

Plankton Detection and Segmentation in Lake Geneva Using Holographic Imaging with YOLO11 Nano Model

Aolaritei Valentin¹, De Laurentis Alberto¹, Le Bras Martin Louis¹
Machine Learning (CS-433) ML4Science, EPFL, Switzerland

Abstract—Plankton are microscopic organisms essential to aquatic ecosystems and global climate regulation. Understanding their distribution and ecology is critical for preserving biodiversity and environmental health.

This study focuses on object detection and segmentation within the water column of Lake Geneva to facilitate plankton classification by domain experts. We implemented the Ultralytics YOLO11 Nano model for robust object segmentation, tailored for detecting diverse small entities.

Our proposed model demonstrated improved metrics (Recall: 1.00, mDice: 0.75) compared to the existing model, offering domain experts a reliable tool for plankton identification. This work contributes to the development of advanced methodologies for aquatic ecosystem research, with potential applications in biodiversity monitoring, climate studies, and environmental management.

I. INTRODUCTION

Plankton are microscopic organisms that play a vital role in global ecology today due to their influence on oxygen generation and impact on climate regulation [1], [2]. They form the bottom of the aquatic food chain and are responsible for approximately 45% of earth's oxygen production [3]. Moreover, changes in their ecology can strongly influence global climate, with important economic consequences [4]. For this reason, their identification and subsequent segmentation are crucial to understand better aquatic ecosystem and preserve its biodiversity and environmental health. The objective of this project was object segmentation throughout the water column of Lake Geneva (local name: *Lac Léman*). Due to limited domain knowledge, we decided to not classify objects into Living (Planktons), Non-Living or Bubbles. Instead, our main goal was to implement a robust detection and segmentation model capable of identifying objects in the dataset. The segmented entities will then be used by domain experts to properly classify them.

Our research was motivated by the need to improve the existing model used in plankton detection, as its generated images were often inaccurate, leading experts to make errors during analysis.

In order to achieve our goal, Ultralytics YOLO11 Nano model was applied on a large dataset which consisted of .tif images stacked on the depth-axis of the water columns. Each image consists of a slice of the water sample.

The paper is organized as follows. Section II describes the data collection process from Lake Geneva, the data segmentation using a state-of-the-art DL model, outlines the pre-processing and data-augmentation methods, and culminates with the presentation of YOLO11 model. Section III describes results obtained using computer vision techniques and their performance compared to the benchmark model. Section IV summarizes our findings and contribution to the research purpose. Finally, Section V addresses some ethical risks and considerations and discusses the impact of our project on direct and indirect stakeholders.

II. MODELS AND METHODS

In this section, we describe the method used to analyze underwater images. First, we start by describing data acquisition process from Lake Geneva (II-A), followed by data segmentation process executed thanks to an advanced DL model developed by Meta (II-B). We then explain in detail our pre-processing and data-augmentation techniques (II-C), together with our approach to image patching and feature-engineering (II-D). Finally, we present the model we will use for object segmentation (II-E).

A. Data Acquisition

Water samples were collected in Lake Geneva from the scientific platform LéXPLORÉ ($46^{\circ}30'0.819''$ N, $6^{\circ}39'39.007''$ E), located 570 meters offshore from the town of Pully, Switzerland [5]. The images were acquired using LISST-Holo2, a submersible digital holographic camera used by oceanographic researchers to capture underwater particles and organisms [6]. The output was stored as .tif file, where each .tif file consisted of 101 consecutive images on the vertical axis stacked together. A total of 108 holograms (made of 101 slices) were collected during six campaigns across different seasons of the year, ensuring temporal diversity in the sampled environment. The maximum depth under the floating platform is approximately 110 meters. Sample distribution between each depth range and campaigns are resumed in **Table I**.

