PLOS ONE

RESEARCH ARTICLE

Enhanced convolutional neural network for plankton identification and enumeration

Kaichang Cheng[1], Xuemin Cheng[1]*, Yuqi Wang[1], Hongsheng Bi[2]*, Mark C. Benfield[3]

1 Graduate School at Shenzhen, Tsinghua University, Shenzhen, Guangdong, P.R. China, **2** Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, Solomons, Maryland, United States of America, **3** Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana, United States of America

* chengxm@sz.tsinghua.edu.cn (XC); hbi@umces.edu (HB)

Citation: Cheng K, Cheng X, Wang Y, Bi H, Benfield MC (2019) Enhanced convolutional neural network for plankton identification and enumeration. PLoS ONE 14(7): e0219570. https://doi.org/10.1371/journal.pone.0219570

Editor: Jie Zhang, Newcastle University, UNITED KINGDOM

Received: March 1, 2019

Accepted: June 26, 2019

Published: July 10, 2019

Data Availability Statement: The minimal underlying in situ plankton dataset, pre-trained models and other data used in the manuscript can be accessed at https://doi.org/10.6084/m9.figshare.8146283. Laboratory protocols are in protocols.io, and its DOI is https://dx.doi.org/10.17504/protocols.io.2u5gey6. The necessary MATLAB functions can also be accessed at https://github.com/KaichangCHENG/PIE-MC/tree/master/EnhancedCNN. Additional raw images and data for validation can be obtained by contacting flyinskyxm@hotmail.com or author Xuemin Cheng (chengxm@sz.tsinghua.edu.cn). The relevant DOIs

Abstract

Despite the rapid increase in the number and applications of plankton imaging systems in marine science, processing large numbers of images remains a major challenge due to large variations in image content and quality in different marine environments. We constructed an automatic plankton image recognition and enumeration system using an enhanced Convolutional Neural Network (CNN) and examined the performance of different network structures on automatic plankton image classification. The procedure started with an adaptive thresholding approach to extract Region of Interest (ROIs) from *in situ* plankton images, followed by a procedure to suppress the background noise and enhance target features for each extracted ROI. The enhanced ROIs were classified into seven categories by a pre-trained classifier which was a combination of a CNN and a Support Vector Machine (SVM). The CNN was selected to improve feature description and the SVM was utilized to improve classification accuracy. A series of comparison experiments were then conducted to test the effectiveness of the pre-trained classifier including the combination of CNN and SVM versus CNN alone, and the performance of different CNN models. Compared to CNN model alone, the combination of CNN and SVM increased classification accuracy and recall rate by 7.13% and 6.41%, respectively. Among the selected CNN models, the ResNet50 performed the best with accuracy and recall at 94.52% and 94.13% respectively. The present study demonstrates that deep learning technique can improve plankton image recognition and that the results can provide useful information on the selection of different CNN models for plankton recognition. The proposed algorithm could be generally applied to images acquired from different imaging systems.

Introduction

Zooplankton play a pivotal role in marine ecosystems by feeding on phytoplankton and serving as important food for fish larvae [1]. Understanding their spatial and temporal dynamics and interactions with their environment remain fundamental questions in plankton ecology [2]. In recent years, imaging techniques have contributed greatly to our understanding of fine-

and URLs of all the data and codes used are also included in the corresponding positions of the revised manuscript.

Funding: This work was supported by The National Key Research and Development Program of China (2017YFC1403602), the Shenzhen Science and Technology Innovation Program (Nos. JCYJ20160428182026575, Nos. JCYJ20170412171011187). Bi was supported by the National Science Foundation (NSF #1602488).

Competing interests: The authors have declared that no competing interests exist.

scale plankton distributions and their interactions with their environments [3]. The number of imaging systems and applications have greatly increased in recent years, e.g., Video Plankton Recorder(VPR) [4], Underwater Vision Profiler (UVP) [5], ZOOplankton VISualization system (ZOOVIS) [6], Shadow Image Particle Profiling Evaluation Recorder (SIPPER) [7], and the In Situ Ichthyoplankton Imaging System (ISIIS) [8]. However, *in situ* plankton images are often acquired under sub-optimal conditions, e.g., particulates, turbidity and currents, which can affect light attenuation and scattering and lead to less than ideal image quality [9]. Moreover, imaging systems are capable of generating very large numbers of unique image files [10]. Extracting useful information from a large number of *in situ* plankton images acquired by underwater imaging systems in a timely manner remains a challenge [10, 11]. This latter challenge is essential if *in situ* imaging systems are to achieve their full potential as operational sampling, monitoring, and forecasting tools.

The need for automated plankton recognition and enumeration has spurred the development of new image processing techniques in the past two decades. A typical procedure involves identification and extraction of Regions of Interest (ROIs), feature description, and finally classification into taxonomic categories [10]. For ROI extraction, a common approach is the Otsu global threshold method [12], in which *in situ* images were converted to binary images based on a single threshold value and then connected pixels, i.e., potential targets, were segmented. Another common approach for ROI extraction is the local threshold method represented by the Sauvola's method [13] in which *in situ* images are converted to binary images using threshold values estimated for different regions. Some techniques also incorporate spatial filtering [14] and color information [15–18] to facilitate ROI extractions. However, for images acquired from highly turbid water, a more complex approach is required. Bi et al. [9] demonstrated that it is more effective to combine Maximally Stable Extremal Regions (MSER) [19] for relatively large targets like jellyfish and the Sauvola's method to segment small targets like copepods.

Effective feature extraction and description are essential to ensure the success of an automated plankton recognition procedure. Early work often applied feature descriptions based on plankton morphology, for example, a combination of different geometric features [20–22], which appeared to perform well for images acquired under laboratory conditions. Zheng et al. [23] used a basic Local Binary Pattern (LBP) method to describe the texture features of plankton and achieved reasonable results on the microscopic benchmark plankton dataset. Recently many feature descriptors based on local features have been applied in plankton recognition, for example, Histogram of Oriented Gradient (HOG) [9, 23], Scale-Invariant Feature Transform (SIFT) [23, 24] and Shape Context [25].

Classification is the final step in automated plankton recognition in which, each ROI is assigned into one of a number of different classes. Early classification was often based on differential distance measurement, e.g., distance between eigenvectors of feature descriptors [20]. With the recent developments in machine learning, more sophisticated approaches such as Artificial Neural Networks (ANN) [26], random forest classifiers [27], Bayesian approaches [28], and Support Vector Machines (SVM) [9, 29, 30] have been applied to plankton classification. Almost all these existing methods are customized descriptors that achieve the invariance by pre-selected rules and are consequently, not flexible enough to accommodate large variations in image quality and content in plankton images, e.g., morphological variation in the target objects caused by non-uniform illumination. Deep learning methods have been used effectively to provide substantial improvements in image processing and feature extraction and appeared to be good candidates for plankton recognition.

Convolutional Neural Networks (CNN) are a common, deep learning approach, which combines feature description and classification to achieve better performance in various

classification tasks. CNN is a weight-sharing network based on image convolution [31]. The convolution result contains a convolution kernel and output. The kernel matches the image features and can be activated for amplification and the output can be used for image classification. The number of CNN models and their capabilities of processing complex images have increased rapidly since the first CNN model (LeNet) was introduced by LeCun et al. [32]. In the past few years, CNNs have moved towards deeper networks to extract complex features and increase accuracy, but increasing network depth often leads to gradient diffusion, which is problematic [32–35]. To overcome the gradient diffusion problem, He et al. [36] introduced a residual block in the neural network, the ResNet. In this new model, convergence speed and identification accuracy were increased by introducing shortcut connections between parameter layers. Another recent development in CNN models is the Dense Convolutional Network (DenseNet) [37], which connects each layer to every other layer in a feed-forward fashion. This alleviates gradient loss and reuses features learned before as depth increases. The disadvantage is that DenseNet implementation can require a large amount of memory.

