



**HAL**  
open science

## Automated plankton image analysis using convolutional neural networks

Jessica Y. Luo, Jean-Olivier Irisson, Benjamin Graham, Cédric Guigand,  
Amin Sarafaz, Christopher Mader, Robert K. Cowen

► **To cite this version:**

Jessica Y. Luo, Jean-Olivier Irisson, Benjamin Graham, Cédric Guigand, Amin Sarafaz, et al.. Automated plankton image analysis using convolutional neural networks. *Limnology and Oceanography : methods*, 2018, 16 (12), pp.814-827. 10.1002/lom3.10285 . hal-03390539

**HAL Id: hal-03390539**

**<https://hal.science/hal-03390539>**

Submitted on 21 Oct 2021

**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.

# Automated plankton image analysis using convolutional neural networks

Jessica Y. Luo<sup>1,2,‡\*</sup>, Jean-Olivier Irisson<sup>3</sup>, Benjamin Graham<sup>4</sup>, Cedric Guigand<sup>2</sup>, Amin Sarafraz<sup>5</sup>, Christopher Mader<sup>5</sup>, Robert K. Cowen<sup>1,2</sup>

<sup>1</sup> Marine Biology and Fisheries, Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, 4600 Rickenbacker Causeway, Miami FL 33149, USA

<sup>2</sup> Hatfield Marine Science Center, Oregon State University, 2030 SE Marine Science Dr. Newport OR 97365, USA

<sup>3</sup> Sorbonne Universités, UPMC Univ Paris 06, CNRS, Laboratoire d'Océanographie de Villefranche (LOV), 06230 Villefranche-sur-Mer, France

<sup>4</sup> Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom

<sup>5</sup> Center for Computational Science, University of Miami, 1320 S. Dixie Highway, Coral Gables FL 33146, USA

‡ present address: National Center for Atmospheric Research, Boulder CO

\* corresponding author: [jluo@ucar.edu](mailto:jluo@ucar.edu)

**keywords:** plankton, imaging systems, automated classification, convolutional neural networks

**22 ABSTRACT**

23           The rise of *in situ* plankton imaging systems, particularly high volume imagers  
24 such as the *In Situ* Ichthyoplankton Imaging System (ISIIS), has increased the need for  
25 fast processing and accurate classification tools that can identify a high diversity of  
26 organisms and non-living particles of biological origin. Previous methods for automated  
27 classification have yielded moderate results that either can resolve few groups at high  
28 accuracy, or many groups at relatively low accuracy. However, with the advent of new  
29 deep learning tools such as Convolutional Neural Networks (CNNs), the automated  
30 identification of plankton images can be vastly improved. Here we describe a image  
31 processing procedure that includes pre-processing, segmentation, classification, and post-  
32 processing, for the accurate identification of 108 classes of plankton using Spatially-  
33 Sparse Convolutional Neural Networks. Following a filtering process to remove images  
34 with low classification scores, average precision was 84% and recall was 40%. Biological  
35 groups, excluding 12 rare taxa, exceeded 93% precision. This method provides proof of  
36 concept for the effectiveness of an automated classification scheme using deep learning  
37 methods, which can be applied to a range of plankton or biological imaging systems, with  
38 the eventual application in a variety of ecological monitoring and fisheries management  
39 contexts.

## 40 INTRODUCTION

41 Much of plankton ecology has been focused upon questions surrounding the  
42 identity, quantity, and spatial-temporal variability of planktonic organisms in aquatic  
43 systems, which has historically been addressed by various net-based sampling systems  
44 (Wiebe and Benfield 2003). Although many advanced nets were designed to overcome  
45 limitations in horizontal or vertical resolution in sampling, the emergence of plankton  
46 imaging systems represents a significant advancement for plankton ecology. Current  
47 imaging systems are able to quantify organisms within fine spatial and temporal scales,  
48 with some systems imaging organisms undisturbed and in their natural environment [e.g.,  
49 Video Plankton Recorder (VPR; Davis et al. 1992), Shadow Image Particle Profiling  
50 Evaluation Recorder (SIPPER; Samson et al. 2001), ZOOplankton VISualization and  
51 Imaging System (ZOOVIS; Benfield et al. 2003), *In Situ* Ichthyoplankton Imaging  
52 System (ISIIS; Cowen and Guigand 2008), Underwater Vision Profiler 5 (UVP5;  
53 Picheral et al. 2010). Note that the SIPPER samples via an intake tube, and is thus not a  
54 truly undisturbed sampler]. Research using plankton imaging systems has led to new  
55 insights into the relationships between species and their fine-scale environment (e.g.  
56 Benfield et al. 2000, Ashjian et al. 2001), with implications ranging from fine-scale  
57 aggregation dynamics (Luo et al. 2014), N<sub>2</sub> fixation (Davis and McGillicuddy 2006),  
58 predator-prey interactions (Greer et al. 2013), carbon export (Petrik et al. 2013), and  
59 global plankton biomass estimates (Biard et al. 2016).