¹GitHub repository: <https://github.com/CS-433/ml-project-2-mva>

Depth (m)	Date (MM-YY)					
	09-23	11-23	03-24	05-23	06-24	07-24
0-10	3	3	3	3	3	3
10-20	3	3	3	3	3	3
20-30	3	3	3	3	3	3
30-40	3	3	3	3	3	3
≈ 60	3	3	3	3	3	3
≈ 80	3	3	3	3	3	3
Total Number of Samples	18	18	18	18	18	18

Table I: Samples Collected per Depth-range for Each Campaign

B. Data Segmentation

Upon agreement with the laboratory supervisors, 35 holograms were segmented after careful consideration. First, we selected holograms from diverse depth-ranges in order to ensure an unbiased and balanced dataset, which can best represent the composition of Lake Geneva. Second, images with minimal noise and a good representation of both living and non-living creatures were prioritized. In this way, even with challenging images, we were sure the model can reliably detect objects, ensuring robustness under varied conditions. The segmentation process was done using Qu-Path [7], an open-source software for bio-image analysis and by checking all the 101 slices for each hologram. Within this application, we leveraged Meta's Segment Anything Model (SAM) [8], a pre-trained model known for its fast and precise segmentation capabilities. Following the segmentation process, results were downloaded and exported as GeoJSON files. A segmented example can be observed in **Figure 1**.

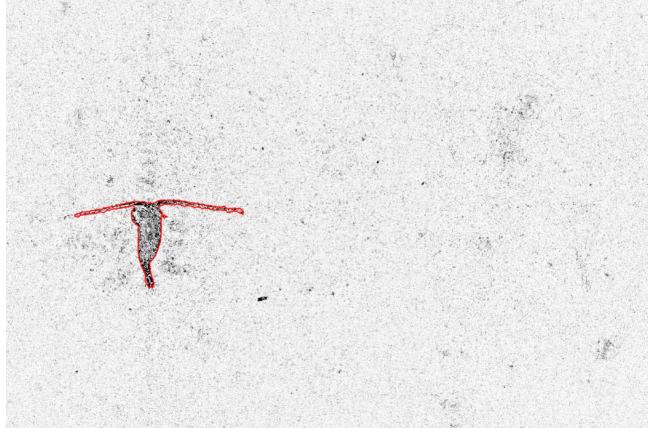


Figure 1: Example of Segmentation Result from SAM Model on the Selected z-Slice

C. Pre-Processing and Data-Augmentation

This stage involved various critical steps in order to refine our dataset and prepare it for the model training.

1) *Bounding-Box Split*: we performed an analysis of the annotated segmentation distribution size, with a primary focus on the bounding-box area of the segmented objects. As can be noticed in **Figure 2**, the distribution of the

dataset was heavily positively skewed with a median area of $144 \mu\text{m}^2$.

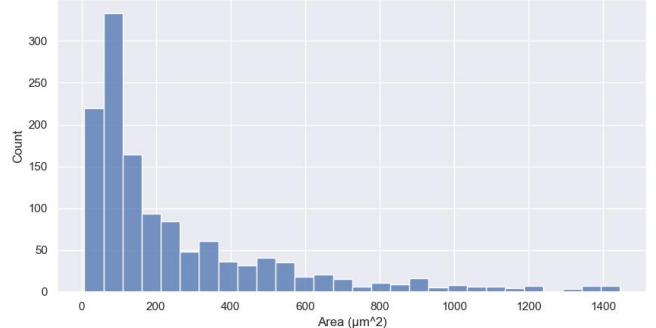


Figure 2: Distribution of the Surface of Segmented Objects Truncated at 95th Percentile for Better Visualisation

While traditional approaches categorize based on the area of the segmented object, we decided to adopt a more elaborate approach.

As a matter of fact, our method took into account the area of the bounding-box. The main reason was that recognizing an elongated, thin object might have had a large bounding-box but a small segmentation area. Such objects were valuable for our analysis and pivotal for researchers' studies. Based on the bounding-box surface distribution, we split the dataset into two subsets using a threshold area of $200 \mu\text{m}^2$, generating more homogenous datasets based on extensive image coverage, which enabled us to investigate better the smallest objects.

2) *Training-Validation-Test Split*: a key aspect of this pre-processing stage was ensuring a robust dataset splitting plan to prevent data leakage and maintain the integrity of model evaluation. To achieve this aim, the following approach was followed.

- Split was conducted at the image level rather than at the z-slice level.
- Each complete image was assigned exclusively to one set, eliminating the risk of slice-level data contamination.

