A New Method for the Detection of Siliceous Microfossils on Sediment Microscope Slides Using Convolutional Neural Networks

Camille Godbillot, Ross Marchant, Luc Beaufort, Karine Leblanc, Yves Gally, Thang D Q Le, Cristele Chevalier, Thibault de Garidel-thoron

To cite this version:

Camille Godbillot, Ross Marchant, Luc Beaufort, Karine Leblanc, Yves Gally, et al.. A New Method for the Detection of Siliceous Microfossils on Sediment Microscope Slides Using Convolutional Neural Networks. Journal of Geophysical Research: Biogeosciences, 2024, 129 (9), pp.e2024JG008047. ⟨10.1029/2024jg008047⟩. ⟨hal-04709563⟩

HAL Id: hal-04709563
https://hal.science/hal-04709563v1

Submitted on 25 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

JGR Biogeosciences

METHOD
10.1029/2024JG008047

A New Method for the Detection of Siliceous Microfossils on Sediment Microscope Slides Using Convolutional Neural Networks

Camille Godbillot[1], **Ross Marchant**[2], **Luc Beaufort**[1], **Karine Leblanc**[3], **Yves Gally**[1], **Thang D. Q. Le**[3], **Cristele Chevalier**[3], **and Thibault de Garidel-Thoron**[1]

[1]Aix-Marseille Université, CNRS, IRD, INRAE, CEREGE, ITEM, Aix-en-Provence, France, [2]Image Analytics, Benalla, VIC, Australia, [3]Aix-Marseille Université, Université de Toulon, CNRS, IRD, MIO, Marseille, France

Key Points:

- Faster Region-based Convolutional Neural Network models are efficient at detecting marine diatom frustules on microscope slide images
- These models can be applied to detect the diatoms on images from a diverse array of environmental conditions
- Adequate sample preparation and image quality enhance model results

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

C. Godbillot,
godbillot@cerege.fr

Citation:

Godbillot, C., Marchant, R., Beaufort, L., Leblanc, K., Gally, Y., Le, T. D. Q., et al. (2024). A new method for the detection of siliceous microfossils on sediment microscope slides using convolutional neural networks. *Journal of Geophysical Research: Biogeosciences*, *129*, e2024JG008047. https://doi.org/10.1029/2024JG008047

Received 2 FEB 2024
Accepted 24 AUG 2024

Author Contributions:

Conceptualization: Camille Godbillot
Formal analysis: Camille Godbillot
Funding acquisition: Cristele Chevalier, Thibault de Garidel-Thoron
Investigation: Camille Godbillot
Methodology: Camille Godbillot, Ross Marchant, Luc Beaufort, Yves Gally
Software: Ross Marchant
Supervision: Thibault de Garidel-Thoron
Validation: Luc Beaufort, Karine Leblanc, Thang D. Q. Le, Thibault de Garidel-Thoron
Writing – original draft: Camille Godbillot
Writing – review & editing: Camille Godbillot, Ross Marchant, Luc Beaufort, Karine Leblanc, Thang

© 2024. The Author(s).
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract Diatom communities preserved in sediment samples are valuable indicators for understanding the past and present dynamics of phytoplankton communities, and their response to environmental changes. These studies are traditionally achieved by counting methods using optical microscopy, a time-consuming process that requires taxonomic expertise. With the advent of automated image acquisition workflows, large image data sets can now be acquired, but require efficient preprocessing methods. Detecting diatom frustules on microscope images is a challenge due to their low relief, diverse shapes, and tendency to aggregate, which prevent the use of traditional thresholding techniques. Deep learning algorithms have the potential to resolve these challenges, more particularly for the task of object detection. Here we explore the use of a Faster Region-based Convolutional Neural Network model to detect siliceous biominerals, including diatoms, in microscope images of a sediment trap series from the Mediterranean Sea. Our workflow demonstrates promising results, achieving a precision score of 0.72 and a recall score of 0.74 when applied to a test set of Mediterranean diatom images. Our model performance decreases when used to detect fragments of these microfossils; it also decreases when particles are aggregated or when images are out of focus. Microfossil detection remains high when the model is used on a microscope image set of sediments from a different oceanic basin, demonstrating its potential for application in a wide range of contemporary and paleoenvironmental studies. This automated method provides a valuable tool for analyzing complex samples, particularly for rare species under-represented in training data sets.

Plain Language Summary Microfossils preserved in ocean sediments are studied to explore the impact of climate change on planktonic communities. The usual way to count these microfossils is slow and requires an expert to identify them on microscope images. In this study, we explore how artificial intelligence can be used on microscope images to detect the microfossils produced by one particular group, diatoms. Our results show that models can be trained to identify these objects, including the ones that were not specifically shown to the model during the training phase. However, the quality of the microscope image, and of the sample preparation beforehand, can affect how well the model works. This new protocol has good potential to be used on diatom images differing in age and geographical origins. Adopting this method could make it possible to rapidly increase the temporal resolution and spatial extent of existing data on diatom diversity, which could thus improve our knowledge of plankton resilience to climate change.

1. Introduction

There exists a significant variability in size, growth rates, nutrient acquisition, and trophic interactions within and between different phytoplankton groups. This diversity exerts a fundamental control on biogeochemical cycles, for instance through its influence on carbon export from the surface ocean and on food web dynamics. Ongoing climate change, through a comprehensive set of processes, impacts phytoplankton diversity and size structure, with consequences both for carbon storage and trophic efficiency (Henson et al., 2021; Passow & Carlson, 2012). As a result, efforts have been made to integrate phytoplankton diversity into Earth system models (Le Quéré et al., 2005), to better predict the effect of different climate scenarios on biogeochemical cycling and food web efficiency.

The study of phytoplankton communities is traditionally achieved by manually counting species identified in sediment or plankton net samples, a time-consuming process that requires specific taxonomic expertise.

D. Q. Le, Cristele Chevalier,
Thibault de Garidel-Thoron

Consequently, existing time series of plankton changes often have low temporal and spatial resolution. Moreover, counting methodologies may vary between research groups, making interlaboratory comparisons difficult (Zingone et al., 2015). As a response to these challenges, a variety of proxies have been proposed to describe the phytoplankton community (Lombard et al., 2019), including (but not limited to) the use of satellite ocean color (Hirata et al., 2008; Mouw et al., 2017; Nair et al., 2008), DNA meta-barcoding analyses (De Vargas et al., 2015), pigment analysis (Claustre, 1994) and fluorescence (Petit et al., 2022). These methods yield information on the planktonic community's size distribution and taxonomic composition over a broad range of sizes. However, their analyses operate at different levels (functional, genetic, etc.), which may only partially align with the traditional approach of morphological taxonomy. In parallel, efforts have been made to develop automated imaging techniques, making it possible to obtain taxonomical and morphological data on both single organisms and the total population (reviewed in Lombard et al., 2019). The palette of methods described above is powerful for monitoring present-day plankton diversity and provides a means of obtaining standardized phytoplankton time series.

