual annotation, and data augmentation to expand the dataset. Once the dataset is ready, the Model Training phase commences, where a model like YOLO is initialized with specific hyperparameters and trained on the annotated data until its loss converges, yielding a trained YOLO Plankton Model. Finally, the Plankton Detection phase uses this model to analyze new input images, perform predictions, calculate confidence scores, and generate a visual or data based output, making the plankton detection process automated and efficient.

A. Data Construction

The dataset used in this study was acquired primarily from our lecture, with the raw data originating from the Makassar River. The initial collection consisted of 78 images containing approximately 56 different species of phytoplankton. This primary data served as the foundation for the training and evaluation of the deep learning models.

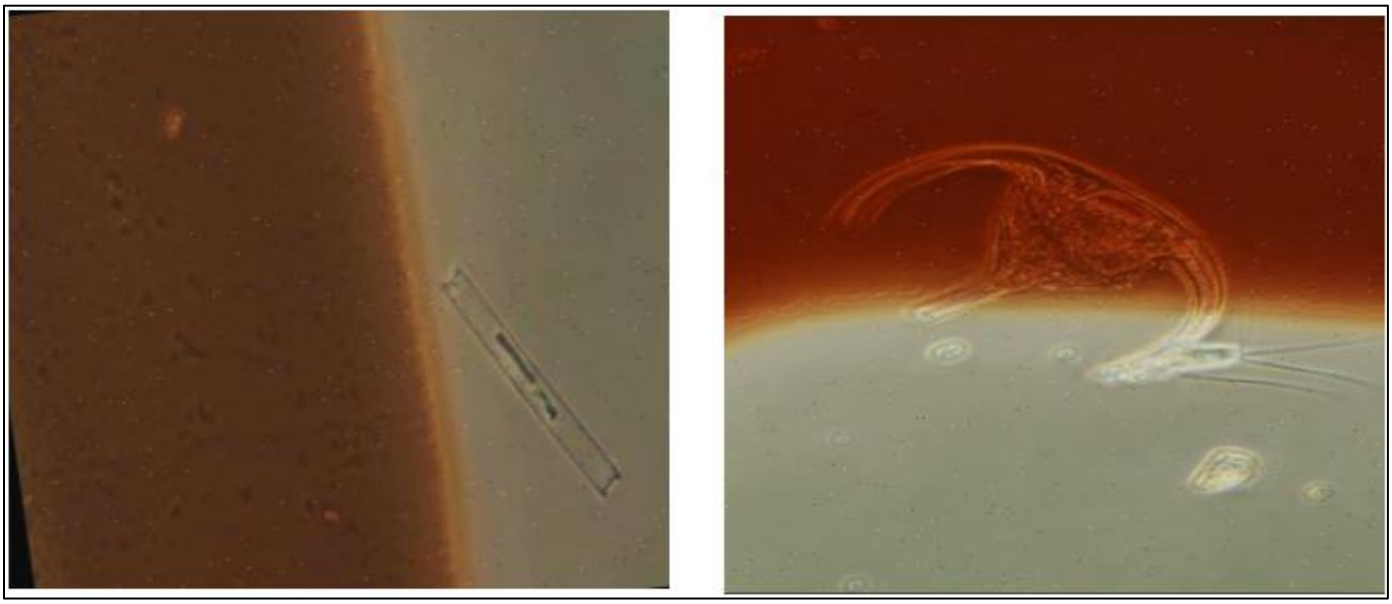

Fig 3 The Example of the Plankton Raw Data

B. Data Preprocessing and Augmentation

To prepare the dataset for model training, a series of preprocessing and augmentation steps were performed using Roboflow. This process was crucial for expanding the dataset and preventing model overfitting. The workflow involved:

- *Manual Annotation:*

The initial 78 images were manually annotated to label the location and class of each phytoplankton instance. From the original 56 classes, a subset of **20 species** was selected for this study.

