Automatic Plankton Detection and Classification on Raw Hologram with a Single Deep Learning Architecture

Romane Scherrer, Rodrigue Govan, Thomas Quiniou, Thierry Jauffrais, Hugues Lemonnier, Sophie Bonnet, Nazha Selmaoui-Folcher

Automatic Plankton Detection and Classification on Raw Hologram with a Single Deep Learning Architecture

Romane Scherrer*[1], Rodrigue Govan[1], Thomas Quiniou[1], Thierry Jauffrais[2], Hugues Lemonnier[2], Sophie Bonnet [3], Nazha Selmaoui-Folcher*[1]

[1] Université de la Nouvelle-Calédonie, ISEA, Nouméa, New-Caledonia.
romane.scherrer@hotmail.fr, nazha.selmaoui@unc.nc

[2] UMR Entropie, Ifremer, Nouméa, New-Caledonia.

[3] UMR MIO, IRD, Marseille, France.

*corresponding authors

Keywords: Object Detection, Deep Learning, Plankton, Holography.

Abstract.
Digital holography is an imaging process that encodes the 3D information of objects into a single intensity image. In recent years, digital holography has been used to detect and count various microscopic objects and has been applied in submersible equipment to monitor the distribution of plankton. To count and classify plankton, conventional methods require a holographic reconstruction to decode the hologram and identify objects. To avoid this iterative and time consuming reconstruction, we demonstrate an object detection based approach that simultaneously performs the detection and classification of all plankton within raw holograms. The proposed method achieved a mean Average Precision (AP@.5) score of 0.981 with 13 plankton species on 512x512 holograms with an inference time of 14 ms per image. Our method can be implemented to detect and count other microscopic objects in raw holograms.

1 Scientific Background

The observation and statistics of marine plankton are essential to measure the health of our oceans. In recent years, several submersible equipment [1] (ISIIS, LISST-Holo, eHoloCam) have been deployed as part of large-scale campaigns to acquire *in situ* images of plankton. Some of these systems use digital holography, a method that enable high resolution images acquisition over a large water column and at high flow rates. Since a hologram encodes the 3D information of all plankton as a single intensity image, a decoding process, called holographic reconstruction, is required to retrieve the sample image from its hologram. Unfortunately, the methods used to process holograms and then count and classify the species are still very time-consuming and manual.

With the multiplication of collected images, various efforts have been made to accelerate and improve the holographic reconstruction, for instance, by adopting a convolutional neural network (CNN) to automatically find the focus [2] or to reconstruct a de-focused hologram without performing an auto-focusing or phase recovery routine [3, 4]. Even though those approaches greatly accelerate the holographic reconstruction, the classification and detection of the objects need to be performed afterwards. As a result, some research works have focused on classifying objects directly on raw holograms, thus avoiding iterative and time-consuming operations required for reconstruction. In [5] a CNN is trained to perform a classification on cropped simulated holograms with 16 different plankton species but a preliminary detection is necessary

to determine the regions of interest (ROIs). [6] use a sliding window detection method with a CNN that performs a binary classification to detect and count cells. However, all the proposed methods often perform a single task on raw holograms, either a detection or a classification. Object detection methods such as [7] Faster-RCNN, YOLO, SSD or RefineDet, offer an alternative by performing real-time localization and classification of all objects within an image in one pass. By labeling the detected objects within a hologram with rectangular bounding boxes, a detection model could greatly improve the application efficiency of digital holography.

The aim of the paper is to show that the classification and localization of plankton can be simultaneously performed on raw holograms with an object detection model. For that purpose, a dataset of labeled in-line holograms will be simulated with 13 different plankton species. The paper is organized as follows. In Section two, the generation of holographic datasets and the object detection models are described. Section three show the performances of the models. Conclusions and perspectives are given in the last section.