The sediment record of the biominerals produced by different phytoplankton groups constitutes a different type of archive. It can be used to study the past sensitivity of the phytoplankton community to environmental change and its implications on the past strength of the biological pump (Kohfeld et al., 2005). The taxonomic composition of phytoplankton is also used as a proxy for various environmental variables (Abrantes et al., 2007; Marino et al., 2014). The study of the sedimentary record of biomineralization is typically achieved using light microscopy on fixed samples. The development of automated image acquisition techniques using electron scanning or optical microscopy has made it possible to acquire large sets of plankton images from sediment samples with limited human intervention. These methods have been used successfully to obtain large image data sets documenting the past production of a variety of fossil organisms in marine and freshwater environments, including coccoliths, radiolarians, foraminifera, or pollen grains (Beaufort et al., 2014; Bourel et al., 2020; Gimenez et al., 2024; Marchant et al., 2020; Tetard et al., 2020). To treat these large image data sets, studies increasingly rely on machine learning algorithms for object detection and identification. While the use of these techniques was once limited to IT experts, their recent integration into user-friendly software such as EcoTaxa (Picheral et al., 2017) or ParticleTrieur (Marchant et al., 2020) now makes it possible for researchers without specific training in plankton identification to obtain plankton counts from their images or to build and train a classification model based on their own image library. This is a handy feature when studying fossil data (Carlsson et al., 2023), for which annotated data sets are rare. Among machine learning techniques, the use of deep-learning algorithms has been generalized (Borowiec et al., 2022), encouraged by the increasing amount of publicly available software libraries and the advent of fast and affordable GPU-based computing systems. The use of convolutional neural networks (CNNs) for image recognition has been developed in particular; CNNs are able to extract the features of interest in an image, which makes them useful and efficient for tasks such as object detection and object classification (Borowiec et al., 2022; Serre, 2019). To date, fully automatic workflows have been successfully developed to study the fossil record of a variety of biominerals (Beaufort & Dollfus, 2004; Tetard et al., 2020); however, there exists to our knowledge no means of performing such studies on the diatom record. This, in part, is due to the difficulties in efficiently detecting them on a microscope image.

Diatoms are a phytoplankton group of particular interest, as they live in most marine and freshwater environments, which makes them responsible for one-fifth of the photosynthesis on Earth (Armbrust, 2009). They are often the dominant phytoplankton group in turbulent and nutrient-rich environments (Kemp & Villareal, 2018; Margalef, 1978). Diatoms produce remarkable species-specific silicified cell walls called frustules. At the end of their life cycle, these biominerals are transported through the water column and accumulate in the sediments, creating deposits that can reach, in some areas, more than several hundred meters in thickness (Armbrust, 2009). The number and large size of these frustules in comparison to the biominerals produced by other phytoplankton groups makes them one of the main contributors to the biological pump (Jin et al., 2006). The blooms formed by some diatom species have also been shown to contribute significantly to carbon export from the surface ocean in more oligotrophic regions such as the Mediterranean Sea (Leblanc et al., 2018). Studies have shown, however, that the diatoms' contribution to carbon export depends on the group's composition (Ragueneau et al., 2006; Tréguer et al., 2018). Climate-induced changes in the diversity of this group could, therefore, have implications for the biological pump. Furthermore, the decrease in diatom abundance in the total phytoplankton assemblage in favor of nanoplankton groups, including coccolithophores, has been linked in some environments to a decrease in primary productivity and carbon export (Iriarte & González, 2004). It is thus essential to describe how the diatom community responds to climate change, both within the group and in comparison to other phytoplankton groups.

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Diatom identification in sediments is traditionally achieved using optical and electron scanning microscopy (SEM). Detecting diatom frustules on a microscope image remains a challenge. Implementing an automated object detection workflow for diatoms is indeed hindered by the difficulty of isolating these minerals from the background image. Traditional thresholding techniques have proven effective to segment calcareous nanofossils, due to their birefringence properties (Beaufort et al., 2021), as well as radiolarians, which exhibit high relief (Tetard et al., 2020). They have also been used to detect diatoms from images of modern-day river samples (Bueno et al., 2017) or, more generally, on frustule images obtained using brightfield microscopy (Kloster et al., 2014). However, the low relief of diatom frustules on sediment slides, the diversity of frustule shapes, and a tendency for frustules to cluster or break complicate the use of existing plankton detection workflows (Kloster et al., 2014). As a result, studies increasingly explore the use of deep neural networks (Kloster et al., 2020, 2023) to extract individual diatom frustules from a raw image. However, most deep learning-based attempts at detecting diatoms on microscope images have focused on samples used for water monitoring (Bueno et al., 2018; Kloster et al., 2023) or for forensic analyses (Yu et al., 2022). These models have been either developed for living cells (Li et al., 2020) or for images acquired using SEM (Yu et al., 2022) and thus are not directly transposable to marine sediment samples, which present a lot of damaged and broken cells.

In this study, we propose and test the use of a Faster R-CNN model (Ren et al., 2017) to detect diatom frustules on images obtained from sediment slides using an automated microscope. We detail how the images in the data set were acquired, how they were annotated to constitute a training and validation library, and how the CNN model performed when compared to a test set. We investigate which factors influence the detection performance.

2. Materials and Methods

2.1. Image Acquisition

We trained an object detection model using images from samples collected from the Lionceau sediment trap series located in the Gulf of Lion in the NW Mediterranean Sea (42°N, 4.5°E, 2,400 m). This series comprises 80 samples collected between 2010 and 2018. The Gulf of Lion is among the most productive areas in the Mediterranean Sea and exhibits a relatively large diversity of siliceous biomineralizers, including diatoms and silicoflagellates (Rigual-Hernández et al., 2013). Each sample's collection period was 2 weeks, on average, and the samples span the entire seasonal cycle. To evaluate how the model performs on an image data set from a different oceanographical setting, we also included images from three sediment samples aged less than 100 years from the core B1404-11 (14.14°S, 76.50°W, 302 m) recovered from the Peruvian upwelling zone in 2014.

Sediment trap and core samples were prepared using the random settling method (Beaufort et al., 2014; Tetard et al., 2020), a protocol that can also be used to study calcareous nanofossils. A couple of milligrams of dried sediment were resuspended in water and ultrasonicated for less than a minute to remove major aggregates. Around 3 ml of the suspension was collected, left to settle for 4 hours on 12 mm × 12 mm coverslips, and left to dry overnight after pipetting the excess water. Subtracting the initial mass from the final mass of the coverslips yielded a mass of sediment on each coverslip within the range of 50–150 µg. Sets of eight coverslips were then mounted on a microscope slide using Norland Optical Adhesive 81. Each slide was mounted consecutively on a Leica DMR6000 B automated transmitted light microscope with 630x magnification using a HCX PL FLUOTAR 63× Leica lens. Images (210 µm × 210 µm) were taken using a Hamamatsu ORCA-Flash4.0 LT camera, controlled via a LabVIEW (National Instruments) interface. We acquired images for around 250 fields of view for each sample. For each field of view, we acquired 15 images of different focal lengths to image a depth of at least 100 µm. Hyperfocused stacks were created from these 15 images using the Helicon Focus 7 software (Helicon Soft).

2.2. Experimental Design for Automatic Diatom Detection Using Deep Learning Approaches

2.2.1. Training and Validation Image Data Set

The use of deep learning methods requires the creation of a labeled set of images for training. For this experiment, we randomly chose 253 images from the Lionceau image set (corresponding to 39 samples). We used the Computer Vision Annotation Tool (CVAT) developed by Intel to label these 8-bit images of 2048 × 2048 pixels.

We drew bounding boxes around all microfossil material identified on the images for a total of 10,293 bounding boxes. We attributed each bounding box to 12 different microfossil categories (i.e., silicoflagellates, diatoms,