Thanks to this approach, we were sure that the models were trained, validated and tested on completely different datasets.

SET	TRAIN	VALIDATION	TEST
% SPLIT	70	20	10
HOLOGRAMS	24	7	4

Table II: Trainining-Validation-Test Split

3) *Flipping Augmentation*: to enhance the model's generalization ability and increase the diversity of the dataset, image augmentation was performed using *Albumentations* library [9] on the training set. For this reason, we applied 4

basic transformations: horizontal flip, vertical flip, diagonal flip, and counter-diagonal flip. By doing so, we were able to expand our dataset significantly, increasing its size by a factor of 5. Principally, all these transformations were applied because they preserved the original image annotations while generating relevant variations for our task.

D. Feature Engineering

1) Image Patching: due to the unbalanced nature of our dataset toward objects with a really small bounding-box area compared to the total size of the images, we decided to introduce an image patching on all the slices of the hologram for the model examining objects below the aforementioned threshold area of $200 \mu\text{m}^2$.

As shown in **Figure 3**, there are several ways to patch images. We decided to use regular patching with a 8×8 grid of 126×190 pixel patches for training and validation set to avoid overfitting. Instead, for the test set and the prediction output the approach was different. It consisted in combining together the method used for training and validation with the other 3 methods to better detect objects on the edges of the regular patches.

By using patching, we were able to train the model on sub-images, which resulted in a higher number of images available and a possible higher detection power for small objects thanks to the bigger bounding-box to sub-image ratio.

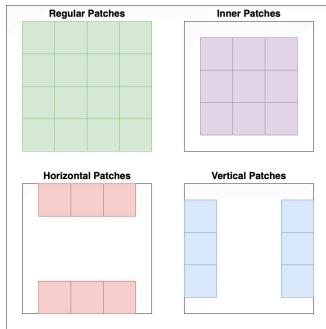


Figure 3: Image Patching Methods

E. Selected Model

We employed the 11th version of the YOLO (You Only Look Once) family [10], a state-of-the-art object detection and segmentation model pre-trained on COCO dataset [11], known for its speed and accuracy on small objects. Our model selection process involved the analysis of different versions of YOLO11 implemented using *Ultralytics* library [12]. More precisely, we started with YOLO11 Small, but then decided to try the YOLO11 Nano version, which showed considerable performance aspects. More precisely, it took almost 1.5 times less to train with negligible accuracy differences. For this reason, we opted to use YOLO11 Nano as our designed model for this project. The revolutionary

architecture of YOLO11, which can be observed in **Figure 4**, combines the qualities of Faster R-CNN with attention mechanisms that improve the model's focus on important regions within an image, by emphasizing spatial relevance in the feature maps.

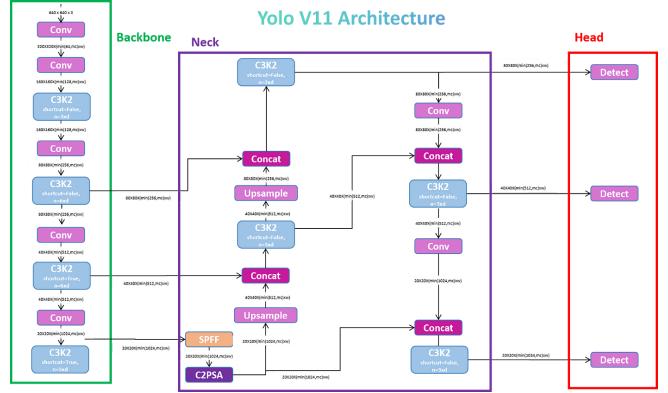


Figure 4: YOLO v11 Architeture

We further investigated YOLO's features, by evaluating its built-in augmentation methods, which were discarded because of a lack of improvement. It's remarkable to notice that YOLO was directly padding the input images and the patches directly to a square size using the longest side (e.g. for patches: 190×190 pixels). The only parameter we had to tune was the number of epochs, which was chosen equal to 100 because it was ensuring to achieve the best metrics possible. Our research was conducted using two different processed versions of the initial dataset:

- "NAIVE" model obtained by using only methods II-C2 and II-C3.
- "BOX-SPLIT" model obtained by using all methods in II-C and II-D.