2 Materials and Methods

2.1 *Hologram Formation*

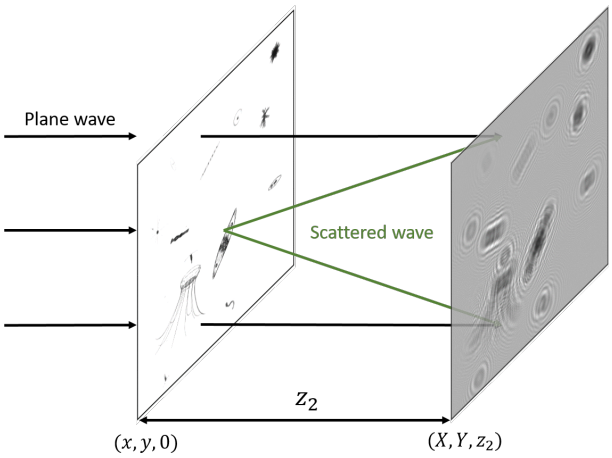

Figure 1: In line-holography.

For an in-line configuration (see Fig. 1), the reference and object waves share the same optical axis and an object can be described by a complex transmission function [8] at a given z plane:

$$t_z(x, y) = \exp[-a(x, y)] \exp[i\phi(x, y)] \tag{1}$$

where $a(x, y)$ describes the absorption of the object and $\phi(x, y)$ is the phase distribution. The transmission function can be used to calculate the wavefront just behind the object $U_{z^+}(x, y)$:

$$U_{z^+}(x, y) = t_z(x, y) U_{z^-}(x, y) \tag{2}$$

where $U_{z^-}(x, y)$ is the incident wave that can be either plane or spherical.
Considering that the object is located at $z = 0$, the exit wave given by Eq. 2 can be rewritten as $U_{0^+}(x, y) = t_0(x, u) U_{0^-}(x, y)$ and is propagated to the detector/hologram plane which is located at $z = z_2$ along the optical axis. This propagation is simulated by the angular spectrum method by calculation of the following transformation:

$$U_{z_2}(X, Y) = TF^{-1}\left[TF\left(U_{0^+}(x, y)\right) \times \exp\left(\frac{2\pi i z_2}{\lambda}\sqrt{1 - (\lambda u)^2 - (\lambda v)^2}\right)\right] \tag{3}$$

where λ is the wavelength and (u, v) are the Fourier domain coordinates. TF^{-1} and TF denoted the inverse Fourier transform and the direct Fourier transform, respectively.

Note that Eq. 3 is often expressed as $U_{z_2}(X,Y) = R(X,Y) + O(X,Y)$ where R and O are the reference and the object waves that interfere at the surface of the recording medium. The recorded hologram at $z = z_2$ is the intensity calculated by:

$$H_{z_2}(X,Y) = |U_{z_2}(X,Y)|^2 = U_{z_2}(X,Y)U^*_{z_2}(X,Y) \tag{4}$$

where * denotes the complex conjugate. As a result, a hologram can be simulated once λ, z_2, $U_{0^-}(x,y)$ and $t_0(x,y)$ are known or set.

2.2 *Transmission Function Simulation*

To generate a dataset of labeled holograms for an object detection task, the complex transmission function $t_0(x,y)$ of several objects in a plane $(x,y,z=0)$ must be simulated first. In that case, two labeled datasets of plankton images will be used as objects. The first dataset consists of shadow images collected by the In Situ Ichthyoplankton Imaging System (ISIIS) which was the subject of a competition on Kaggle [1]. This open source dataset consists of 121 marine plankton species among which 10 species with a number of images greater than 1000 were selected for our simulations. The second dataset (custom) consists of optical microscopy images of 3 phytoplankton species from New Caledonia (noted P1, P16 and P17). The plankton was imaged with a brightfield microscope at $\times 10$ magnification. The images were thresholded, segmented into (ROIs) and manually labeled. Fig. 2 presents the number of images per species. Note that for each datasets, the ROI segments are labeled and saved as grayscale images where the background has an uniform value equal to 1. As a result, an image $I(x,y)$ can be converted into an absorption $a(x,y) = -1 \times I(x,y) + 1$ so that $t_0(x,y) = \exp[-a(x,y)]\exp[i\phi(x,y)]$ inside the object support and $t_0(x,y) = 1$ where there is no object. In the latter situation, the incident wave remains undisturbed.