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

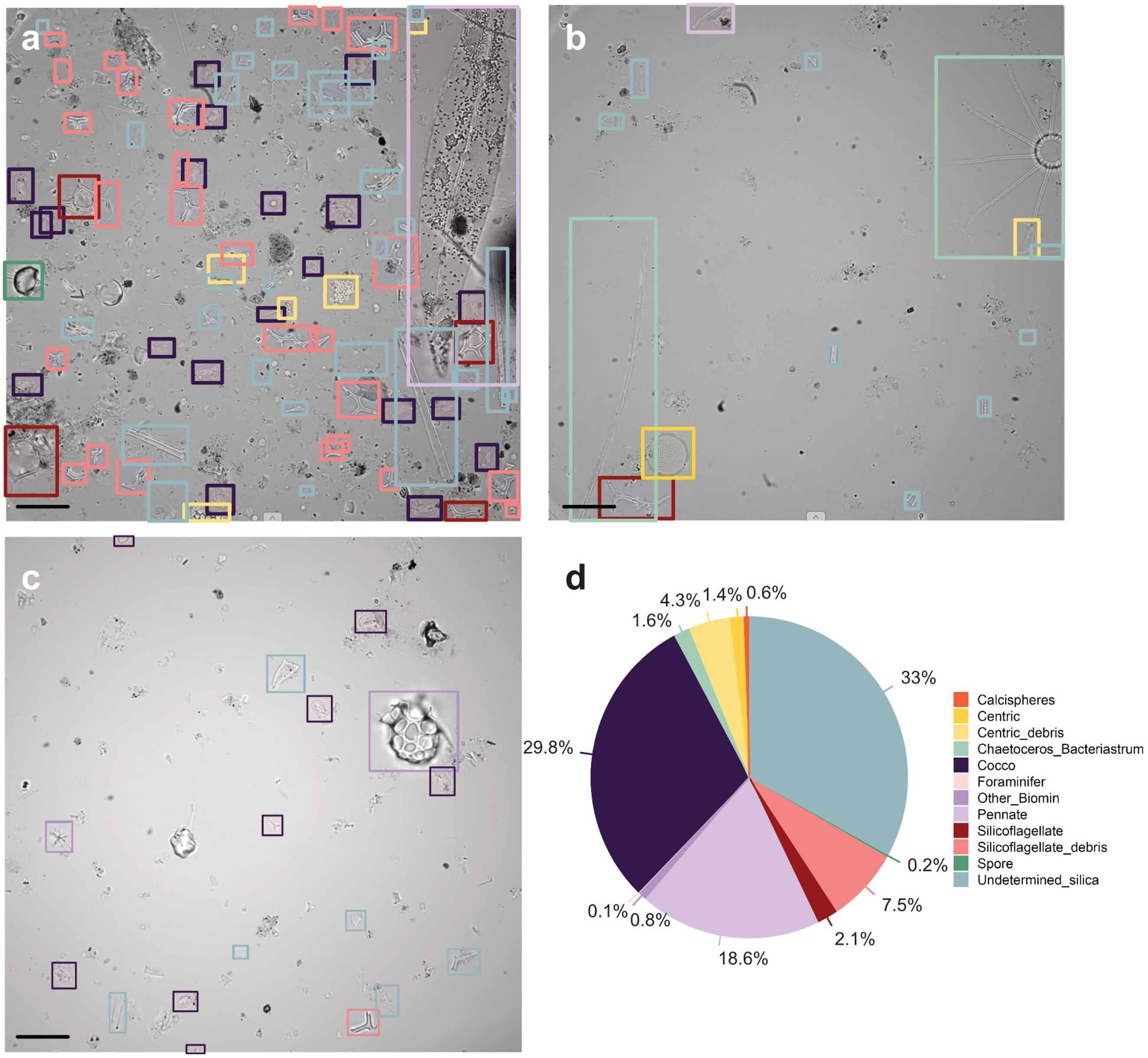

Figure 1. Computer Vision Annotation Tool screenshots of different microscope images annotated (a–c) and relative contribution of different microfossil material to the total bounding box data set (d). The color code for the bounding boxes is reported in panel (d). The models were not trained for object classification, however the training set is dominated by small fractured elements ("Undetermined_silica"), pennates (often fragmented), and large coccoliths ("Cocco"). The scale is 20 µm.

coccoliths, other biominerals, etc.; Figure 1) but pooled all the boxes into a single "microfossil" category for training. Our data set is dominated by small fragmented elements ("Undetermined_silica"), pennates (often fragmented), and large coccoliths ("Cocco") (Figure 1).

2.2.2. First Test Image Data Set—Unlabeled Images From the Mediterranean

To test the model, a second set of randomly selected images of the Lionceau sediments was manually annotated on CVAT to generate a set of ground-truth bounding boxes. This test set includes 66 annotated images for a total of 2,165 bounding boxes. These bounding boxes were attributed to one of the 12 categories used for the training set to better constrain the objects that the model might overlook. In addition, descriptive tags were associated with the images of the data set to investigate how the image or sample quality could impact the detection model. These descriptive tags account for images with one or several characteristics, including low concentration of elements,

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

high concentration of fragmented elements, presence of aggregates, image out of focus, and image darker in appearance due to the use of a different set of light parameters on the microscope. Images that did not fall into these categories were labeled as "good quality."

2.2.3. Second Test Image Data Set—Unlabeled Images From a Peruvian Sediment Core

A second test set was created to test the potential for a CNN-based model to perform diatom frustule detection for samples originating from various oceanographical settings and time periods. Specifically, we explore how the detection model trained on images from Mediterranean sediment traps generalizes to a set of images from a sediment core of the Peruvian upwelling. Although these images were acquired using a similar methodology, they contain microfossils from species that are not present in the Mediterranean data set. Similarly to the test set from the Mediterranean sediment traps, we labeled 37 images, drawing 1,868 ground-truth bounding boxes, which were attributed to 12 categories.

2.2.4. Model Training

For this experiment we chose the commonly used Faster R-CNN (faster region-based convolutional neural network) object detection model with a ResNet50 backbone (He et al., 2016). Faster R-CNN models were indeed shown to perform better on diatom images from microscope observations than other deep-learning algorithms such as You Only Look Once (YOLOv3), which did not perform very well on small objects, and Single -Shot Multi-box Detector (SSD) (Li et al., 2020). Code to train Faster R-CNN model is available on the PyTorch torchvision library (https://pytorch.org/vision/stable/index.html) and we used the example code with no modifications to the model architecture (Ren et al., 2017), and some minor changes to the training routine, such as increasing the number of detections per image.

Faster R-CNN feeds the initial image through to a backbone CNN to generate a feature map of the image. The features propose regions that may correspond to objects, and these regions are subsequently pooled before being classified into object classes using fully connected prediction layers. In our case, we use the ResNet50 CNN pre-trained on the COCO data set (Lin et al., 2014) as the backbone. Faster R-CNN provides bounding box co-ordinates and class labels for the detected objects.

The model was trained on the 253 images in the training/validation set using a split of 80% of the images as training images and the remaining 20% as validation images. Importantly, all bounding boxes drawn on CVAT were pooled into a single "microfossil" category for model training. This was intended as a way of generating a model capable of detecting any microfossil on a microscope image regardless of its species or type, including rare species for which collecting a detailed data set can be a time-consuming task.

The model was trained end-to-end (i.e., not using transfer learning) using either the Adam optimizer with an initial learning rate of 0.001 (Kingma & Ba, 2017) or Stochastic Gradient Descent (Bottou & Bousquet, 2011; LeCun et al., 2015) with an initial learning rate of 0.005, momentum of 0.9, and weight decay of 0.0005. The learning rate was monitored using the Adaptive Learning Rate Scheduler (ALRS) from Marchant et al. (2020). This method monitors the loss over the last N epochs, and if the slope of the loss is not significantly different from 0 (i.e., no longer decreasing) the learning rate is reduced (typically by half). After four reductions, training is automatically stopped. Note, we monitor the training loss, which is the average of the classification and region proposal network losses, and the validation data set is only used for hyper-parameter tuning.

Data augmentations were applied to the images during training, including random horizontal flip, random vertical flip, and random photometric distortion (brightness, contrast, saturation, hue). We experimented with different training set sizes, optimization methods, and ALRS parameters (Figures S1–S5 in Supporting Information S1). The code for the training routine is available at https://github.com/microfossil/particle-object-detection.