The application of CNNs to plankton recognition and enumeration is still relatively new. Ouyang et al. [38] implemented a multi-size image sensing module and a deep CNN to identify 121 plankton species from 2015 National Data Science Bowl. Li et al. [39] employed deep ResNet to identify these 121 types of plankton, too. However, both studies used images acquired under ideal imaging conditions. In contrast, processing *in situ* plankton images is more challenging. Luo et al. [40] applied a sparse CNN model to process images acquired from the In Situ Ichthyoplankton Imaging System (ISIIS). The advantage of a sparse CNN is that each sparse convolutional layer can be performed with a few convolution kernels followed by a sparse matrix multiplication, which leads to higher computation efficiency [41].

The effectiveness of a CNN model for plankton recognition could affected by the presence of substantial noise due to less than ideal imaging conditions, and ambiguous features in the descriptors, boundaries, and internal features. Meanwhile, low plankton abundances, i.e., low number of occupied pixels by potential target objects would clearly exacerbate the feature problem. For example, an *in situ* image from the ZOOplankton VISualization (ZOOVIS) system has dimensions of 2448×2044 pixels and even large planktonic organisms such as small jellyfish have average pixel dimensions $< 200\times200$ [42]. After multiple layers of convolution and pooling, the detailed features will continuously disappear, resulting in a gradual reduction in the number of valid features retained. Lastly, the diversity of plankton in the oceans is high and many plankton species share similar morphological features. The lack of clearly defined features could be exacerbated by the less than ideal imaging conditions which makes it difficult to apply a model with a deep network structure. CNNs have an unrivaled feature description capability, and the key for an successful implementation of a CNN for *in situ* plankton images is to reduce the impact of unambiguous features at the input end of CNNs, i.e., improve morphological feature for better feature description, and adopt a more plankton-targeted classifier at the output end of CNNs.

In this paper, we implemented an end-to-end *in situ* plankton image identification and enumeration method. To reduce the impact of unambiguous features, we developed a specialized adaptive ROI extraction and feature enhancement procedure. For better classification, we built a multi-class SVM model to achieve global optimization of the learned feature among different target groups. To identify the proper CNN model for plankton recognition, we compared the performance of several readily available CNNs on the same plankton dataset including AlexNet [31], VGGNets [43], GoogleNet [44], and ResNet. AlexNet model [31] was proposed in 2012, which contains 5 convolution layers and 3 fully connected layers, and this model used big data, Graphics Processing Unit (GPU), Rectified Linear Units (ReLU), and dropout techniques to accelerate network convergence speed while simultaneously preventing

overfitting. VGGNets [43] have fewer parameters by performing multiple continuous small-scale convolutions instead of one step of large-scale convolution to gain more non-linear expressions and achieve a better performance with less parameters. GoogLeNet model [44] used the inception structure to replace simple traditional operations of convolution and activation to better tackle the large variation in ROIs to achieve better performance. The core idea of this model is to use wide inception structure to make models automatically adapt to features at different scales. ResNet implemented shortcut connections and showed advantages in convergence speed and identification accuracy [35, 36]. The introduction of shortcut guarantees that the model makes full use of network residual information, which makes the topology of the network more complex, and improves the performance of the model with a much deeper layers.

Methods

Data description

The plankton images used in this study were obtained by the ZOOVIS underwater *in situ* imaging system that was deployed in the southeastern Bering Sea in May 2017. The datasets used for training and testing were obtained from these *in situ* images (685,520 in total) using the ROI extraction procedure described below. The segmented images were manually separated into different taxa and verified by one of the authors (HB). The sample datasets (Fig 1) used in the experiments included 6 plankton categories (chaetognatha, copepoda, medusae, euphausiids, fish larvae, and limacina) and an 'other' category to accommodate zooplankton other than the six primary categories and non-zooplankton particles. Constructing and training CNN models requires tens of thousands of images which is problematic for rare plankton classes. For example, there were insufficient chaetognatha, medusae, euphausiids and fish larvae present in the data to establish a training set with an adequate number of images. Under these circumstances, we rotated and mirrored existing ROIs and artificially created ROIs to mimic real ROIs. For example, the back view of a copepods could be rotated to represent different orientations. The treatment details for rare classes varied slightly due to difference in their occurrence (euphausiids > chaetognatha and medusae > fish larvae). Each euphausiid ROI was rotated by 90 degree and 180 degree clockwise, and mirrored from up to down, which led to a threefold increase in sample size. Each chaetognatha and medusae ROI of them was rotated by 90 degree and 180 degree, and mirrored from up to down and left to right, which leads to a fourfold increase in sample size. The number of fish larvae ROI had the least occurance, so in addition to the rotation by 90 degree and 180 degree, and mirror from up to down and from left to right, the contrast of each ROI of this category is adjusted with a wider dynamic range of grayscale so as to mimic different imaging environment, which led to a fivefold increase in sample size. All this sample expansion methods are best chosen to mimic the real underwater *in situ* imaging environment. Using this approach we were able to ensure that

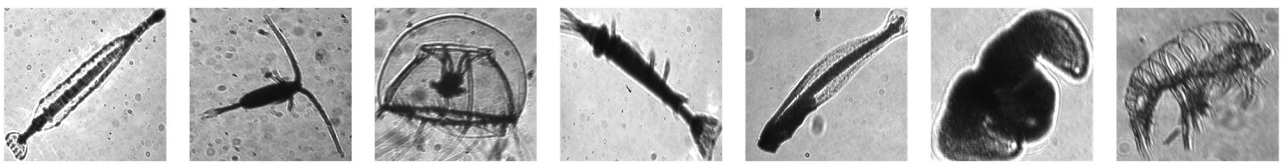

Fig 1. Example images of 7 classes in our *in situ* plankton dataset. (*a*) chaetognatha; (*b*)copepoda; (*c*) medusae; (*d*) euphausiids; (*e*) fish larvae; (*f*) Limacina; (*g*) other zooplankton or objects.

https://doi.org/10.1371/journal.pone.0219570.g001

there were 2048 ROIs in the training set and 512 ROIs in the testing set for each class and therefore we could train a balanced and unbiased classifier. To ensure an independent test, all the samples used in the training set did not appear in the testing set. The expanded training data set and testing test are to take into account various motion patterns, different oritentations and various underwater imaging environment. A more inclusive sample library could mimic plankton details in real water bodies and allow more accurate and realistic plankton feature description. Finally, we adjusted the contrast level (δ) to mimic different imaging conditions. We chose 5 contrast levels: 3.1, 3.3, 3.5, 3.7, and 3.9 based on imaging conditions presented in the dataset. At this point, the number of samples in the training set was 10,240 for each class. The dataset of in situ plankton images is available online: DOI: 10.6084/m9.figshare.8146283.

Description of the procedure

Image convolution is a key process in CNNs in which image features are activated, amplified, and outputted under suitable convolution kernel parameters. The effectiveness of features of the image will directly affect the final output. To overcome the problems of *in situ* plankton images discussed previously, the proposed procedure started from designing an algorithm that can effectively extract potential target objects from images with highly variable contents and different levels of edge sharpness and object contrast ratio (Fig 2). The potential target objects (ROIs), were first located and segmented. Subsequently, local grayscale values were used to enhance the local features of ROIs to allow more effective feature description and reduce gradient loss during the convolution process. Through the image convolution, the features of the enhanced ROIs were extracted after the fully connected layer was achieved in the network. The extracted feature was used as the input for the SVM model classification. In summary, the procedure includes 4 modules: adaptive ROI extraction, ROI enhancement, feature extraction by the CNN, and the multi-class SVM model.