It is important to underline that the YOLO model was run on full-image size for NAIVE model.

Instead, for the BOX-SPLIT model, it was run on patches for images containing only objects with bounding-box area below $200 \mu\text{m}^2$ ("SMALL") and on full-image size for those containing objects with area over the threshold ("BIG"). The patch predictions for this model were then recombined sequentially to have a single image output for all the models.

III. RESULTS AND DISCUSSION

In this section, we introduce our results for both models and compare them to the existing one, which is currently used by researchers. The mAP (Mean Average Precision), which is the most common metric for object detection, measures both precision and recall across multiple Intersection over Union (IoU) thresholds. This metric was used to evaluate the performance of both models and to detect the best YOLO weights during validation. The mAP results on the validation set are recorded in **Table III**.

MODEL	NAIVE	SMALL	BIG
mAP-50	0.379	0.424	0.559

Table III: mAP-50 for the Validation Set

The segmentation coordinates obtained after running YOLON were then used to reconstruct the mask on patches for SMALL model and on the full-images for NAIVE and BIG models.

We noticed that same objects were detected on adjacent z-slices. This issue was solved by selecting mask connected components on these adjacent z-slices and then applying the Normalized Variance focus measure to determine which had the best overall performance [13]. This metric was used to compensate for the differences in the average image brightness among different images and is defined as:

$$FM = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N \frac{(I(x, y) - \bar{I})^2}{\bar{I}}$$

Upon selecting the best slice, we merged the patches from the SMALL model all together to have a single output. Then, we recombined it with the results of the BIG model in order to build the unique BOX-SPLIT model. It was then time to assess our results compared to the BENCHMARK model used by researchers.

We evaluated precision, recall and mDice coefficients by comparing the overlapping area between the annotated ground truth that we acquired and the predictions made by the three models on test set as shown in **Table IV**.

MODEL	BENCHMARK	NAIVE	BOX-SPLIT
PRECISION	0.05	0.30	0.08
RECALL	0.79	0.86	1.00
mDICE	0.47	0.69	0.75

Table IV: Best Model Performance Evaluation on Test Set

We could then consider the BOX-SPLIT as the best model, because even if it presented lower results in precision, it had the best recall and mDice results which are more robust indicators in our problem settings.

This outcome can be easily explained by the small number of annotations we, as humans, were able to take and detect compared to the ones assessed by the three models, resulting in an highly biased number of false positive predictions. Besides that, BOX-SPLIT was the best in detecting true positives, because it was able to find all of them and not decompose big objects into several smaller ones as the BENCHMARK.

A more robust and meaningful evaluation became evident when comparing the BENCHMARK model's image reconstruction to the BOX-SPLIT one, which demonstrated a

remarkable improvement in object segmentation, as can be seen in **Figure 5** and **Figure 6**.



Figure 5: Example of Full-Image Output of BOX-SPLIT Model



Figure 6: Example of Full-Image Output of BENCHMARK Model

IV. CONCLUSION

Despite the possibility of achieving better performance with a more detailed and extensive data acquisition, and with an higher numbers of holograms available, our results for both the models were still exceptionally good compared to the pre-existent model.

It is also important to underline that our final predictions will always be analyzed by a human eye, so the most important feature is visual output and not only classification or segmentation metrics.

Finally, we believe that an higher sample size would strongly increase the metrics for BOX-SPLIT model compared to the NAIVE one, allowing also the possibility to experiment segmentation with a different version of YOLO model, such as the XL one.

V. ETHICAL RISKS

This project was carefully evaluated for ethical risks, and none were found. Our aim was to develop a model able to segment Lake Geneva's aquatic micro-environment. By doing so, we provide domain researchers a tool for scientific understanding and environmental conservation. The key stakeholders which were considered during the development of this project are environment and domain experts.

As mentioned in Section I, Planktons are responsible for approximately 45% of Earth's oxygen production and are fundamental to climate regulation. Moreover, they are also responsible for pollution mitigation and support biodiversity [14]. To assess risks, we analysed each step of this project. Since our data was sourced by experts, we believe strict procedures were followed. Hence, no actions impacted directly plankton population or their environment.