2.2.5. Model Testing

We used the object detection model that was most performant on the training and validation set to infer a set of "modeled" bounding boxes on our test sets (Mediterranean and Peruvian), which we compared to the "ground-truth" bounding boxes drawn manually on the test sets. The comparison of the sets of ground-truth bounding boxes with the modeled set was achieved using the bounding box coordinates. We calculated the "intersection over union" (IoU; Table 1) metric for all possible pairs of ground-truth and modeled bounding boxes of a single

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
Definition of the Metrics Used to Evaluate the Performance of the Object Detection Model

Metric	Formula	Description
Intersection over Union (IoU)	$\frac{BB_1 \cap BB_2}{BB_1 \cup BB_2}$	The ratio of area of overlap between bounding boxes 1 (BB_1) and 2 (BB_2) to the total area represented by BB_1 and BB_2.
Precision	$\frac{\sum t_p}{\sum t_p + \sum f_p}$	The number of correct predictions to the total number of predictions made by the model
Recall	$\frac{\sum t_p}{\sum t_p + \sum f_n}$	The number of correct predictions to the total number of initial ground-truth bounding boxes

image. In general, a higher Intersection over Union (IoU) score indicates a better performance, an IoU score of 1.0 being the perfect case. We matched the best pairs of bounding boxes using a cost algorithm (R package RcppHungarian (Silverman, 2022)). True positives (t_p) were calculated for the data set as the number of bounding box pairs with an IoU metric greater than or equal to 0.5. Unattributed ground-truth bounding boxes were counted as false negatives (f_n), while unattributed modeled bounding boxes were counted as false positives (f_p). With these different metrics, we could generate the model's precision and recall metrics (Table 1).

3. Results and Discussion

3.1. Training Outputs

The different model outputs can be compared using standard COCO object detection metrics (Table 2). To account for the large number of microfossils in each image, we increased the maximum number of objects from 100 to 300. The precision metric for an IoU of 0.50 is, on average, 0.717 (±0.006) across the different model runs. The mean recall for all bounding boxes across an IoU range of 0.50–0.95 is, on average, 0.537 (±0.007). The model performs better across all trainings on objects that are comparatively larger than others (i.e., area larger than 96 × 96 pixels), with precision and recall scores being consistently better for large objects than for medium (i.e., area between 32 × 32 pixels and 96 × 96 pixels) objects (Table 2).

Results show that increasing the training set size from 10,293 bounding boxes to 12,458 (+21%) (using the added annotations from the test set) slightly decreases model performance for all model metrics observed. Changing the optimizing method from SGD, used as default, to Adam increases model precision and recall on medium-sized objects, but it decreases the overall model precision for an IoU of 0.5 as well as the model recall for all objects. Changing the ALRS epochs parameter (i.e., the learning rate the scheduler watches to check if training is not improving) from 10 to 20 does not change the model precision for an IoU of 0.5 and decreases the model recall when all objects are considered.

3.2. Model Performance on a Test Set

The model trained using the base parameters (Model 1 in Table 2), which showed the best overall performance on the trap sediment images, was then applied to a test set of images from the same sediment trap series. To do so, we first used a detection threshold of 0.5 to generate the modeled bounding boxes, which corresponds to the probability score below which the detection software rejects detection. Bounding boxes were matched when the IoU metric was over 0.50 (see example outputs and cropping results in Figure 2). Of the 2,165 ground-truth bounding boxes on which the model was tested, the model recognized 1,569 objects (recall = 0.725) when a detection threshold of 0.5 was used. The model generated 565 false positive bounding boxes (precision = 0.735). Lowering the detection threshold to 0.3 increases the number of true positives (1,599) and, thus, recall (0.739) (Table 3). However, using a detection threshold of 0.3 also increases the number of false positive bounding boxes (623) and thus leads to lower model precision (0.72).

3.2.1. Controls on Model Recall

To ensure that all objects present on the raw images are efficiently cropped to be used in a further classification algorithm, we are most interested in maximizing model recall—that is, its skill at detecting any object—and identifying which of the ground-truth boxes the model misses. In detail, not all microfossil categories are detected the same way by the model (Figure 3). Results show that recall scores are better for high relief particles

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
Outputs From the Different Model Runs Performed Using Faster Region-Based Convolutional Neural Network With the Training and Validation Data Set

Input						Output			
Model number	Test description	Total images (training + validation)	Total bounding boxes	ALRS epochs	Optimizer	Precision: • IoU = 0.5 • IoU = 0.75	Average precision: @0.50:0.95 • Medium • Large	Recall: Max detections = 300 Area = all	Recall: Max detections = 300 • Medium • Large
1	Base parameters	253 (203 + 50)	10,293	Default (10)	Default (SGD)	• **0.720** • 0.531	• 0.381 • **0.574**	**0.545**	• 0.475 • 0.649
2	Increase data set size	319 (256 + 63)	12,458	Default (10)	Default (SGD)	• 0.706 • 0.508	• 0.350 • 0.571	0.525	• 0.446 • 0.637
3	Change optimizer from "SGD" to "Adam"	253 (203 + 50)	10,293	Default (10)	Adam	• 0.716 • **0.543**	• **0.382** • 0.561	0.539	• **0.479** • 0.627
4	Change ALRS epochs from 10 to 20	253 (203 + 50)	10,293	20	Default (SGD)	• **0.720** • 0.516	• 0.376 • 0.540	0.540	• 0.466 • **0.650**

Note. Output variables are standard COCO object detection metrics, modified to detect up to 300 objects in a single image. Average precision (AP) is the precision metric averaged over a range of IoUs from 0.50 to 0.95, for medium (area between 32 × 32 pixels and 96 × 96 pixels) and large (area larger than 96 × 96 pixels) objects. Recall scores are all calculated over a range of IoUs from 0.50 to 0.95. The best model result for each metric considered is shown in bold.

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

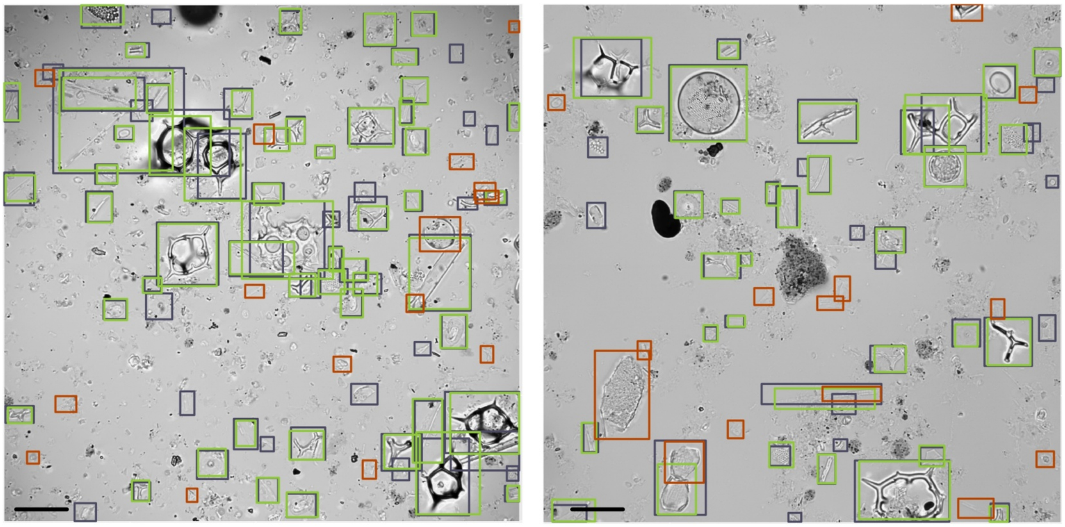

Figure 2. Prediction results on a test set using a detection threshold of 0.3. Ground-truth bounding boxes are represented in dark blue. True positive bounding boxes are in green, and false positives are shown in orange. The scale is 20 μm. Images are acquired in optical microscopy from the Lionceau sediment trap series in the Gulf of Lion. Left: sample collected in July 2011, Right: sample collected in April 2014.

such as silicoflagellates, with 100% particles identified, than for *Chaetoceros* spp. (70%) which exhibit lower relief and are prone to dissolution due to less robust frustules. The lowest recall scores are obtained for calcispheres and centric diatoms.

When considering the total category area instead of category counts, it appears that the model captures more than 80% of the ground-truth area in each category, except for calcispheres (57%), which are only present as fragments in this test data set. The model also exhibits low recall values for total coccolith area (78%) and for the total area of silica not attributed to a specific producer (75%), despite these classes dominating the training set (Figure 1). These results and the raw model outputs suggest that the object detection model works best for the larger elements across all categories. For these larger elements (i.e., large diatom frustules and silicoflagellates), these recall scores above 0.8 are comparable to the pixel-based recall results obtained using object segmentation methods on microscope images of freshwater diatoms by Kloster et al. (2023).