Adaptive ROI extraction

The objective of ROI extraction is to separate the target object from a complex and noisy background and to reduce the interference of the background noise and improve the feature extraction in the subsequent image convolution within the CNNs (Fig 3(*a*)). Low levels of edge

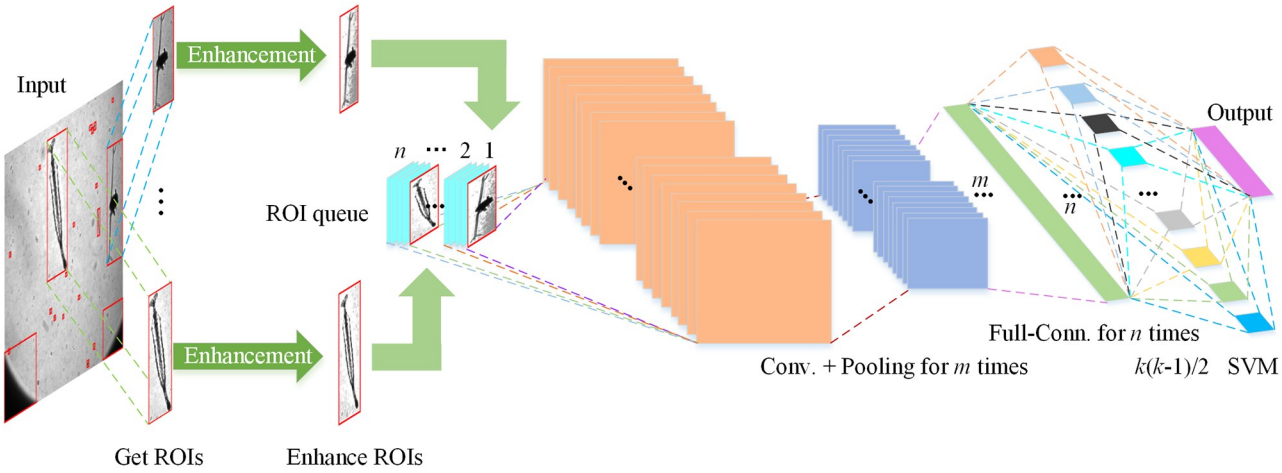

Fig 2. A flow chart illustrating the different steps and modules in the proposed automated plankton identification and enumeration procedure.

https://doi.org/10.1371/journal.pone.0219570.g002

(*a*) (*b*) (*c*)

Fig 3. Examples of binarization with global threshold and local threshold methods. (*a*) Original image with uneven illumination, the contrast of which is adjusted only for illustration purpose due to the heavy darkness of original image, and the procedure directly use the original image; (*b*) Result from binarization with global threshold method; (*c*) Result from binarization with Sauvola's method based on sliding window.

https://doi.org/10.1371/journal.pone.0219570.g003

sharpness and ambiguous morphological features make it difficult to extract ROIs. For example, the commonly used Otsu's global threshold method missed many target objects (Fig 3(*b*)). A combination of Maximally Stable Extremal Regions (MSER) [19] and Sauvola's local binarization method [13] appeared to a good choice [9]. A caveat is that both methods need to manually specify parameters to extract ROIs. MSER requires manual specification of the maximum area variation between extremal regions and step size between intensity threshold levels. Sauvola's local binarization requires a manual specification of suitable window size and a fixed coefficient k in Sauvola's method (Eq 3) to obtain intact target regions. Another problem is that both MSER and Sauvola's method have trouble in segmenting targets with multi-parts and different levels of edge intensity. Uneven illumination and varying contrast ratio within the same image can lead to excessive ROIs for both methods.

To overcome these issues, we described the image contrast ratio using Mean Signal-to-Noise Ratio (*MSNR*), which is defined by Eqs 1 and 2.

$$M = \frac{(x_1 + x_2 + \ldots + x_n)}{n}, \tag{1}$$

$$MSNR = \max(M - x_i)^2, i = 1, 2, \ldots, n, \tag{2}$$

where n represents the number of pixels, M represents the mean value of pixels in the entire image, x_i represents the i-th pixel, and max is the maximum pixel value. For *in situ* plankton images, when the *MSNR* is small, the changes in the contrast ratio and the sharpness of the whole image are low, and it is not easy to distinguish target objects from the background and particulates. When the *MNSR* is large, the whole image is clear, and it is relatively easy to extract target objects.

In the present study, we used a threshold value of 0.1 for *MSNR*. For images with a high contrast ratio, $MSNR > 0.1$, we used the MSER method to segment these images. For images with a low contrast ratio, $MSNR \leq 0.1$, which is often the case for *in situ* plankton images, we used the Sauvola's method, a local threshold segmentation approach to extract ROIs. First, each pixel was considered as a center, and a sliding window was used for pixel-by-pixel sliding on the image with a step of 1 pixel. The length and width of the sliding window was 1~3% of the entire image size. Within every sliding window, we first employed the Sauvola's method to

obtain the local threshold value within the sliding window (Eq 3).

$$T(x, y) = m(x, y)[1 + k(\frac{\delta(x, y)}{R} - 1)], \quad (3)$$

where $T(x, y)$ represents the threshold value for the sliding window at (x, y) calculated from the local contrast ratio, R represents the maximum standard deviation of all pixels in the image that may occur which is 128 for a grayscale image [13], k is a fixed coefficient which usually has a value of 0.34, $m(x, y)$ is the mean pixel value of all pixels in the sliding window and $\delta(x, y)$ is the standard deviation of all pixels in the sliding window. The necessary source codes in MATLAB language for ROI extraction are available here: https://github.com/KaichangCHENG/PIE-MC/tree/master/EnhancedCNN.

The standard Sauvola's method only binarizes the central pixel during each sliding and the sliding step length is 1 pixel and each window has overlapping positions. Based on the analysis, when the changes in pixel values within the region are large, the contrast ratio is large, and the regional standard deviation $\delta(x, y)$ will approach the maximum standard deviation R, i.e., $T(x, y) \approx m(x, y)$. However, in regions with a lower contrast ratio, $T(x, y)$ is significantly lower than $m(x, y)$. The single-pixel sliding Sauvola's method allows an adaptive threshold value, ensuring that the threshold value is between the potential target object and background noise and distinguish them effectively. For example, we applied this technique on the same *in situ* image for Otsu's global thresholding approach and the single-pixel Sauvola's method performed much better in which the ROIs were separated and extracted from the background effectively (Fig 3(*c*)).

ROI enhancement

After extracting the ROIs for the potential target objects, we implemented a procedure to enhance the morphological features of ROIs and suppress background noise.

Target feature enhancement. Due to the complexity of plankton images, conventional spatial domain filtering and frequency domain filtering were not effective. The main inherent problem was too many breakpoints in the target region, making it difficult to extract intact targets. To solve this problem, we employed a denoising algorithm based on breakpoint connections in the spatial domain to achieve target feature enhancement. Based on using the above method for ROI extraction, we took every pixel as a center and use a small rectangular window that was 1~3% of the complete ROI to segment it into many small units. Subsequently, Eq 4 was used to calculate threshold values for every small unit. White pixels were considered as valid pixels and the number of valid pixels within this rectangular region (labeled as N_{valid}) was compared with the threshold value to determine whether this central pixel is a biological boundary feature point:

$$T_{value} = floor(\sqrt{2[floor(\sqrt{N_{rect}})]^2}) - 2, \quad (4)$$

where T_{value} represents the adaptive threshold value, N_{rect} is the number of all pixels in the rectangular window that we set, and the *floor* equation represents a rounding down. The valid pixels in $0.75N_{valid}$ had to lie within one of the rectangular windows or between two adjacent rectangular windows, otherwise that region was regarded as background or noise (Fig 4(*a*)).