60 Though *in situ* plankton imaging systems were developed with a goal of reducing  
61 processing time (very lengthy for physical net samples, which requires sorting and expert  
62 identification), in reality, analyzing plankton images currently still requires extensive and

63 time-consuming manual classification and expert taxonomic knowledge. The tradeoff is:  
64 human operators' time vs. classification accuracy vs. taxonomic resolution. Manual  
65 processing time is typically not reported in papers, but as an example, the manual  
66 analysis of 50+ taxa of gelatinous zooplankton within 5,500 m<sup>3</sup> of water, imaged in  
67 750,000 frames (13.5 inch square frames with 66 μm pixel resolution; each frame with up  
68 to 50 organisms) required the equivalent of three full man-years (Luo et al. 2014).  
69 Classification of pre-processed image segments is slightly faster; Faillettaz et al. (2016)  
70 reported a manual classification rate of 10,000 images day<sup>-1</sup> into 10-15 biotic and abiotic  
71 classes, but a single multiday cruise can easily generate upwards of 50 million image  
72 segments. In general, automated classification efforts currently consist of identifying  
73 small numbers of classes [five to seven classes (Davis et al. 2004; Hu and Davis 2006),  
74 and three classes (Bi et al. 2015)], but even so, few reach an acceptable benchmark of  
75 classification accuracy, commonly set at 67-83% (Culverhouse et al. 2003; Hu and Davis  
76 2005). Alternatively, computer-assisted classification is generally used to achieve higher  
77 accuracies, which consists of a computer generated set of automated classifications  
78 followed by fully validating all images manually (Gorsky et al. 2010; Ohman et al. 2012).  
79 Consequently, it is still very difficult and time-consuming to extract high-accuracy data  
80 on many types of plankton, particularly in highly diverse areas, which limits the utility of  
81 many underwater imaging systems.

82         The issues with classification accuracy and speed have been pronounced with the  
83 *In Situ* Ichthyoplankton Imaging System (ISIIS; Cowen and Guigand 2008), which is a  
84 high resolution, large volume imager designed for sampling mesozooplankton. It  
85 typically images at a rate of 150-185 L s<sup>-1</sup>, depending on tow speed. Compared with other

86 plankton imaging systems (VPR: 10-17 mL s<sup>-1</sup>, ZOOVIS: 3.6 L s<sup>-1</sup>, SIPPER, 9.2 L s<sup>-1</sup>,  
87 UVP5: 8-20 L s<sup>-1</sup>), ISIIS records at 10-1000 times the sampling volume, which has  
88 allowed for studies on rare organisms such as larval fish (Cowen et al. 2013) or large  
89 gelatinous zooplankton (McClatchie et al. 2012; Luo et al. 2014). Particularly in sub-  
90 tropical zones such as the northern Gulf of Mexico, ISIIS can simultaneously record in-  
91 focus, clear images of hundreds of species, ranging from protists, diatom chains, and  
92 copepods, to shrimps, larval fish, and medusae. Therefore, in order to properly classify  
93 organisms within ISIIS images, a classifier that could handle not just a few (< 10) classes,  
94 but many classes (e.g. 30-150) is necessary.

95       Methods for the automated analysis of zooplankton images had early beginnings  
96 in statistical approaches, e.g., discriminant analysis (Jeffries et al. 1980, 1984), but  
97 quickly progressed to using machine learning techniques such as artificial neural  
98 networks (ANNs; Simpson et al. 1992; Culverhouse et al. 1996). Culverhouse et al.  
99 (1996) designed their ANN system explicitly for dinoflagellates, and were able to  
100 identify species with *ca.* 72% accuracy, which was comparable to human classification  
101 (Culverhouse et al. 2003). For a slightly broader range of taxonomic classes (five to  
102 seven classes of phytoplankton and zooplankton), an ANN-type network was combined  
103 with a support vector machine (SVM) in a dual classification method for VPR images,  
104 resulting in classification precision rates between 23-95% when tested on the original  
105 training set (Hu and Davis 2005, 2006). For images from the UVP5, an extensive review  
106 of different classifiers (including ANN and SVM classifiers) resulted in the adoption of a  
107 Random Forest (RF) algorithm, which consistently performed the best, even superseding  
108 the SVM classifier (ZooProcess with PkID; Gorsky et al. 2010; Gasparini 2013).