Model precision and recall may be affected by the quality of the sample image. We pooled all the ground-truth bounding boxes into categories corresponding to the descriptive tags listed in the Methods section. Due to multiple tags being sometimes attributed to the same image, some bounding boxes are present in different categories. We calculated precision and recall scores for each descriptive tag (Figure 4). Good-quality images display the best precision and recall scores (0.775 and 0.766 respectively). Changes in microscope acquisition parameters, which can yield darker images and/or less contrasted images, decrease precision (0.70) and, to a lesser extent, recall (0.75). A low concentration of objects of interest, which in the images observed can be attributed to some extent to dissolution, reduces precision (0.692) and recall (0.734). The presence of fragmented elements also reduces precision (0.691) and recall (0.71) scores. This could be explained by the fact that the model

Table 3
Performance of the Object Detection Model Applied to Different Test Sets

Test set	Model used	TP bounding boxes	FN bounding boxes	FP bounding boxes	Precision	Recall
Mediterranean sediment trap	Detection model with base parameters applied with a 0.5 detection threshold	1,569	596	565	0.7352	0.7247
Mediterranean sediment trap	Detection model with base parameters applied with a 0.3 detection threshold	1,599	566	623	0.720	0.739
Peruvian sediments	Detection model with base parameters applied with a 0.3 detection threshold	1,408	460	524	0.729	0.754

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

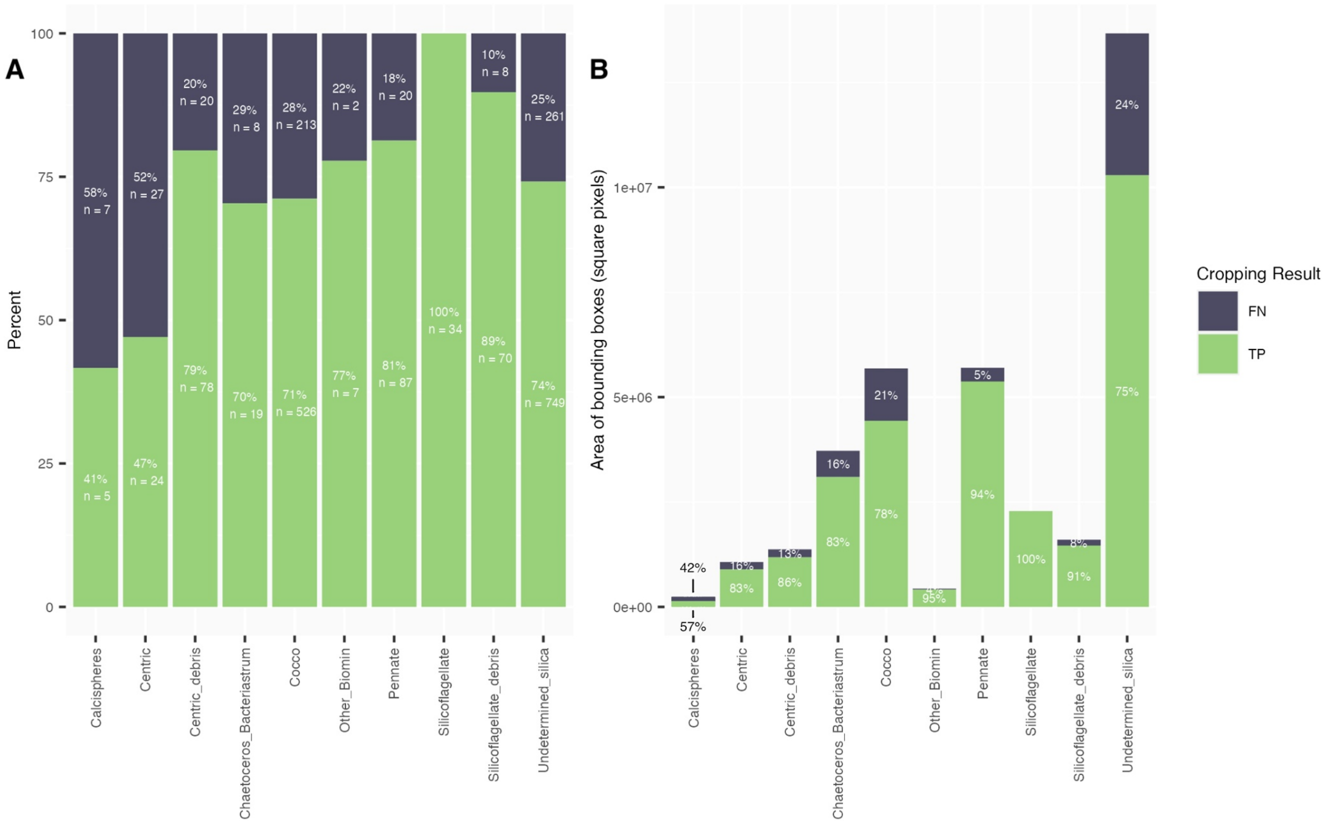

Figure 3. Object detection performance on a test set using the base model with a detection threshold of 0.3. (a) Stacked bar plot of the percentage of false negative (FN) and true positive bounding boxes obtained, per ground-truth box category. (b) Bar plot of the total predicted bounding box area per ground-truth bounding box category. The color code is the same as for panel (a). Model recall is highest for silicoflagellates and pennates. When considering the bounding box area instead of total counts, the model appears to perform better, suggesting that small elements are more likely to lead to false negatives.

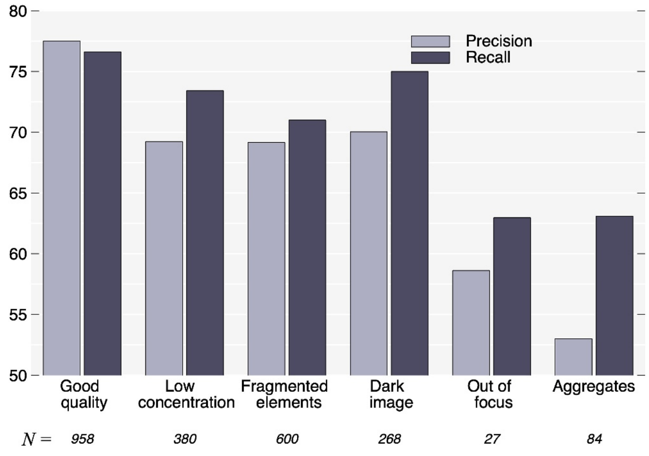

Figure 4. Object detection performance on a test set, per descriptive tag. The set of predicted bounding boxes was generated using the base model with a detection threshold of 0.3. Acquiring images with a different set of light parameters on the microscope has little effect on model recall, but may lead to a reduction in model precision. However, the detection model does not perform as well when images are out of focus, and/or have a low concentration of elements (usually linked to dissolution), and/or include fragments and/or aggregates.

appears to perform more poorly on small objects. Images that are out of focus generate more false positives and negatives, with precision and recall scores of 0.586 and 0.63. The presence of aggregates on the images, which lead to more objects being superimposed, also impacts the quality of object detection, with precision and recall scores of 0.53 and 0.631.

3.2.2. Model Precision

Model precision is low when the model generates many false positive bounding boxes. One issue with low model precision in a detection task is the generation of multiple bounding boxes per microfossil (Figure 2). This can lead to certain particles being transferred multiple times to a classification algorithm and thus counted twice. We reviewed the false positive bounding boxes generated using a detection threshold of 0.3 (Figure 5) into five categories to evaluate this bias. Results show that 30 of the 623 false positives are duplicate bounding boxes, which represents 1.4% of the ground-truth bounding boxes. Therefore, it appears that the generation of multiple bounding boxes of the same microfossil occurs marginally in our data set.