The reasons for these settings are twofold. When a valid pixel is the boundary or internal feature point of plankton, there will be many identical feature points nearby. Therefore, the number of valid pixels will be greater than the length of the diagonal line of the rectangular window. From the perspective of pixel density, this can be interpreted as a form of expression for high frequency information. Secondly, if that pixel is a feature point, then its surrounding

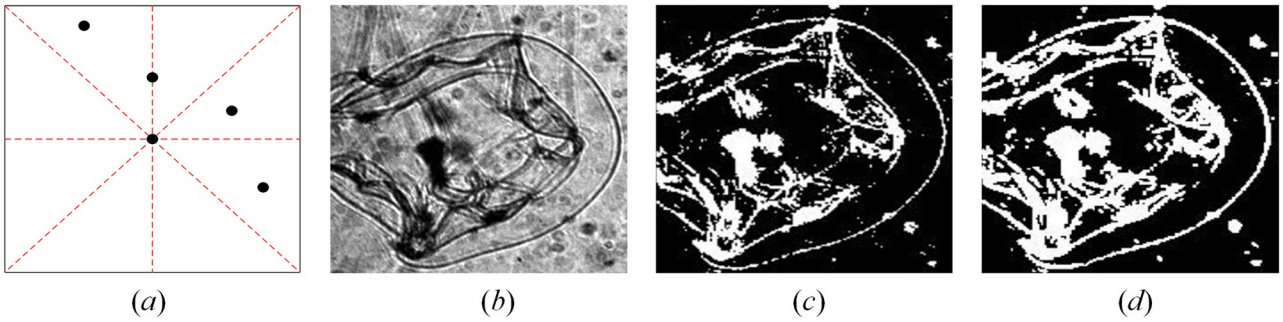

Fig 4. Enhancing target features. (*a*) Illustration of valid pixel. The black point in the center of the black rectangle is the pixel to be confirmed, and the $0.75N_{valid}$ valid pixels (other black points) around it lie within one of the rectangular windows or between two adjacent rectangular windows, so the central pixel is a part of the target; (*b*) Original ROI, the contrast of which is adjusted only for illustration purpose due to the heavy darkness of original ROI, and the procedure directly use the original ROI; (*c*)Example from binarization with Sauvola's method; (*d*) Example after enhancement with the edge roughening method.

https://doi.org/10.1371/journal.pone.0219570.g004

pixels will be distributed around it according to specific rules and will not be scattered randomly within the rectangular window. Therefore, we can carry out filter denoising of the image from the angle of the spatial domain. Fig 4(*c*) and 4(*d*) shows the raw extracted ROI using the sliding window described in the previous section and the results after thickening of the target boundaries through determination of valid pixels, respectively.

Background suppression. After obtaining the intact plankton feature image, the pixel intensity for the segmented ROI was stored in an array $\mathbf{P}_{back}$ and the locations of target object in array $\mathbf{P}_{back}$ were all set to 0. The remaining nonzero pixel intensity in $\mathbf{P}_{back}$, denoted as p, were arranged according to the pixel values, and the boundary threshold T_b was set (Eq 5).

$$T_b = p_{\min} + \frac{p_{\max} - p_{\min}}{\delta}, \delta \in [3, 4], \tag{5}$$

where $p_{\min}$ represents the lowest pixel value in the region, $p_{\max}$ represents the largest pixel intensity value in the region, and δ is the boundary parameter, which is usually a number between 3 and 4.

Grayscale transformation $p' = p + 5 \cdot (p - T_b) \cdot (\delta - 3)$ was carried out based on the difference between the pixel value and the threshold value T_b to get new background intensity p', where p is the nonzero pixel values in $\mathbf{P}_{back}$. Pixel values lower than the threshold value were suppressed and pixel values greater than threshold value were artificially amplified due to the changes in relative intensity. Through this approach, the differences between different pixels were enhanced and morphological features were better reflected. Finally, the new transformed pixel intensity p′ were used to replace p that were recorded in array $\mathbf{P}_{back}$. Note that changes in boundary parameter affect morphological features. For example, when the boundary parameter ranged from 3.1 to 3.9, the morphological features of potential object were enhanced substantially (Fig 5). In the present study, we set the boundary parameter to 3.7 to achieve an optimum result for ROI enhancement (Fig 6). The necessary source codes in MATLAB language for ROI enhancement are also available here: https://github.com/KaichangCHENG/PIE-MC/tree/master/EnhancedCNN.

CNN models for feature learning

CNNs are commonly used in pattern recognition with superior feature learning capabilities. When applied to plankton recognition, it is important to determine the best suitable network structure to overcome issues in plankton recognitions. For example, many plankton are small

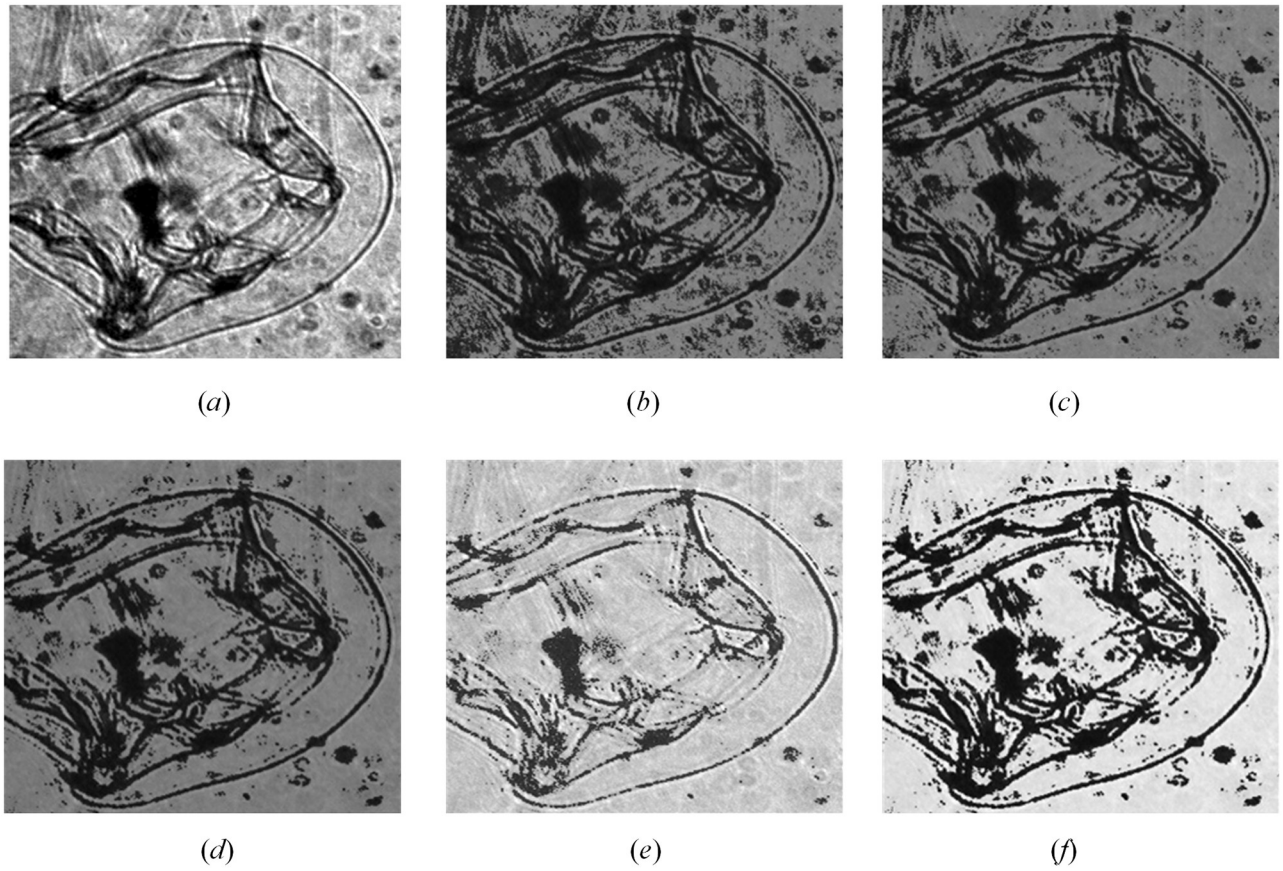

Fig 5. Effects of different δ on ROI enhancement. (*a*) Original ROI, the contrast of which is adjusted only for illustration purpose due to the heavy darkness of original ROI, and the procedure directly use the original ROI; (*b*) $\delta = 3.1$; (*c*) $\delta = 3.3$; (*d*) $\delta = 3.5$; (*e*) $\delta = 3.7$; (*f*) $\delta = 3.9$.

https://doi.org/10.1371/journal.pone.0219570.g005

with mesozooplankton ranging from 200 μm to 2,000 μm and microzooplankton ranging from 20 μm to 200 μm. They can have a wide range of morphological features and sometimes it is difficult to distinguish them from non-living particulates in the water column. In the present study, we tested common network structures for plankton recognition including AlexNet, VGG16, VGG19, GoogLeNet, and ResNet and compared their performance. To examine the impact of ROI enhancement at the front end of the model and the multi-class SVM model at the back of the model, we directly used the convolutional and fully connected layers in CNN models for the feature learning and we used samples in our training set to fine-tune the CNNs. In the back end of the classification model, we used the output of the fully connected layer in these classical CNNs as extracted ROI features, the input for the multi-class SVM model training and classification.