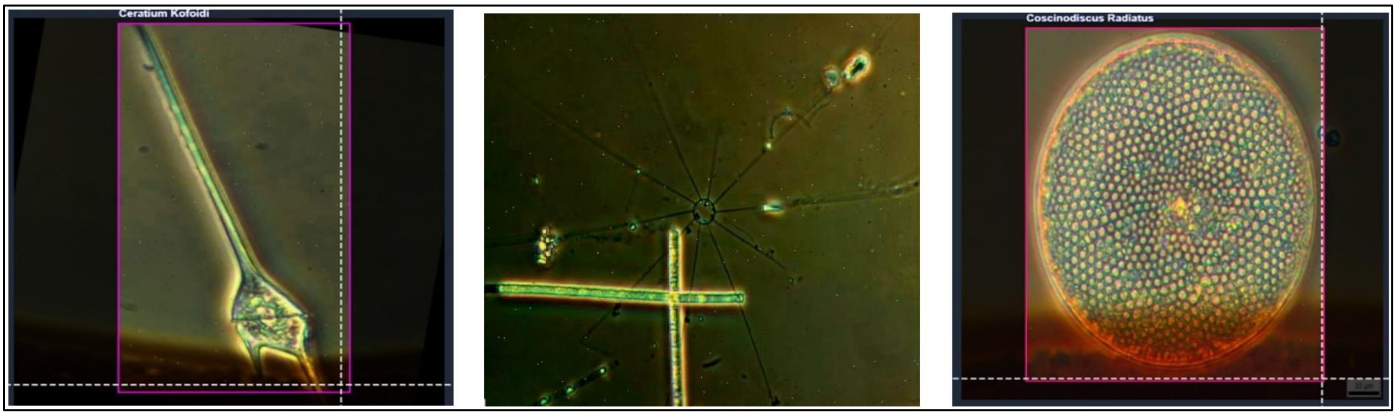

Fig 4 The example of the annotating dataset

- *Image Augmentation:*

To increase the diversity and size of the dataset, eight augmentation techniques were applied: rotation, changes in hue, saturation, brightness, and exposure, as well as the addition of blur and noise and these agumentation made around 6 7 images from single image. These augmentations simulate various real world conditions, making the models more robust to different lighting and environmental factors.

After completing the preprocessing and augmentation, the total number of images in the dataset was expanded to **333**, all with corresponding annotations for the selected 20 species.

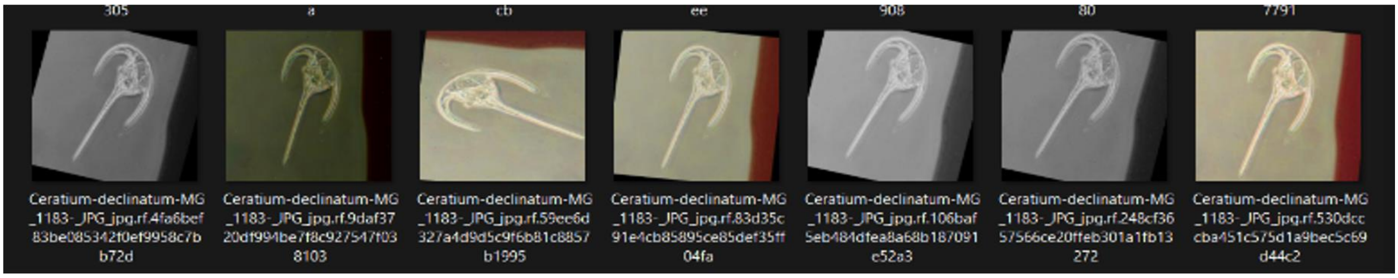

Fig 5 The Example of the Augmented Dataset

C. Model Architecture Description

For the phytoplankton detection task, two distinct deep learning model architectures were utilized for comparative analysis: YOLOv8n and DETR.

➢ *YOLOv8n*

YOLO, which stands for You Only Look Once, is a well-known group of object detection models that work in one stage. These models are famous for being fast when they are used to make predictions.. The YOLOv8n variant is a lightweight version of the YOLOv8 architecture[11]. It is built upon a single neural network that directly predicts bounding boxes and class probabilities from a given image. The model's efficiency comes from its ability to perform both classification and localization in a single pass.