We find that 30 of the f_p bounding boxes generated by the model partially overlap a ground-truth bounding box. However, they are not counted as t_p due to low IoU scores (<0.5): the predicted bounding box may capture only part of the microfossil ($n = 16$, Figure 5) or be too large or shifted ($n = 14$) compared to the ground-truth bounding boxes. Additionally, 37 of the f_p bounding boxes

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

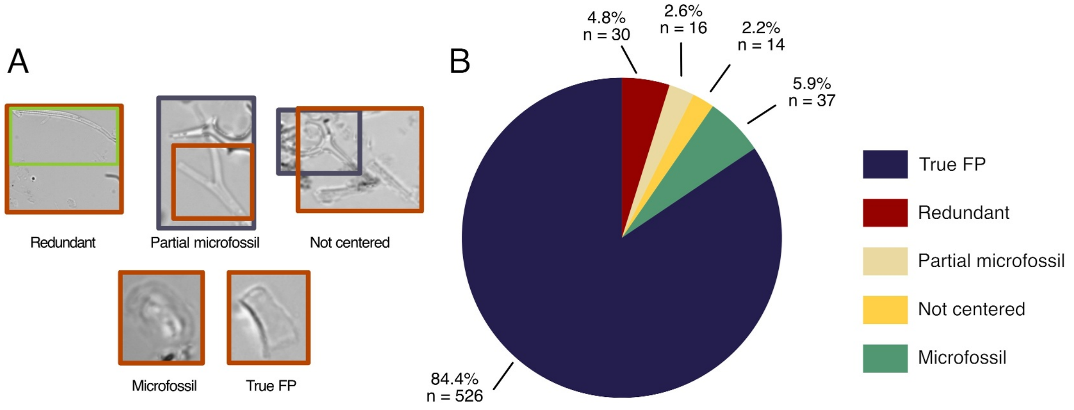

Figure 5. False positive outputs for the test data set. (a) Classification scheme for the false positive bounding boxes; see Figure 2 for color scheme explanation. (b) Distribution of false positive bounding boxes.

are actual microfossils that were overlooked during annotation. The remaining 526 f_p bounding boxes do not correspond to a microfossil and can easily be eliminated in a classification step.

3.3. Transfer to a New Data Set

The model was initially trained on a relatively limited number of images from a single oceanic setting, and could thus be expected to be less accurate when applied to images featuring different species and preservation states. Indeed, deep neural networks have a limited capability of transferring to image data sets which differ from the training set (Serre, 2019). To evaluate how the model performed on images acquired by the same method but with samples from a different oceanic setting, we constituted a second test set with images from a sediment core retrieved off the Peruvian coast. We applied the base model with a detection threshold of 0.3 and compared the resulting bounding boxes to a set of ground-truth bounding boxes using the same matching method as for the Mediterranean sediment trap test data set (see Methods Section 2.2.5.).

The comparison of these two bounding box data sets yields precision and recall scores of 0.729 and 0.754 respectively. The model thus performs slightly better on this new test set of images from the Peruvian setting than on the test set of images from the Mediterranean trap (Table 3). Multiple factors might explain this result, such as (a) the smaller abundance of coccoliths in these samples, which were shown to be less well captured by the model, (b) the good preservation of these diatom-rich sediments formed in an environment with a high accumulation rate, and (c) a smaller amount of out-of-focus images than in the Mediterranean test data set.

As for the Mediterranean data set, the model performs better on the largest particles present on the images, especially on silicoflagellates. Interestingly, the model detects plankton species and morphospecies that are absent from the training set, including the silicoflagellate *Octactis* genus and the diatom genus *Actinoptychus* (Figure 6). However, the model performs poorly on detecting diatom resting spores, which are nearly absent from the training set (Figure 1). These results suggest that this CNN-based object detection protocol has the potential to be used across a variety of image data sets from sediment samples but could still benefit from being trained specifically on some particles not captured by our training set.

4. Conclusions

In this study, we developed an object detection workflow using Faster R-CNN model which allows for the automatic detection of diatom frustules on sample slides. Given their non-birefringent nature, wide variety in shapes and sizes, and tendency to overlap, diatom frustules remain challenging to detect and crop with traditional thresholding techniques. The method showed promising results for detecting diatom frustules on microscope images, with a precision score reaching 0.72 for an IoU of 0.5, and a corresponding recall score of 0.58 for IoU scores ranging from 0.5 to 0.95. The comparison of modeled bounding boxes with a test set showed that, for the generally accepted IoU threshold of 0.5, the recall for the model reaches an acceptable level for analyzing real and complex samples, with a value of 0.73. Large and high relief particles are well detected by the Faster R-CNN (Figure 3) despite making up a smaller fraction of the training data set. Our workflow can deal with

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

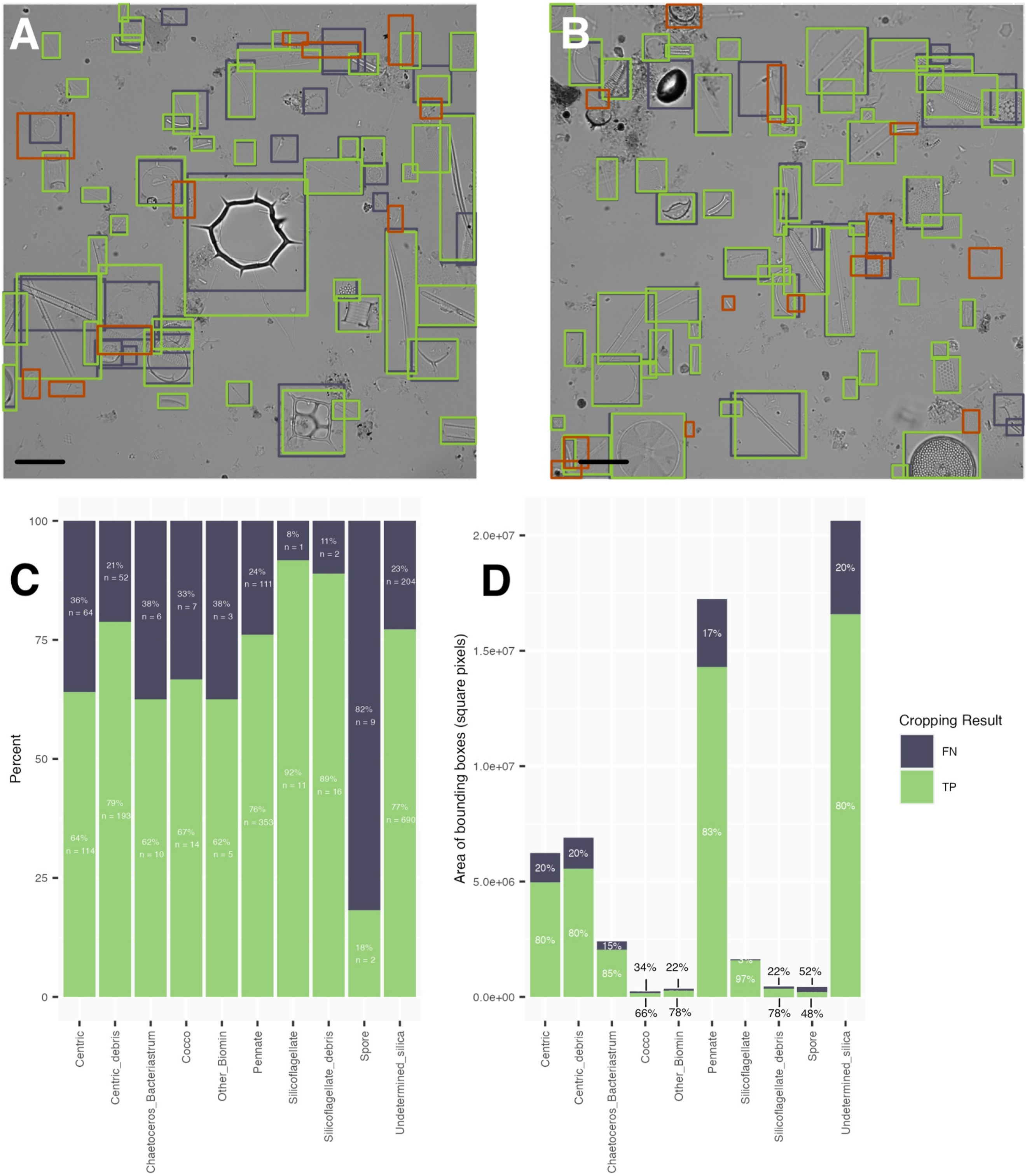

Figure 6. Prediction results on an image data set from the Peruvian upwelling zone. (a)–(b): Ground-truth bounding boxes are represented in dark blue. True positive (TP) bounding boxes are in green, and false positives are shown in orange. The scale is 20 µm. (c) Stacked bar plot of the percentage of false negative and TP bounding boxes obtained, per ground-truth box category. (d) Bar plot of the total predicted bounding box area per ground-truth bounding box category. The color code is the same as for panel (a).