Multi-class SVM classification

The advantage of SVM is that it uses support vectors to identify optimal hyperplanes in a feature space so that the distance between positive and negative samples in the training set is maximized. In addition, flexible intervals can be used to increase fault tolerance and improve the robustness and classification accuracy. Regarding plankton recognition, species within the same class may show large variation, e.g., different copepod species vary in size and morphological features, and therefore a model with high fault tolerance would be beneficial for

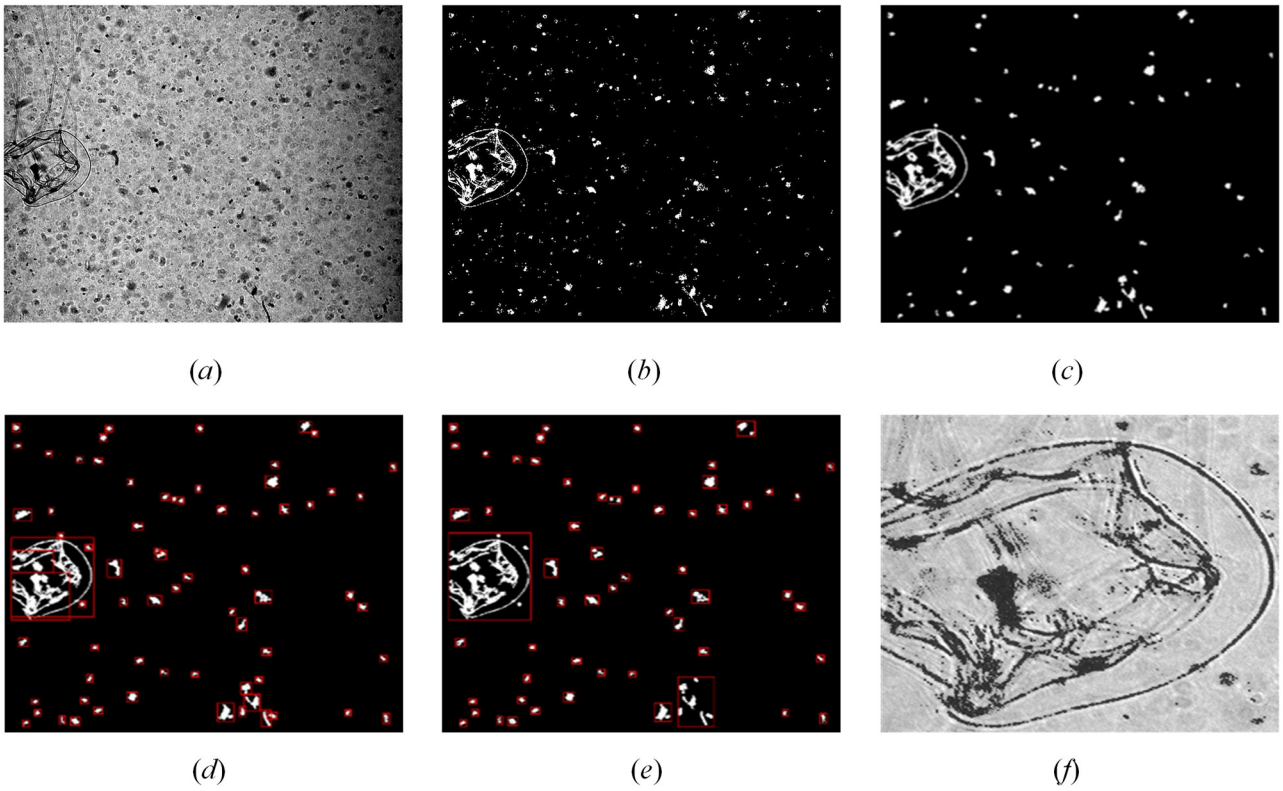

Fig 6. Effects of ROI enhancement with different steps. (*a*) Original plankton image, the contrast of which is adjusted only for illustration purpose due to the heavy darkness of original image, and the procedure directly use the original image; (*b*) The effect of binarization with Sauvola's method; (*c*) The effect of denoising and edge roughening; (*d*) Extraction of ROIs based on connected domain; (*e*) Extraction of ROIs with rectangular merging method based on RPN [45]; (*f*) Final enhanced ROI after background suppression, $\delta = 3.7$.

https://doi.org/10.1371/journal.pone.0219570.g006

classification. In this paper, we employed a multi-class SVM model (Fig 7) to classify target objects using features extracted from selected CNNs.

In the proposed multi-class SVM, a one vs. one classification will train 1 classifier between every two classes. Therefore, there will be $k(k-1)/2$ classifier functions for a problem with k classes. We used a simple linear classifier, $f(\mathbf{X}) = \mathbf{W}^{\mathrm{T}}\mathbf{X} + b$ to map each ROI to different classes, where $\mathbf{W}$ is a weight vector, $\mathbf{X}$ is a feature vector, and b is the bias. When the trained model is used for classification of unknown samples, every classification function will be used to determine its class and the probability of its class. The class of the unknown sample corresponds to the class with the highest probability. As a multi-class SVM model will carry out a 1 vs. 1 comparison between every two classes, it will increase computational demands for identification but effectively increase classification accuracy as compared with a one vs. all SVM model. There were 7 different groups of plankton in the present study groups. Therefore, the computational demand for the 1 vs. 1 multi-class SVM in this study was not increased too much. More importantly, the structural diversity of non-target objects was extremely high, which required a classification model with high accuracy.

Results and discussion

To validate the accuracy and efficiency of the enhanced CNN proposed in this study, we conducted multi-group comparison and validation experiments. The experiments were conducted

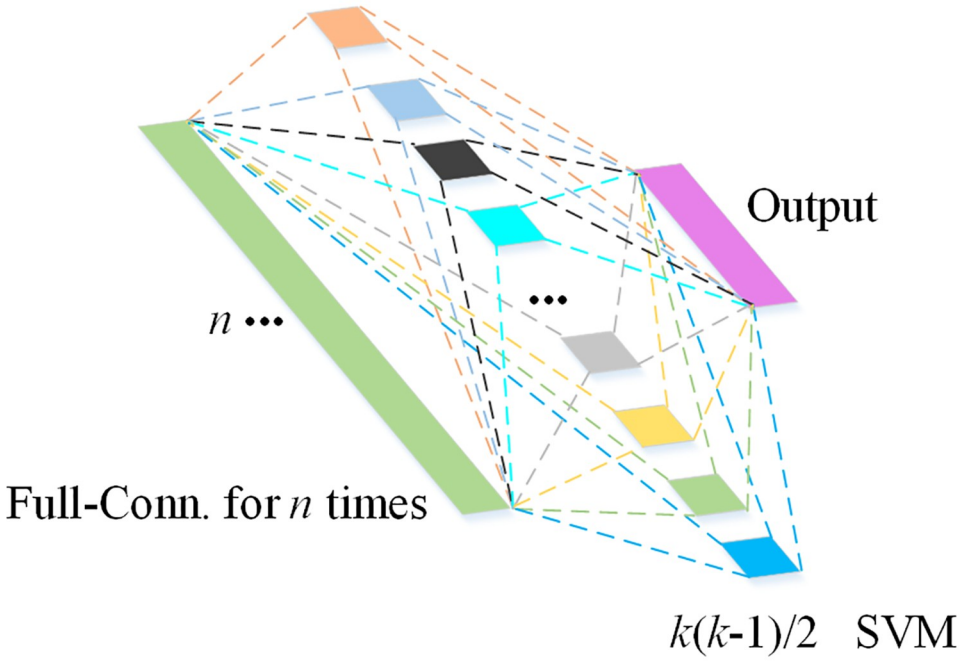

Fig 7. Illustration the multi-class SVM model starting from pairwise classification to the final output.

https://doi.org/10.1371/journal.pone.0219570.g007

using MATLAB (Release 2018a). The accuracy, recall, and elapsed time were the mean values of 5 replicates.