To simulate $t_0(x,y)$ with various objects, the transmission functions of several plankton images can be randomly placed on a $N \times N$ empty (all-ones) image. By doing so, the (x,y)-axis coordinates of the bounding boxes are randomly set. Moreover, the plankton images are already saved as ROIs so that the bounding boxes width and height are the images dimensions. Since the images are classified per species, the labels of a simulated $t_0(x,y)$ for an object detection task (classes and bounding boxes coordinates) can be completely set.

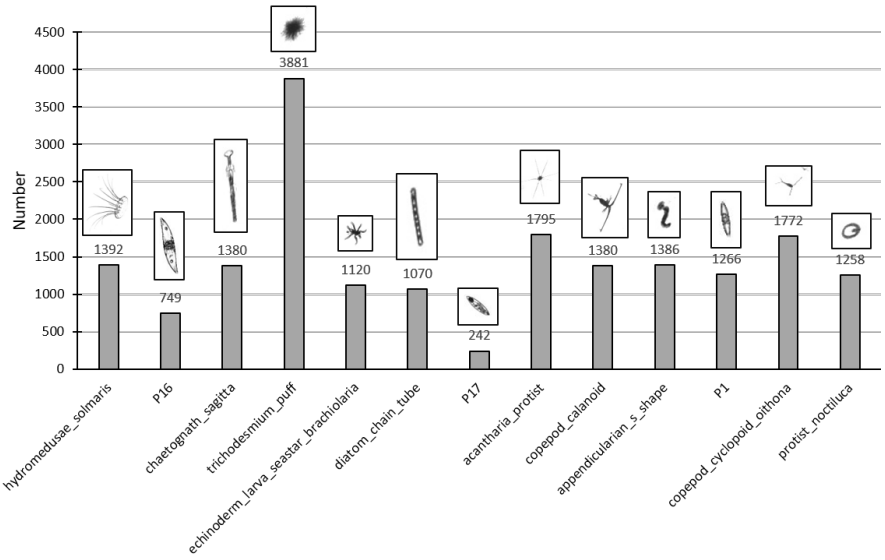

Figure 2: Number of images per species.

[1] https://www.kaggle.com/c/datasciencebowl/

2.3 *Hologram Simulation*

Before simulating the $t_0(x, y)$ and the holograms, the plankton images from the two presented datasets were randomly split, per class, in a 80:20 ratio for training and testing, respectively. We have considered that the plankton are pure amplitude objects so that $\phi(x, y) = 0$. The simulation of $t_0(x, y)$ proceeds as follows. First, for each simulated $t_0(x, y)$, 13 plankton images (one per species) are randomly selected. The images are then randomly rotated and flipped with 4 possible rotations (0°,90°,180° or 270°) and 3 possible flips (None, horizontal or vertical). Then, the transmission functions are individually modified so that $t_{plankton}(x, y) = \exp[-C \times a(x, y)]$ where C is a random constant and $C \in [0.5, 1]$. Next, the 13 transmission functions are randomly placed without overlapping on a 512×512 empty image to generate $t_0(x, y)$. Finally,the hologram $H_{z2}(X, Y)$ is simulated with Eq. 3 and Eq. 4. Both the holograms and the $t_0(x, y)$ are normalized between 0 and 1 and saved.

For the simulations, we have considered $\lambda = 530nm$ (green), $z_2 = 0.8mm$ and an incident plane wave $U_{0-}(x, y) = 1$. 8000 and 2000 holograms were simulated for training and testing, respectively. Fig. 3 presents an example of a simulated and labeled $t_0(x, y)$ and its corresponding hologram.