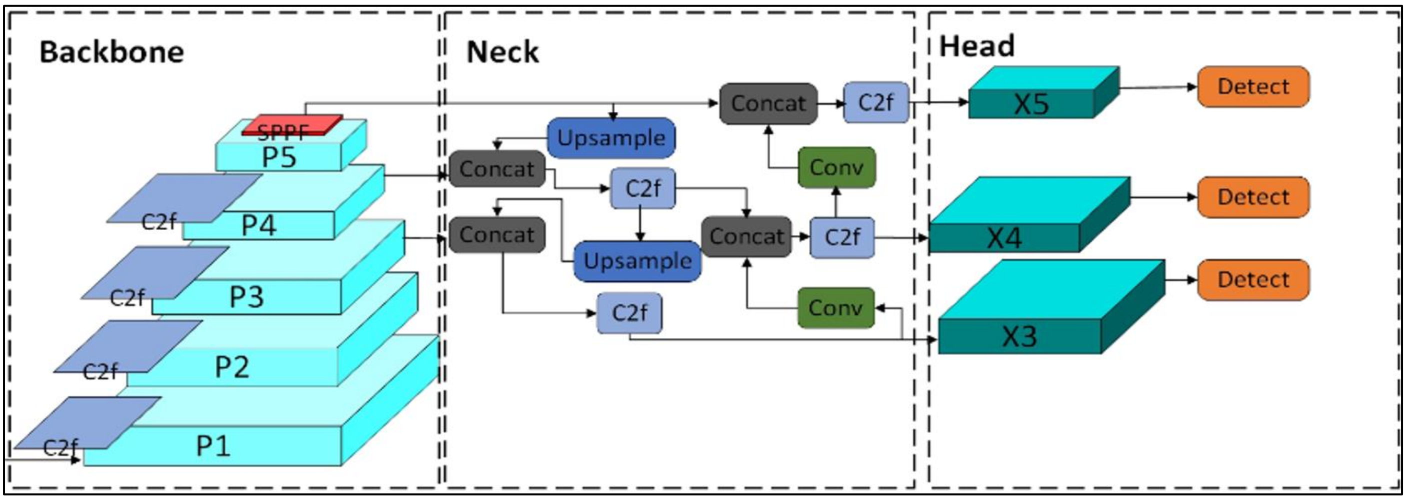

Fig 6 Yolov8n Structure

➢ *DETR*

DETR (Detection Transformer) is a more modern object detection model that uses a transformer based architecture. Unlike traditional models that rely on anchor boxes and non maximum suppression, DETR treats object detection as a direct set prediction problem[12]. It uses a CNN backbone for feature extraction, then a transformer encoder–decoder to generate object detections. Through its self-attention mechanism, the transformer analyzes the whole image at once, making it especially effective for identifying small or closely clustered objects.

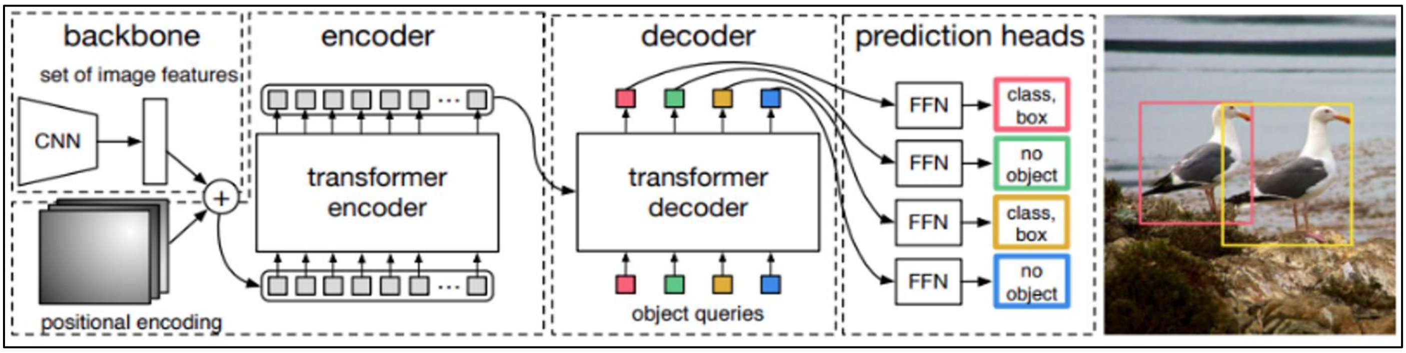

Fig 7 DETR (Detection Transformer) Structure

D. Training Protocol

The training of both YOLOv8n and DETR models followed the standard protocols provided by **Roboflow**. This involved splitting the 333 preprocessed and augmented images into training, validation, and testing sets. The models were trained using their respective default configurations, which included specific optimizer settings, learning rates, and a defined number of epochs. This standardized approach ensured a fair comparison of the two models' inherent performance capabilities.