The first experiment (Table 1) was designed to examine whether incorporate the multi-class SVM would improve the performance of the CNN models. The fully connected layers from the selected CNN models were used to describe sample features and were used as the input for the multi-class SVM model. For CNNs with fully connected layers (AlexNet, GoogLeNet, VGG16 and VGG19), the output of corresponding fully connected layers is a one-dimensional feature vector, which was used as the input of multi-class SVM model. While there is no real fully connected layer in ResNet50 model, the output of the last dropout layer in this model is a one-dimensional feature vector, which could be used as the input of multi-class SVM model. The baseline based on Histogram of Oriented Gradient (HOG) features and SVM classifier had relatively low performance on *in situ* plankton ROIs with precision and recall rates of approximately 60%. The selected classic CNN models without the multi-class SVM performed much better on the same set *in situ* plankton training set with both precision and recall rates ranging 85% ~ 88%. Within this group, the ResNet50 model performed the best. Model128_5 and model48_5 proposed by Ouyang py et al. [38] designed for the classification of 2015 National Data Science Bowl with a good imaging quality and did not perform well for *in situ* plankton images, which is only 71.26% and 67.16%, respectively. When the selected CNN models were combined with the multi-class SVM, i.e., some fully connected layers and Softmax classification layer in classic CNN models were replaced by the multi-class SVM model, both the precision and recall rates increased to 88% ~ 92%. We did not use the last fully connected layer of each classic CNN model because it was a special vector that included the corresponding scores of all the classes to be predicted in the model, not the learned features. From Table 1, we can see that the classification performance of experiment using specified features learned from rear fully connected layers were not as good compared with those that directly used the first layer, except for the AlexNet Model. Therefore, we used

Table 1. Results of model performance.

Treatments	Model (Detailed structure)	Precision (%)	Recall (%)	Time (ms/sample)
Baseline	HOG + Multi-class SVM	61.33	60.24	16.82
CNN models without SVM	AlexNet[a]	85.55	85.06	25.95
	GoogLeNet[b] [38]	86.87	87.01	84.10
	VGG16[b] [38]	86.18	86.87	341.59
	VGG19[b] [39]	87.03	87.11	410.76
	ResNet50[b] [39]	88.22	88.46	172.26
	model128_5[b] [38]	71.26	71.87	186.32
	model48_5[b] [38]	67.16	67.89	92.45
CNN models with SVM	AlexNet-fc1 + Multi-class SVM	88.63	88.15	29.05
	AlexNet-fc2 + Multi-class SVM	89.01	88.62	33.09
	GoogLeNet-fc′ + Multi-class SVM	90.45	90.15	87.70
	VGG16-fc1 + Multi-class SVM	91.33	91.13	371.27
	VGG16-fc2 + Multi-class SVM	90.18	90.75	375.62
	VGG19-fc1 + Multi-class SVM	91.86	91.42	425.72
	VGG19-fc2 + Multi-class SVM	90.88	90.99	428.81
	ResNet50-fc′ + Multi-class SVM	92.47	92.76	176.76

Note: Symbol fc1 and fc2 indicate that learned features are from the corresponding fully connected layers according to the order in CNN models. Some models only have one fully connected layer, and outputs of this fully connected layer are the corresponding scores of all the classes to be predicted in the model, not learned features. For GoogLeNet and ResNet50, we use fc′ to indicate that the learned features are from the last dropout layer in the CNN, the output of which is a vector and similar to the output features of fully connected layer in other CNN models. HOG indicates histogram of gradients as feature descriptor and SVM represents support vector machine for classification.

[a]No one used this model to identify plankton images before, and the result reproduced based on *in situ* plankton dataset made by our own group is only for reference.

[b]These models were used for the classification of 2015 National Data Science Bowl with a good imaging quality firstly, and the results in the table were reproduced based on *in situ* plankton dataset made by our own group.

https://doi.org/10.1371/journal.pone.0219570.t001

the output from the fully connected layer of the classical CNN, which was proved to have a better performance in Models 7 to 14, as the input for the multi-class SVM model. The necessary pre-trained models in MATLAB language are also available from https://doi.org/10.6084/m9.figshare.8146283.

Because the original ROIs contained a lot of noise, the second experiment (Table 2) was performed to examine the impact of ROI enhancement on the precision and recall rates using

Table 2. Effect of ROI enhancement in plankton identification and enumeration.

No.	Model (Detailed structure)	Precision (%)	Recall (%)	Time (ms/sample)
1	OriginalROI + AlexNet-fc2 + Multi-class SVM	89.01	88.62	33.09
2	OriginalROI + GoogLeNet-fc′ + Multi-class SVM	90.45	90.15	87.70
3	OriginalROI + VGG16-fc1 + Multi-class SVM	91.33	91.13	371.27
4	OriginalROI + VGG19-fc1 + Multi-class SVM	91.86	91.42	425.72
5	OriginalROI + ResNet50-fc′ + Multi-class SVM	92.47	92.76	175.76
6	EnhancedROI + AlexNet-fc2 + Multi-class SVM	90.44	90.13	34.13
7	EnhancedROI + GoogLeNet-fc′ + Multi-class SVM	92.04	92.15	88.90
8	Enhanced ROI + VGG16-fc1 + Multi-class SVM	93.65	93.43	411.25
9	EnhancedROI + VGG19-fc1 + Multi-class SVM	93.99	93.48	427.57
10	EnhancedROI + ResNet50-fc′ + Multi-class SVM	94.52	94.13	178.42

https://doi.org/10.1371/journal.pone.0219570.t002

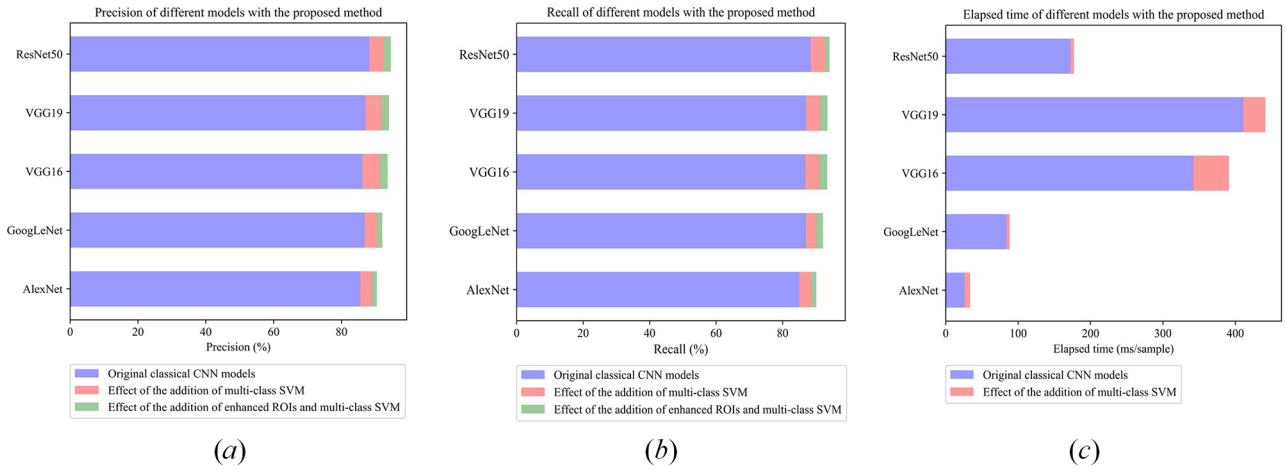

Fig 8. Impacts of multi-class SVM and enhanced ROIs. (*a*) Precision of different models with the proposed method; (*b*) Recall of different models with the proposed method; (*c*) Elapsed time of different models, the increasement of elapsed time (in light coral) is the average of corresponding models based on original ROIs and enhanced ROIs.

https://doi.org/10.1371/journal.pone.0219570.g008

the same selected CNNs, feature output and the SVM classifier. We used different δ values to enhance the original ROI, and then the original samples and all the enhanced samples were combined together to train the classifier. Note that the spatial domain-based breakpoint connection ROI enhancement method was only applied on samples in the training set, while samples in testing set used for performance test were not enhanced. The classification precision, recall, and time consumed generally increased by 1~2% (Fig 8).