overlapping particles, however the proposed method is not completely efficient on large aggregates (Figure 4). Ensuring properly focused microscope images and reducing the number of images containing fragments and dissolved elements minimizes potential errors in the detection process.

Decoupling the detection from the identification makes our method interesting for the detection of rare species which may be under-represented in the training data set. A possible solution to combine detection and identification into a single deep-learning model would be to test the use of virtual slides to artificially increase the instances of rare microfossils in the training set (Venkataramanan et al., 2023), however this solution would still require the acquisition of a large image data set to cover the full diversity spectrum.

Prospects for this particular workflow include incorporating this detection and cropping algorithm into a more general identification workflow, specific to siliceous biominerals. The microfossil crops generated from the microscope images by the workflow can be uploaded into open-access computer programs such as ParticleTrieur (Marchant et al., 2020) for labeling and training a classification model. The microscope slides processed in this workflow are also suitable for analysis by the SYRACO software to obtain coccolith counts and morphology (Beaufort & Dollfus, 2004) and radiolarian analysis (Tetard et al., 2020). This makes it possible to study the changes in species composition within and between different groups of marine phytoplankton, a critical step when studying biogeochemical cycles and changes in planktonic communities.

This workflow has the potential to detect microfossils present on sediment samples from different spatial and temporal settings, both from sediment traps and sediment coring systems (Figure 6). Future work will focus on enriching the image database used for training to include microfossils that require better detection, to obtain a model applicable to any sedimentary sample containing siliceous biominerals.

Data Availability Statement

The code for the object detection workflow is available using the https://doi.org/10.5281/zenodo.10591771 (geometrikal, 2024). The data sets for image annotation as well as the training results for the different models can be found online on Zenodo with the following https://doi.org/10.5281/zenodo.12665681.

Acknowledgments
We would like to thank Xavier Durrieu de Madron of the CEFREM for providing the Lionceau sediment trap samples. We would like to thank Jean-Charles Mazur and Sandrine Conrod for help with the sample preparation, and Julien Lecubin and the SIP team at OSU Pythéas for help setting up the servers. This work received support from the French government under the France 2030 investment plan, as part of the Initiative d'Excellence d'Aix-Marseille Université–A*MIDEX (AMX-19-IET-012).

References

Abrantes, F., Lopes, C., Mix, A., & Pisias, N. (2007). Diatoms in Southeast Pacific surface sediments reflect environmental properties. *Quaternary Science Reviews*, *26*(1–2), 155–169. https://doi.org/10.1016/j.quascirev.2006.02.022

Armbrust, E. V. (2009). The life of diatoms in the world's oceans. *Nature*, *459*(7244), 185–192. https://doi.org/10.1038/nature08057

Beaufort, L., Barbarin, N., & Gally, Y. (2014). Optical measurements to determine the thickness of calcite crystals and the mass of thin carbonate particles such as coccoliths. *Nature Protocols*, *9*(3), 633–642. https://doi.org/10.1038/nprot.2014.028

Beaufort, L., & Dollfus, D. (2004). Automatic recognition of coccoliths by dynamical neural networks. *Marine Micropaleontology*, *51*(1–2), 57–73. https://doi.org/10.1016/j.marmicro.2003.09.003

Beaufort, L., Gally, Y., Suchéras-Marx, B., Ferrand, P., & Duboisset, J. (2021). Technical note: A universal method for measuring the thickness of microscopic calcite crystals, based on bidirectional circular polarization. *Biogeosciences*, *18*(3), 775–785. https://doi.org/10.5194/bg-18-775-2021

Borowiec, M. L., Dikow, R. B., Frandsen, P. B., McKeeken, A., Valentini, G., & White, A. E. (2022). Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*, *13*(8), 1640–1660. https://doi.org/10.1111/2041-210X.13901

Bottou, L., & Bousquet, O. (2011). The tradeoffs of large-scale learning. In S. Sra, S. Nowozin, & S. J. Wright (Eds.), *Optimization for machine learning* (pp. 351–368). The MIT Press. https://doi.org/10.7551/mitpress/8996.003.0015

Bourel, B., Marchant, R., De Garidel-Thoron, T., Tetard, M., Barboni, D., Gally, Y., & Beaufort, L. (2020). Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains. *Computers & Geosciences*, *140*, 104498. https://doi.org/10.1016/j.cageo.2020.104498

Bueno, G., Deniz, O., Pedraza, A., Ruiz-Santaquiteria, J., Salido, J., Cristóbal, G., et al. (2017). Automated diatom classification (Part A): Handcrafted feature approaches. *Applied Sciences*, *7*(8), 753. https://doi.org/10.3390/app7080753

Bueno, G., Déniz, O., Ruiz-Santaquiteria, J., Olenici, A., Cristobal, G., Pedraza, A., et al. (2018). Lights and pitfalls of convolutional neural networks for diatom identification. In P. Schelkens, T. Ebrahimi, & G. Cristóbal (Eds.), *Optics, photonics, and digital technologies for imaging applications V* (p. 14). SPIE. https://doi.org/10.1117/12.2309488

Carlsson, V., Danelian, T., Tetard, M., Meunier, M., Boulet, P., Devienne, P., & Ventalon, S. (2023). Convolutional neural network application on a new middle Eocene radiolarian dataset. *Marine Micropaleontology*, *183*, 102268. https://doi.org/10.1016/j.marmicro.2023.102268

Claustre, H. (1994). The trophic status of various oceanic provinces as revealed by phytoplankton pigment signatures. *Limnology & Oceanography*, *39*(5), 1206–1210. https://doi.org/10.4319/lo.1994.39.5.1206

Computer Vision Annotation Tool (CVAT) (2.8.2). (2023). CVAT.ai Corporation [Computer software]. https://github.com/opencv/cvat

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237), 1261605. https://doi.org/10.1126/science.1261605

geometrikal. (2024). microfossil/particle-object-detection: Paper Version (paper_v1). *Zenodo*. https://doi.org/10.5281/zenodo.10591772

Gimenez, B., Joannin, S., Pasquet, J., Beaufort, L., Gally, Y., De Garidel-Thoron, T., et al. (2024). A user-friendly method to get automated pollen analysis from environmental samples. *New Phytologist*, *243*(2), 797–810. https://doi.org/10.1111/nph.19857

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Henson, S. A., Cael, B. B., Allen, S. R., & Dutkiewicz, S. (2021). Future phytoplankton diversity in a changing climate. *Nature Communications*, *12*(1), 5372. https://doi.org/10.1038/s41467-021-25699-w

Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T. J., & Barlow, R. G. (2008). An absorption model to determine phytoplankton size classes from satellite ocean colour. *Remote Sensing of Environment*, *112*(6), 3153–3159. https://doi.org/10.1016/j.rse.2008.03.011

Iriarte, J., & González, H. (2004). Phytoplankton size structure during and after the 1997/98 El Niño in a coastal upwelling area of the northern Humboldt Current System. *Marine Ecology Progress Series*, *269*, 83–90. https://doi.org/10.3354/meps269083