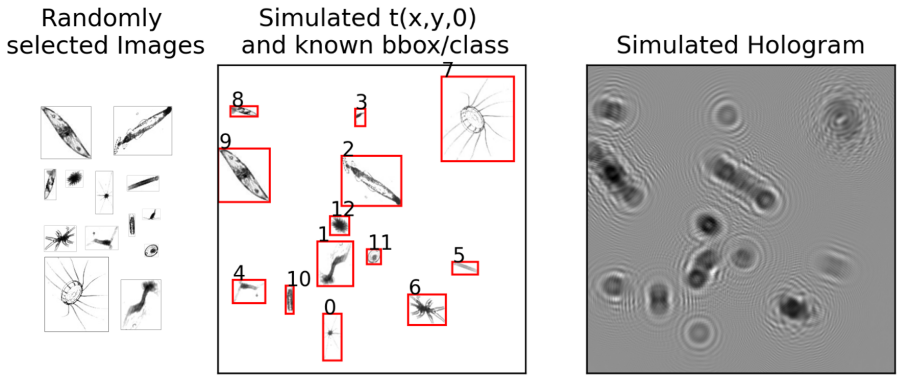

Figure 3: Simulation example. 13 plankton images are used to generate a labeled hologram.

2.4 *Object Detection Models*

To perform object detection on raw holograms, we chose two YOLOv5 [9] models that were pre-trained on the COCO dataset, namely YOLOv5s [2] (the smallest) and YOLOv5x (the largest) with $7.3M$ and $87.7M$ parameters, respectively. The models were trained on 8000 holograms during 400 epochs with a batch size of 8 and tested on 2000 holograms. The SGD optimizer was used with an initial learning rate equal to 0.01. To further evaluate the object detection performances on raw holograms, two models were also trained on the transmission functions $t_0(x, u)$ with the same hyperparameters.

The experiments were carried out on a 2.9 GHz Intel Core i7 PC with 64 GB of RAM and a Nvidia GTX 2060 GPU. The training took 8 hours for the small model and 2 days for the larger one.

3 **Results**

In this section, we report the object detection performances with the well-known average precision (AP) metrics [7]. We recall that the AP@.5 and AP@.75 are the average precision computed with an intersection over union (IoU) threshold $t = 0.5$ and $t = 0.75$, respectively. The AP@[.5:.95] is reported by computing the mean AP@ with 10 different IoU thresholds [.5:.05:.95]. The mean inference time was evaluated on 2000

[2]https://github.com/ultralytics/yolov5

holograms and was carried out on a GTX 2060 GPU. It includes FP16 inference, post-processing and non-maximum suppression. The Tab. 1 summarizes the performances of the object detection tasks performed on the raw holograms and on the transmission functions.

Table 1: Performances.

Model	Inputs	AP@.5:.95	AP@.5	AP@.75	Speed
YOLOv5s	Holograms	0.820	0.976	0.928	4 ms
	$t_0(x, y)$	0.967	0.985	0.985	
YOLOv5x	Holograms	0.855	0.981	0.955	14 ms
	$t_0(x, y)$	0.980	0.993	0.989	

For the models trained on the holograms, the AP@.5 are 0.976 and 0.981 for YOLOv5s and YOLOv5x, respectively. For the models trained on the transmission functions, the AP@.5 are slightly better with 0.985 and 0.993 for YOLOv5s and YOLOv5x, respectively. The AP@[.5:.95] are significantly higher on $t_0(x, y)$ than on holograms (eg. 0.980 vs. 0.855 for YOLOv5x) but the AP@.75 are still high on holograms (0.928 and 0.955 for YOLOv5s and YOLOv5x, respectively). Those results suggest that the detectors trained on the holograms are efficient for a IoU threshold $\leq .75$ but that their performances start to decline at a higher threshold.

Fig. 4 shows the confusion matrix of YOLOv5x at IoU.5 on the test holograms and one of its predictions. One can notice that the diffraction pattern of an object spreads beyond its bounding box. In fact, the further away the object is from the camera, the more this effect will be visible on the hologram. Because of this and the lack of sharp edge, a detector trained on holograms was expected to have difficulty in determining the object boundaries with a high IoU. Although the models trained on $t_0(x, y)$ perform better than those on holograms, the performances are still promising: the species and location of all the plankton within a 512x512 hologram can accurately be determined with a true positive rate between 91% and 100% for the 13 classes. Moreover, the prediction can be performed in real-time (4 ms for YOLOv5s and 12 ms for for YOLOv5x).