Subsequently, by developing a machine learning model which can reliably detect aquatic objects, we firmly believe this may lead to a better understanding of plankton distribution, supporting environmental preservation and biodiversity conservation.

Domain experts, such as marine biologists and climate change researchers, rely on data to make marine management decisions [15]. As a matter of fact, understanding plankton spatial and temporal distribution plays an important role in assessing environmental status and mitigating potential impacts of human activity. To mitigate potential risks, we ensured that our model was robust and unbiased in its object segmentation. This will help researchers identify plankton distribution. Plus, thanks to more accurate data, researchers can better understand the environment and take more precise decisions regarding biodiversity conservation, climate change mitigation, and pollution contrast.

Ultimately, through cautious consideration of these stakeholders and thorough evaluation of the project's methodologies and data sources, we confidently ruled out ethical risks.

VI. ACKNOWLEDGEMENTS

We would like to express our gratitude to Professor Natacha Tofield-Pasche for giving us the possibility to work on this project and for the support in the initial stage of the research.

Last but not least, we would like to express our heartfelt gratitude to our mentor Mallory Wittwer for his invaluable support and guidance throughout the course of this project. We truly appreciate the time and effort you dedicated to mentoring us, and we are grateful for the opportunity to have worked under your guidance.

REFERENCES

- [1] C. B. Field, M. J. Behrenfeld, J. T. Randerson, and P. Falkowski, "Primary production of the biosphere: integrating terrestrial and oceanic components," *Science*, vol. 281, no. 5374, pp. 237–240, 1998.
- [2] F. P. Chavez, M. Messié, and J. T. Pennington, "Marine primary production in relation to climate variability and change," *Annual review of marine science*, vol. 3, no. 1, pp. 227–260, 2011.
- [3] M. J. Behrenfeld, J. T. Randerson, C. R. McClain, G. C. Feldman, S. O. Los, C. J. Tucker, P. G. Falkowski, C. B. Field, R. Frouin, W. E. Esaias *et al.*, "Biospheric primary production during an enso transition," *Science*, vol. 291, no. 5513, pp. 2594–2597, 2001.
- [4] V. P. Pastore, T. G. Zimmerman, S. K. Biswas, and S. Bianco, "Annotation-free learning of plankton for classification and anomaly detection," *Scientific reports*, vol. 10, no. 1, p. 12142, 2020.
- [5] A. Wüest, D. Bouffard, J. Guillard, B. W. Ibelings, S. Lavanchy, M.-E. Perga, and N. Pasche, "Léxplore: A floating laboratory on lake geneva offering unique lake research opportunities," *Wiley Interdisciplinary Reviews: Water*, vol. 8, no. 5, p. e1544, 2021.
- [6] Z. Liu, S. Giering, T. Takahashi, T. Thevar, M. Takeuchi, N. Burns, B. Thornton, J. Watson, and D. Lindsay, "Advanced subsea imaging technique of digital holography: in situ measurement of marine microscale plankton and particles," in *2023 IEEE Underwater Technology (UT)*. IEEE, 2023, pp. 1–9.
- [7] P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Domrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman *et al.*, "Qupath: Open source software for digital pathology image analysis," *Scientific reports*, vol. 7, no. 1, pp. 1–7, 2017.
- [8] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.
- [9] 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, 2020.
- [10] J. Redmon, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.
- [11] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2015. [Online]. Available: <https://arxiv.org/abs/1405.0312>
- [12] G. Jocher, J. Qiu, and A. Chaurasia, "Ultralytics yolo," Ultralytics. [Online]. Available: <https://ultralytics.com>

- [13] I. Helmy, A. Hamdy, D. Eid, and A. S. Abouelfetouh Elshaer, “Autofocusing optimal search algorithm for a telescope system,” 07 2021.
- [14] V. Doumeizel and J. R. Dolan, “The launch of the plankton manifesto in september 2024,” pp. 525–526, 2024.
- [15] J. F. Tweddle, M. Gubbins, and B. E. Scott, “Should phytoplankton be a key consideration for marine management?” *Marine Policy*, vol. 97, pp. 1–9, 2018.