Jin, X., Gruber, N., Dunne, J. P., Sarmiento, J. L., & Armstrong, R. A. (2006). Diagnosing the contribution of phytoplankton functional groups to the production and export of particulate organic carbon, $CaCO_3$, and opal from global nutrient and alkalinity distributions. *Global Biogeochemical Cycles*, *20*(2), 2005GB002532. https://doi.org/10.1029/2005GB002532

Kemp, A. E. S., & Villareal, T. A. (2018). The case of the diatoms and the muddled mandalas: Time to recognize diatom adaptations to stratified waters. *Progress in Oceanography*, *167*, 138–149. https://doi.org/10.1016/j.pocean.2018.08.002

Kingma, D. P., & Ba, J. (2017). *Adam: A Method for stochastic optimization* (arXiv:1412.6980). *arXiv*. http://arxiv.org/abs/1412.6980

Kloster, M., Burfeid-Castellanos, A. M., Langenkämper, D., Nattkemper, T. W., & Beszteri, B. (2023). Improving deep learning-based segmentation of diatoms in gigapixel-sized virtual slides by object-based tile positioning and object integrity constraint. *PLoS One*, *18*(2), e0272103. https://doi.org/10.1371/journal.pone.0272103

Kloster, M., Kauer, G., & Beszteri, B. (2014). SHERPA: An image segmentation and outline feature extraction tool for diatoms and other objects. *BMC Bioinformatics*, *15*(1), 218. https://doi.org/10.1186/1471-2105-15-218

Kloster, M., Langenkämper, D., Zurowietz, M., Beszteri, B., & Nattkemper, T. W. (2020). Deep learning-based diatom taxonomy on virtual slides. *Scientific Reports*, *10*(1), 14416. https://doi.org/10.1038/s41598-020-71165-w

Kohfeld, K., Le Quéré, C., Harrison, S. P., & Anderson, R. F. (2005). Role of marine biology in glacial-interglacial CO_2 cycles. *Science*, *308*(5718), 74–78. https://doi.org/10.1126/science.1105375

Leblanc, K., Quéguiner, B., Diaz, F., Cornet, V., Michel-Rodriguez, M., Durrieu de Madron, X., et al. (2018). Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. *Nature Communications*, *9*(1), 953. https://doi.org/10.1038/s41467-018-03376-9

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Le Quéré, C., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., et al. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, *11*(11), 2016–2040. https://doi.org/10.1111/j.1365-2486.2005.1004.x

Li, Q., Sun, X., Dong, J., Song, S., Zhang, T., Liu, D., et al. (2020). Developing a microscopic image dataset in support of intelligent phytoplankton detection using deep learning. *ICES Journal of Marine Science*, *77*(4), 1427–1439. https://doi.org/10.1093/icesjms/fsz171

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). Microsoft COCO: Common objects in context. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer vision – ECCV 2014* (Vol. 8693, pp. 740–755). Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48

Lombard, F., Boss, E., Waite, A. M., Vogt, M., Uitz, J., Stemmann, L., et al. (2019). Globally consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science*, *6*, 196. https://doi.org/10.3389/fmars.2019.00196

Marchant, R., Tetard, M., Pratiwi, A., Adebayo, M., & De Garidel-Thoron, T. (2020). Automated analysis of foraminifera fossil records by image classification using a convolutional neural network. *Journal of Micropalaeontology*, *39*(2), 183–202. https://doi.org/10.5194/jm-39-183-2020

Margalef, R. (1978). Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanologica Acta*, *1*(4), 493–509.

Marino, M., Maiorano, P., Tarantino, F., Voelker, A., Capotondi, L., Girone, A., et al. (2014). Coccolithophores as proxy of seawater changes at orbital-to-millennial scale during middle Pleistocene Marine Isotope Stages 14-9 in North Atlantic core MD01-2446. *Paleoceanography*, *29*(6), 518–532. https://doi.org/10.1002/2013PA002574

Mouw, C. B., Hardman-Mountford, N. J., Alvain, S., Bracher, A., Brewin, R. J. W., Bricaud, A., et al. (2017). A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the global ocean. *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00041

Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M.-H., et al. (2008). Remote sensing of phytoplankton functional types. *Remote Sensing of Environment*, *112*(8), 3366–3375. https://doi.org/10.1016/j.rse.2008.01.021

Passow, U., & Carlson, C. A. (2012). The biological pump in a high CO_2 world. *Marine Ecology Progress Series*, *470*(2), 249–271. https://doi.org/10.3354/meps09985

Petit, F., Uitz, J., Schmechtig, C., Dimier, C., Ras, J., Poteau, A., et al. (2022). Influence of the phytoplankton community composition on the in situ fluorescence signal: Implication for an improved estimation of the chlorophyll-a concentration from BioGeoChemical-Argo profiling floats. *Frontiers in Marine Science*, *9*, 959131. https://doi.org/10.3389/fmars.2022.959131

Picheral, M., Colin, S., & Irisson, J.-O. (2017). EcoTaxa, a tool for the taxonomic classification of images [Computer software]. http://ecotaxa.obs-vlfr.fr

Ragueneau, O., Schultes, S., Bidle, K., Claquin, P., & Moriceau, B. (2006). Si and C interactions in the world ocean: Importance of ecological processes and implications for the role of diatoms in the biological pump. *Global Biogeochemical Cycles*, *20*(4), 1–15. https://doi.org/10.1029/2006GB002688

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Rigual-Hernández, A. S., Bárcena, M. A., Jordan, R. W., Sierro, F. J., Flores, J. A., Meier, K. J. S., et al. (2013). Diatom fluxes in the NW Mediterranean: Evidence from a 12-year sediment trap record and surficial sediments. *Journal of Plankton Research*, *35*(5), 1109–1125. https://doi.org/10.1093/plankt/fbt055

Serre, T. (2019). Deep learning: The good, the bad, and the ugly. *Annual Review of Vision Science*, *5*(1), 399–426. https://doi.org/10.1146/annurev-vision-091718-014951

Silverman, J. (2022). RcppHungarian: Solves minimum cost bipartite matching problems (0.2) [Computer software]. https://CRAN.R-project.org/package=RcppHungarian

Tetard, M., Marchant, R., Cortese, G., Gally, Y., de Garidel-Thoron, T., & Beaufort, L. (2020). Technical note: A new automated radiolarian image acquisition, stacking, processing, segmentation and identification workflow. *Climate of the Past*, *16*(6), 2415–2429. https://doi.org/10.5194/cp-16-2415-2020

Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., et al. (2018). Influence of diatom diversity on the ocean biological carbon pump. *Nature Geoscience*, *11*(1), 27–37. https://doi.org/10.1038/s41561-017-0028-x

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Venkataramanan, A., Faure-Giovagnoli, P., Regan, C., Heudre, D., Figus, C., Usseglio-Polatera, P., et al. (2023). Usefulness of synthetic datasets for diatom automatic detection using a deep-learning approach. *Engineering Applications of Artificial Intelligence*, *117*, 105594. https://doi.org/10.1016/j.engappai.2022.105594

Yu, W., Xiang, Q., Hu, Y., Du, Y., Kang, X., Zheng, D., et al. (2022). An improved automated diatom detection method based on YOLOv5 framework and its preliminary study for taxonomy recognition in the forensic diatom test. *Frontiers in Microbiology*, *13*, 963059. https://doi.org/10.3389/fmicb.2022.963059

Zingone, A., Harrison, P. J., Kraberg, A., Lehtinen, S., McQuatters-Gollop, A., O'Brien, T., et al. (2015). Increasing the quality, comparability and accessibility of phytoplankton species composition time-series data. *Estuarine, Coastal and Shelf Science*, *162*, 151–160. https://doi.org/10.1016/j.ecss.2015.05.024

21698961, 2024, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008047 by Camille Godbillot - Cochrane Peru , Wiley Online Library on [25/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License