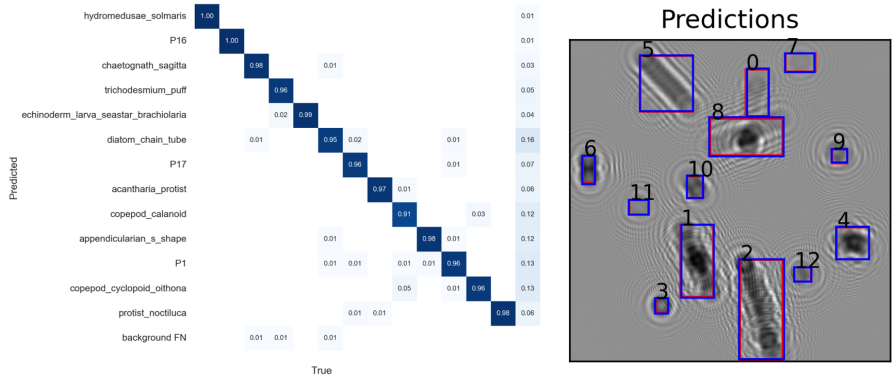

Figure 4: Confusion matrix at IoU.5 and model predictions (blue : ground-truth , red : predicted).

4 Conclusion

In this paper, we have generated a simulated and labeled holograms dataset from optical and shadow plankton images for an object detection task. 13 different species were considered. Two versions of YOLOv5 are trained on the holograms and on the transmission functions, respectively. Even though the AP of the models trained on the holograms are lower than those trained on the transmission functions, the difference

in AP@.5 is less than 1.2%. With a AP@.5 score of 0.981, a YOLOv5x model can perform detection and classification of all plankton within a 512x512 raw holograms in a single pass in $14ms$. Although this method was validated with plankton images, it can be implemented to localize, count and identify other microscopic objects in raw holograms.

While the results on simulated holograms are promising, it is often complicated and time consuming to put together a large dataset of real labeled holograms to train a detector. When a small labeled dataset is available, it might be beneficial to pre-train a detector with a large amount of simulated holograms and then use a transfer learning method to fine tune the model on the small dataset. Another approach would be to rely on an intensive data augmentation. Some works in the literature use de-focused back-propagated holograms as inputs of a deep learning model rather than raw holograms. By back-propagated the holograms on several planes near the the correct global focus, the dataset could be significantly enlarged. For future work, we intend to study the object detection performances on a de-focused holograms dataset in which we will address the problem of overlapping objects.

References

[1] X. Liu, X. Liu, H. Zhang, Y. Fan, H. Meng, "Research progress of digital holography in deep-sea in situ detection," *Seventh Symposium on Novel Photoelectronic Detection Technology and Applications* , 117636J, 2021.

[2] T. Shimobaba, T.Kakue and T. Ito. "Convolutional Neural Network-Based Regression for Depth Prediction in Digital Holography". IEEE 27th International Symposium on Industrial Electronics (ISIE), pp. 1323–1326, 2018.

[3] Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng and A. Ozcan. "Phase recovery and holographic image reconstruction using deep learning in neural networks". *Light: Science and Applications*, vol.7, no.2, pp. 17141,2018.

[4] Y. Wu, Y. Rivenson, Y. Zhang, H. Günaydin, X. Lin and A. Ozcan. "Extended depth-of-field in holographic image reconstruction using deep learning based auto-focusing and phase-recovery". *Optica*,vol.5, pp. 704-710, 2018.

[5] Y. Zhang, Y. Lu, H. Wang, P. Chen and R. Liang. "Automatic classification of marine plankton with digital holography using convolutional neural network". *Optics and Lasers in Engineering*, vol.139, pp. 106979, 2021.

[6] C. Trujillo and J. Garcia-Sucerquia. "Automatic detection and counting of phase objects in raw holograms of digital holographic microscopy via deep learning". *Optics and Lasers in Engineering*, vol.120, pp. 13–20, 2019.

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

[8] T. Latychevskaia and H. Fink, "Practical algorithms for simulation and reconstruction of digital in-line holograms". *Appl. Opt*, vol. 54, pp. 2424-2434, 2015.

[9] J. Redmon, S. Divvala, R. Girshick and A. Farhadi. "You only look once: Unified, real-time object detection". *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016.