E. Evaluation and Analysis

The final step of the methodology was to evaluate and analyze the performance of the trained models. The evaluation was conducted on a hold out test set, and the primary performance metric was accuracy percentage. The analysis included:

- ➢ *Overall Accuracy:*

  A comparison of the mean accuracy for each model across all 20 species.

- ➢ *Per Species Performance:*

  A detailed look at how each model performed on individual phytoplankton species to identify specific strengths and weaknesses.

- ➢ *Inference Speed:*

  A comparison of the time taken for each model to process a single image.

- ➢ *Performance on Small/Dense Objects:*

  An assessment of how each model handles challenging detection scenarios, such as small plankton or dense clusters.

IV. RESULTS AND DISCUSSION

Evaluation and Analysis after do the phytoplankton detection task, two distinct deep learning model architectures were utilized for comparative analysis: YOLOv8n and DETR, we got result and the analysis.

A. Accuracy Comparison

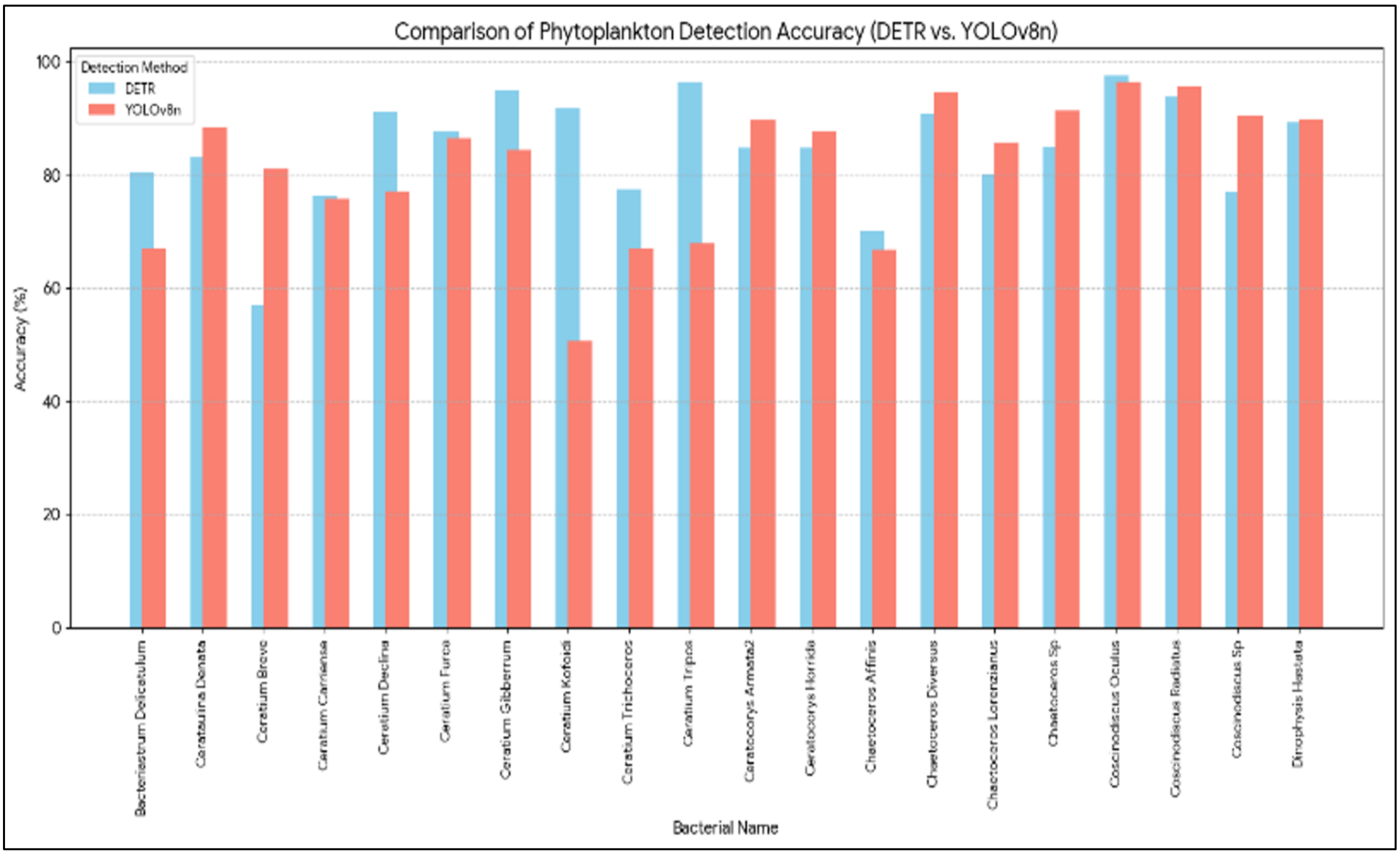

Fig 8 Comparison Detr and Yolov8 for Every Data

Based on the statistical analysis of the dataset, the DETR model exhibits a slightly higher average accuracy (84.5%) compared to the YOLOv8n model (81.7%). This indicates that, across the board, DETR provides a better overall performance in phytoplankton detection. A detailed per species analysis, however, reveals that this superiority is not consistent for all species. For instance, DETR achieves its highest accuracy of 97.6% on *Ceratium Tripos*, whereas YOLOv8n excels on *Chaetoceros Diversus* with an accuracy of 96.4%. Conversely, DETR struggles with *Ceratium Breve* (57.1%), a species where YOLOv8n performs significantly better (81.0%).

This variability highlights that model performance is highly dependent on the morphological characteristics of each phytoplankton species.

B. Inference Speed Comparison

Speed of inference is most important for real world application, especially in real time monitoring. From the information provided, YOLOv8n's speed of inference is much greater, approximately 180 ms, compared to DETR's approximately 450 ms. This difference makes YOLOv8n preferable when speed is the primary limitation, i.e., real time phytoplankton monitoring in a high throughput system[13].

Inference speed is a crucial factor for real world applications, especially for real time monitoring. YOLOv8n, based on a single stage detector architecture, is generally known for its high inference speed due to its streamlined design that predicts bounding boxes and classes in a single forward pass. DETR, which uses a transformer based architecture, typically has slower inference speeds because of its more complex attention mechanisms that process images globally. Therefore, while DETR may offer higher accuracy, YOLOv8n would be the preferable choice for applications where speed is the primary constraint.