109 However, even with the success of the RF algorithm, most UVP5 images are still fully  
110 validated by human operators, though there have been some recent efforts towards  
111 decreasing the amount of manual labor required through the use of filtering methods  
112 (Faillettaz et al. 2016). In recent years, SVMs have continued to be used, for systems  
113 such as the SIPPER (active learning with an SVM reduces human labeling efforts; Luo et  
114 al. 2005) and ZOOVIS (SVM classifier using three classes with > 80% precision; Bi et al.  
115 2015), while other groups have continued on with RF methods, sometimes with many  
116 more classes (47 classes, Laney and Sosik 2014; 114 classes, Schmid et al. 2016).  
117 Nonetheless, for all of these classification algorithms, a highly specific set of pre-  
118 measured features was crucial for successfully training the classifier; this set of extracted  
119 features could not dynamically change, nor be automatically determined by the classifier  
120 itself. Furthermore, while the accepted benchmark for plankton classification accuracy  
121 (67-83%, Culverhouse et al. 2003) had been met in many classes by different classifiers,  
122 for biological questions particularly surrounding rare or cryptic species, a high amount of  
123 error is often untenable.

124 Here we present the results of a process to develop an automated classification  
125 algorithm for ISIIS images using Convolutional Neural Networks (CNNs), a relatively  
126 new class of methods that has revolutionized the computer vision field in recent years  
127 (Krizhevsky et al. 2012; LeCun et al. 2015), and which falls within the general category  
128 known as *deep learning*. As opposed to conventional machine-learning techniques such  
129 as artificial neural networks, Random Forest, and support vector machines, deep learning  
130 tools do not require extensive domain expertise (e.g., plankton imaging) and the careful  
131 engineering of feature extractors for classification. Instead, they are able to process

132 natural data in their raw form, and automatically discover the representations that are best  
133 suited for classification. We describe a whole image processing ‘pipeline’, which  
134 includes pre-processing, segmentation, classification, and post-processing. Then, as a  
135 proof of concept, we apply it to a set of ISIS images collected in the northern Gulf of  
136 Mexico. While the described method is highly tuned to images collected by a particular  
137 instrument, CNNs in general (as well as the machine learning competition we ran to  
138 generate this solution) are highly versatile, and can be applied to many types of images  
139 within the biological sciences.

For Review Only

## 140 **METHODS**

### 141 **Description of instrument**

142           The *In Situ* Ichthyoplankton Imaging System (ISIIS; Cowen and Guigand 2008)  
143 utilizes shadowgraph imaging with a line-scan camera to capture silhouette images of  
144 particles in a sampled parcel of water. This backlighting technique, with early application  
145 by Arnold and Nuttal-Smith (1974) and Ortner et al. (1979, 1981), allows for the fine  
146 taxonomic resolution of transparent organisms (e.g. gelatinous zooplankton) and the  
147 coarse taxonomic resolution of small, opaque organisms (e.g. copepods). The camera  
148 used is a 2048-pixel line-scan camera that images over a 13x13-cm field of view and 50-  
149 cm depth of field, with a resultant 66- $\mu\text{m}$  pixel resolution. The output of the imaging is  
150 recorded as a continuous image that is parsed into square frames (2048 x 2048 pixels) at  
151 17 frames  $\text{s}^{-1}$ . While sampling, we target a ship speed of 2.5  $\text{m s}^{-1}$ , which results in an  
152 ISIIS sampling rate of 169  $\text{L s}^{-1}$ . However, in practice, this sampling rate can vary from  
153 150-185  $\text{L s}^{-1}$  with corresponding ship speeds of 2.25-2.75  $\text{m s}^{-1}$ . The recorded data are  
154 ported to the surface via a fiber-optic wire, time-stamped, and saved onto a ship-based  
155 computer or raid array.

156

### 157 **Field sampling**

158           In July-August 2011, ISIIS was deployed over eight, 6-hr transects during two  
159 oceanographic cruises onboard the NOAA ship *McArthur II* in the northern Gulf of  
160 Mexico (Fig 1). The sampling plan was designed to capture images of species present  
161 during the day and night, at various locations and depths, and over multiple months. ISIIS

162 sampled in tow-yo undulations from the surface to 130 m depth at the offshore sites, and  
163 from the surface to 40-60 m depth at the inshore sites.