When compared to the original classical CNN model, the combination of feature enhancement and multi-class SVM classification layers can increase the classification accuracy and recall of the model by 3~6%, while the computing time increased only by 7~10%. Results suggested that ResNet50 combined with the multi-class SVM performed the best with precision and recall rates >94% and average processing time ~176.764 ms/sample. The misclassification rates among the selected 7 classes also declined using the proposed procedure, i.e., EnhancedROI + ResNet50-fc′ + Multi-class SVM when compared to the results using ResNet50 model alone. Fig 9 showed the corresponding confusion matrix of 7 classes for original finetuned ResNet50 Model and the optimal-performance model proposed in this paper, namely, EnhancedROI + ResNet50-fc' + Multi-class SVM. Clearly, the latter had a much better performance. From the confusion matrix, we can see that limacine is the easiest to identify due to its simple and fixed shape. Copepoda, medusae and euphausiids are relatively harder to identify for their various motion patterns. Chaetognatha, fish larvae and other category are the hardest to identify because of their flexible shapes and motion patterns. Detailly, chaetognatha and fish larvae always have a long and thin body, behaviors of which are always similar and indistinguishable, so the identification results were relatively poor. Other category contains zooplankton other than the six primary categories and non-zooplankton particles which have various behaviors and shapes, so the identification performance is the worst of the 7 classes.

Conclusions and prospects

In summary, we examined the effectiveness of different CNNs models in describing plankton features and results suggested that the ResNet50 performed the best among the 6 selected CNN models. The advantage of ResNet50 in describing plankton likely rises from its relatively wide network structure which allows a better description of plankton, often with relatively

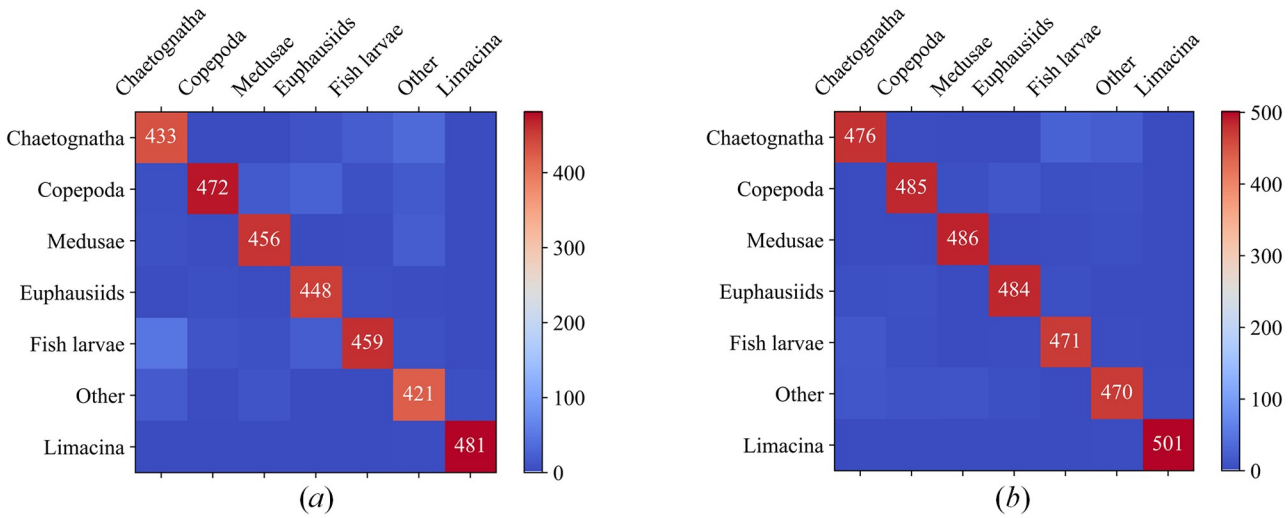

Fig 9. Confusion matrix of the final results of original ResNet50 and the selected optimal-performance model. (*a*) Confusion matrix of original finetuned ResNet50; (*b*) Confusion matrix of EnhancedROI + ResNet50-fc′ + Multi-class SVM.

https://doi.org/10.1371/journal.pone.0219570.g009

small size and < 100 pixels. The inclusion of a multi-class SVM classification model improved the robustness and classification accuracy of the proposed procedure. Finally, a dedicated ROI enhancement helped to remove the background nose and allows more effective feature description which subsequently improved the performance of the proposed procedure with both precision and recall rates >94%. We concluded that the selected ResNet50 model structure combined with the ROI enhancement and the multi-class SVM classification model could effectively identify and enumerate plankton for optical plankton imaging systems like ZOOVIS and other *in situ* plankton imaging systems.

Acknowledgments

This work was supported by The National Key Research and Development Program of China (2017YFC1403602), the Shenzhen Science and Technology Innovation Program (Nos. JCYJ20160428182026575, Nos. JCYJ20170412171011187). Bi was supported by the National Science Foundation (NSF #1602488).

Author Contributions

Conceptualization: Xuemin Cheng, Mark C. Benfield.

Data curation: Kaichang Cheng, Xuemin Cheng.

Formal analysis: Kaichang Cheng, Hongsheng Bi.

Funding acquisition: Xuemin Cheng.

Investigation: Kaichang Cheng, Yuqi Wang.

Methodology: Kaichang Cheng.

Project administration: Xuemin Cheng, Mark C. Benfield.

Resources: Hongsheng Bi.

Software: Kaichang Cheng.

Supervision: Xuemin Cheng, Hongsheng Bi, Mark C. Benfield.

Validation: Kaichang Cheng.

Visualization: Kaichang Cheng.

Writing – original draft: Kaichang Cheng, Yuqi Wang.

Writing – review & editing: Kaichang Cheng, Hongsheng Bi.