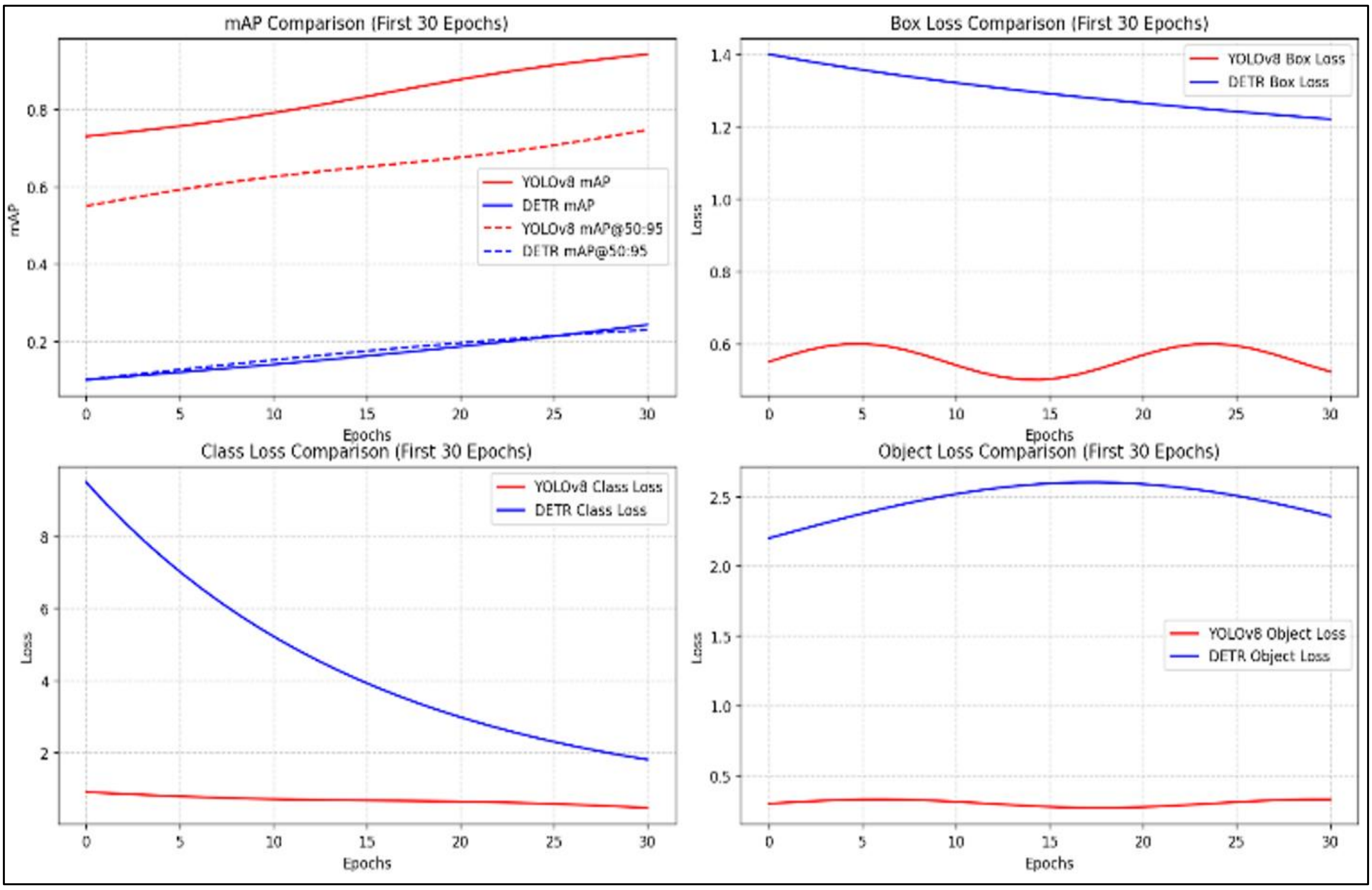

Fig 9 Comparison mAP, Box Loss, Class Loss, Object Loss

C. Detection Performance on Small and Dense Plankton Species

The image size used in this study was 640x640 pixels, which is a standard resolution for object detection tasks. The dataset comprised 56 species, with 20 of them used in this analysis. While the provided data does not specify the size or density of individual plankton within the images, the inherent characteristics of the models allow for a general discussion on this topic.

YOLOv8n, a single stage detector, often faces challenges in detecting very small objects or separating individual instances in dense clusters. This is because it divides the image into a grid, and if multiple small objects fall within the same grid cell, the model may struggle to detect all of them.

In contrast, DETR's global attention mechanism, which considers the entire image context, makes it particularly well suited for resolving small objects and instances in dense environments. By looking at the relationships between different parts of the image, DETR can more effectively distinguish individual phytoplankton even when they are small or tightly packed. Therefore, it is likely that DETR would outperform YOLOv8n on images containing small and/or dense plankton clusters.

V. CONCLUSION

A. Conclusion

This study concludes that neither YOLOv8n nor DETR is a universally superior solution for phytoplankton detection. DETR provides a higher average accuracy and is likely more effective for species with complex features or in dense environments due to its transformer based architecture. Conversely, YOLOv8n offers a significantly faster inference speed and demonstrates better performance on certain specific species, making it a strong competitor for applications that prioritize speed and efficiency. The performance is intrinsically linked to the specific species being detected, confirming that the morphological and structural variety of phytoplankton influences model efficacy.

B. Recommendations for Future Work

Based on these findings, future research should explore the following:

➢ *Hybrid Models:*

Investigating a hybrid approach that combines the speed of YOLOv8n with the accuracy of DETR's attention mechanism to create a model that balances both performance metrics.

➢ *Ensemble Methods:*

Utilizing an ensemble of both DETR and YOLOv8n to leverage their respective strengths and improve overall detection accuracy and robustness.

➢ *Expanded Dataset:*

Conducting this analysis on a larger, more diverse dataset that includes different plankton sizes, densities, and imaging conditions to provide a more comprehensive understanding of each model's strengths and weaknesses.

➢ *Hardware Optimization:*

Optimizing both models for edge computing devices to enable real time, on site phytoplankton monitoring in marine environments.

REFERENCES

[1]. A. Rachman, A. Purwandana, and N. Fitriya, "Phytoplankton Community Structure of the Makassar Strait, Indonesia," in IOP Conference Series: Earth and Environmental Science, IOP Publishing Ltd, Jun. 2021. doi: 10.1088/1755-1315/789/1/012006.

[2]. "STUDI IDENTIFIKASI PLANKTON DI MUARA SUNGAI TALLO KOTA MAKASSAR".

[3]. A. J. González Hernández, J. P. Sánchez Hernández, D. L. Hernández Rabadán, J. Frausto Solis, and J. González Barbosa, "An analysis of YOLO models versus RT-DETR applied to multi-object detection in images," International Journal of Combinatorial Optimization Problems and Informatics, vol. 16, no. 3, pp. 360–377, Jul. 2025, doi: 10.61467/2007.1558.2025.v16i3.778.

[4]. 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.

[5]. A. Lumini, L. Nanni, and G. Maguolo, "Deep Learning for Plankton and Coral Classification." [Online]. Available: https://www.kaggle.com/c/datasciencebowl

[6]. J. Y. Luo et al., "Automated plankton image analysis using convolutional neural networks," Limnol Oceanogr Methods, vol. 16, no. 12, pp. 814–827, Dec. 2018, doi: 10.1002/lom3.10285.

[7]. J. Zhang, W. Yang, Z. Lu, and D. Chen, "HR-YOLOv8: A Crop Growth Status Object Detection Method Based on YOLOv8," Electronics (Switzerland), vol. 13, no. 9, May 2024, doi: 10.3390/electronics13091620.

[8]. Vivi Afifah and S. Erniwati, "YOLOv8 for Object Detection: A Comprehensive Review of Advances,Techniques, and Applications," IJACI : International Journal of Advanced Computing and Informatics, vol. 2, no. 1, pp. 53–61, Jul. 2025, doi: 10.71129/ijaci.v2i1.pp53-61.

[9]. L. Li, Z. Liang, T. Liu, C. Lu, Q. Yu, and Y. Qiao, "Transformer-Driven Algal Target Detection in Real Water Samples: From Dataset Construction and Augmentation to Model Optimization," Water (Switzerland), vol. 17, no. 3, Feb. 2025, doi: 10.3390/w17030430.

[10]. H. K. Choi, C. K. Paik, H. W. Ko, M. C. Park, and H. J. Kim, "Recurrent DETR: Transformer-Based Object Detection for Crowded Scenes," IEEE Access, vol. 11, pp. 78623–78643, 2023, doi: 10.1109/ACCESS.2023.3293532.

[11]. J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," Dec. 01, 2023, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/make5040083.

[12]. Y. Shuai, J. Shi, Y. Li, S. Zhou, L. Zhang, and J. Mu, "YOLO-SW: A Real-Time Weed Detection Model for Soybean Fields Using Swin Transformer and RT-DETR," Agronomy, vol. 15, no. 7, Jul. 2025, doi: 10.3390/agronomy15071712.

[13]. P. Tomiło, P. Oleszczuk, A. Laskowska, W. Wilczewska, and E. Gnapowski, "Effect of Architecture and Inference Parameters of Artificial Neural Network Models in the Detection Task on Energy Demand," Energies (Basel), vol. 17, no. 21, p. 5417, Oct. 2024, doi: 10.3390/en17215417.