References

1. Banse K. Zooplankton: Pivotal role in the control of ocean production. ICES J Mar Sci. 1995; 52(3–4):265–77. https://doi.org/10.1016/1054-3139(95)80043-3
2. Haury LR, McGowan JA, Wiebe PH. Patterns and Processes in the Time-Space Scales of Plankton Distributions. In: Steele JH, editor. Spatial Pattern in Plankton Communities NATO Conference Series (IV Marine Sciences); Boston, MA: Springer; 1978.
3. Wiebe PH, Benfield MC. From the Hensen net toward four-dimensional biological oceanography. Prog Oceanogr. 2003; 56:7–136. https://doi.org/10.1016/S0079-6611(02)00140-4
4. Davis CS, Gallager SM, Marra M, Stewart WK. Rapid visualization of plankton abundance and taxonomic composition using the Video Plankton Recorder. Deep-Sea Res Pt II. 1996; 43:1947–70. https://doi.org/10.1016/S0967-0645(96)00051-3
5. Picheral M, Guidi L, Stemmann L, Karl DM, Iddaoud G, Gorsky G. The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. Limnol Oceanogr METHODS 2010; 8:462–73. https://doi.org/10.4319/lom.2010.8.462
6. Trevorrow MV, Mackas DL, Benfield MC. Comparison of multifrequency acoustic and in situ measurements of zooplankton abundances in Knight Inlet, British Columbia. J Acoust Soc Am. 2005; 117:3574–88. https://doi.org/10.1121/1.1920087 PMID: 16018461
7. Luo T, Kramer K, Goldgof DB, Hall LO, Samson S, Remsen A, et al. Recognizing plankton images from the shadow image particle profiling evaluation recorder. IEEE Trans Syst Man Cybern B Cybern. 2004; 34:1753–62. https://doi.org/10.1109/TSMCB.2004.830340 PMID: 15462442
8. Cowen RK, Guigand CM. In situ ichthyoplankton imaging system (ISIIS): system design and preliminary results. Limnol Oceanogr Methods. 2008; 6:126–32. https://doi.org/10.4319/lom.2008.6.126
9. Bi H, Guo Z, Benfield MC, Fan C, Ford M, Shahrestani S, et al. A Semi-Automated Image Analysis Procedure for In Situ Plankton Imaging Systems. PLoS ONE. 2015; 10(5):e0127121. https://doi.org/10.1371/journal.pone.0127121 PMID: 26010260
10. Benfield MC, Grosjean P, Culverhouse PF, Irigoien X, Sieracki ME, Lopez-Urrutia A, et al. RAPID: Research on Automated Plankton Identification. Oceanogr. 2007; 20:172–87. https://doi.org/10.5670/oceanog.2007.63
11. MacLeod N, Benfield M, Culverhouse PF. Time to automate identification. Nature. 2010; 467:154–5. https://doi.org/10.1038/467154a PMID: 20829777
12. Otsu N. A threshold selection method from gray-level histogram. IEEE Trans Syst Man Cybern. 1979; 9:62–6. https://doi.org/10.1109/TSMC.1979.4310076
13. Sauvola J, Pietikäinen M. Adaptive document image binarization. Pattern Recognit. 2000; 33:225–36. https://doi.org/10.1016/S0031-3203(99)00055-2
14. Crimmins TR. Geometric filter for speckle reduction. Appl Opt. 1985; 24:1438–443. https://doi.org/10.1364/AO.24.001438 PMID: 18223734
15. Galdran A, Pardo D, Picón A, Alvarez-Gila A. Automatic red-channel underwater image restoration. J Vis Commun Image Represent. 2015; 26:132–45. https://doi.org/10.1016/j.jvcir.2014.11.006
16. Zhang S, Wang T, Dong J, Yu H. Underwater image enhancement via extended multi-scale Retinex. Neurocomputing. 2017; 245:1–9. https://doi.org/10.1016/j.neucom.2017.03.029
17. Chen Z, Zhang Z, Bu Y, Dai F, Fan T, Wang H. Underwater object segmentation based on optical features. Sensors. 2018; 18(1):196. https://doi.org/10.3390/s18010196 PMID: 29329245
18. Li X, Zhang M. Underwater color image segmentation method via RGB channel fusion. Opt Eng. 2017; 56(2):023101. https://doi.org/10.1117/1.OE.56.2.023101
19. Matas JCO, Urban M, Pajdla T. Robust wide baseline stereo from maximally stable extremal regions. Image & Vision Computing. 2004; 22(10):761–7.
20. Tang X, Lin F, Samson S, Remsen A. Binary plankton image classification. IEEE J Oceanic Eng. 2006; 31(3):728–35. https://doi.org/10.1109/JOE.2004.836995

21. Li Z, Zhao F, Liu J, Qiao Y. Pairwise nonparametric discriminant analysis for binary plankton image recognition. IEEE J Oceanic Eng. 2014; 39(4):695–701. https://doi.org/10.1109/JOE.2013.2280035

22. Corgnati L, Marini S, Mazzei L, Ottaviani E, Aliani S, Conversi A, et al. Looking inside the ocean: Toward an autonomous imaging system for monitoring gelatinous zooplankton. Sensors. 2016; 16(12):2124. https://doi.org/10.3390/s16122124 PMID: 27983638

23. Zheng H, Wang R, Yu Z, Wang N, Gu Z, Zheng B. Automatic plankton image classification combining multiple view features via multiple kernel learning. BMC Bioinformatics. 2017; 18(16):570. https://doi.org/10.1186/s12859-017-1954-8 PMID: 29297354

24. Tsechpenakis G, Guigand C, Cowen RK. Image Analysis Techniques to Accompany a new In Situ Ichthyoplankton Imaging System. OCEANS 2007; Aberdeen, UK: IEEE; 2007. p. 1–6.

25. Ling H, Jacobs DW. Shape classification using the inner-distance. IEEE Trans Pattern Anal Mach Intell. 2007; 29:286–99. https://doi.org/10.1109/TPAMI.2007.41 PMID: 17170481

26. Culverhouse PF, Simpson RG, Ellis R, Lindley JA, Williams R, Parisini T, et al. Automatic classification of field-collected dinoflagellates by artificial neural network. Mar Ecol Prog Ser. 1996; 139:281–7. https://doi.org/10.3354/meps139281

27. Gorsky G, Ohman MD, Picheral M, Gasparini S, Stemmann L, Romagnan J, et al. Digital zooplankton image analysis using the ZooScan integrated system. J Plankton Res. 2010; 32:285–303. https://doi.org/10.1093/plankt/fbp124

28. Ye L, Chang CY, Hsieh CH. Bayesian model for semi-automated zooplankton classification with predictive confidence and rapid category aggregation. Mar Ecol Prog Ser. 2011; 441:185–96. https://doi.org/10.3354/meps09387

29. Hu Q, Davis CS. Automatic plankton image recognition with co-occurrence matrices and Support Vector Machine. Mar Ecol Prog Ser. 2005; 295:21–31.

30. Hu Q, Davis CS. Accurate automatic quantification of taxa-specific plankton abundance using dual classification with correction. Mar Ecol Prog Ser. 2006; 306:51–61.

31. Krizhevsky A, Sutskever I, Hinton GE, editors. ImageNet classification with deep convolutional neural networks. 25th International Conference on Neural Information Processing Systems; 2012; Lake Tahoe, Nevada.

32. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998; 86(11):2278–324. https://doi.org/10.1109/5.726791

33. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw Learn Syst. 1994; 5:157–66. https://doi.org/10.1109/72.279181 PMID: 18267787

34. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. J Mach Learn Res. 2010; 9:249–56.

35. He K, Sun J. Convolutional neural networks at constrained time cost. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Boston, MA, USA: IEEE; 2015. p. 5353–60.

36. He K, Zhang XY, Ren SQ, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV: IEEE; 2016. p. 770–8.

37. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Honolulu, Hawaii, USA: IEEE; 2017. p. 2261–9.

38. Ouyang P, Hu H, Shi ZZ. Plankton classification with deep convolutional neural networks. 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference; Chongqing, P.R. China: IEEE; 2016. p. 132–6.

39. Li X, Cui ZY. Deep residual networks for plankton classification. OCEANS 2016 MTS/IEEE Monterey; 19–23 Sept. 2016; Monterey, CA, USA: IEEE; 2016. p. 1–4.

40. Luo JY, Irisson J, Graham B, Guigand C, Sarafraz A, Mader C, et al. Automated plankton image analysis using convolutional neural networks. Limnol Oceanogr Methods. 2018; 16:814–27. https://doi.org/10.1002/lom3.10285

41. Liu B, Wang M, Foroosh H, Tappen M, Pensky M. Sparse convolutional neural networks. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Boston, MA, USA: IEEE; 2015. p. 806–14.

42. Bi HS, Cook S, Yu H, Benfield MC, Houde ED. Deployment of an imaging system to investigate fine-scale spatial distribution of early life stages of the ctenophore *Mnemiopsis leidyi* in Chesapeake Bay. J Plankton Res. 2013; 35(2):270–80. https://doi.org/10.1093/plankt/fbs094

43. Simonyan K, Zisserman A, editors. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations (ICLR 2015); 2015 May. 7–9,2015; San Diego, CA, USA.

44. Szegedy C, Liu W, Jia YQ, Sermanet P, Scott ER, Anguelov D, et al. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Boston, MA: IEEE; 2015. p. 1–9.

45. Ren SQ, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell. 2017; 39(6):1137–49. https://doi.org/10.1109/TPAMI.2016.2577031 PMID: 27295650