164

### 165 **Image pre-processing**

166         ISIIS uses a line-scanning camera with a single row of pixels, each with its own  
167 unique light sensitivity characteristics; consequently, raw ISIIS images have a slight non-  
168 uniformity in gray-level across the image, despite the uniform distribution of incoming  
169 light. Furthermore, any dust or particles on the lens appears as vertical lines in the raw,  
170 square frame (Fig. 3a). These lines and image non-uniformities are corrected in a  
171 radiometric calibration called “flat-fielding” in which we calculate a calibration frame  
172 (Fig. 3b) that is subtracted from the raw frame. The calibration is calculated per frame;  
173 since the objects of interest occupied only a small amount of the frame (based on initial  
174 tests, we assumed it to be <20%), we ignored those outliers and calculated a column-  
175 averaged frame for calibration. Thus, the resultant frame after flat-fielding is devoid of  
176 vertical non-uniformities that could bias the segmentation and classification (Fig. 3c).

177         Next, in order to equalize the image histogram, we normalized the contrast within  
178 each frame using the OpenCV 2.4 ‘equalizeHist’ command  
179 (<https://docs.opencv.org/2.4/modules/imgproc/doc/histograms.html>). The histogram  
180 normalized frame allowed for the better detection of regions of interest (ROIs) for  
181 segmentation (Fig. S1).

182         Finally, due to the fact that sampling included coastal waters with high turbidity  
183 (from the Mississippi River plume), we calculated a signal-to-noise (SNR) ratio for each  
184 frame in order to filter out the highly noisy frames captured in turbid waters. The SNR

185 was computed by first calculating a cleaned-up frame, or the “signal-frame”, which was  
 186 simply done by applying a 3x3 median filter to the histogram normalized frame. The  
 187 difference between the histogram normalized frame and the signal frame was then  
 188 considered the “noise-frame”. SNR was then calculated as the log of the ratio between  
 189 vector norm ( $l^2$ -norm) values of the two images:

$$190 \quad SNR = 20 * \log_{10} \left( \frac{|F_{signal}|}{|F_{noise}|} \right) \quad (\text{eq. 1})$$

191 where  $F_{signal}$  is the signal-frame, and  $F_{noise}$  is the noise-frame, and the vector norm values  
 192 was calculated using the OpenCV 2.4 ‘norm’ function, and is defined as:

$$193 \quad |\mathbb{X}| = \sqrt{\sum_{k=1}^n |x_k|^2}, \text{ where } \mathbb{X} = [x_1, x_2, x_3, \dots, x_n] \quad (\text{eq. 2})$$

194 After calculating the SNR on a set of representative images, we found a clear difference  
 195 in SNR values between frames captured from turbid vs. not turbid waters (Fig. S2). Thus,  
 196 we used a threshold cutoff of SNR=25 to discard extremely noisy images, which were  
 197 approximately 26% of all frames originally captured. Note that through earlier efforts, we  
 198 have found that images from highly turbid waters often require manual identification of  
 199 images, so we sought to limit this study to images captured from more typically oceanic  
 200 waters. This exclusion of noisy images should be considered an effect of the sampling  
 201 environment, rather than the image processing method, as images collected in oceanic  
 202 waters rarely had high SNRs.

203

## 204 **Segmentation**

205 Pre-processed frames were then segmented using the ISIIS image segmentation  
 206 software (Tsechpenakis et al. 2007, 2008, Iyer 2012,  
 207 <http://cs.iupui.edu/~gavriil/vital/MVISIIS>). The segmentation software uses an

208 unsupervised machine learning technique, K-harmonic means clustering, to detect ROIs  
209 from the raw images. Iyer (2012) tested six different clustering methods for segmentation  
210 (K-means, iterative K-means, fuzzy C-means, Isodata, Spectral algorithm, and K-  
211 harmonic means), and chose the K-harmonic means method because it achieved the  
212 highest accuracy rates (95%) at relatively fast speeds and was easily implemented for  
213 parallel processing. In our implementation, we found that the segmentation process was  
214 further improved after the addition of the image histogram equalization step (see Image  
215 Pre-processing). Finally, the segmented images were given a unique name that refers to  
216 its time-stamp and location within the original frame. This naming convention allows for  
217 each image to be quickly associated with neighboring images, the original frame,  
218 shipboard GPS, and the environmental data recorded by the instrument.

219

## 220 **Automated image classification**

### 221 *Convolutional Neural Networks*

222 Segmented images were classified using convolutional (or deep) neural networks  
223 (CNNs), a method that is able to process images directly and automatically discover the  
224 characteristics within the images that are best suited for classification. Deep neural  
225 networks make use of the fact that natural images can be analyzed in a hierarchical  
226 fashion, with lower-level features organizing to form higher-level features (e.g. pixels to  
227 edges, edges to body parts, and body parts into organisms), and have been used in  
228 numerous applications from speech and face recognition (Lawrence et al. 1997, Hinton et  
229 al. 2012) to predictions of galaxy morphology (Dieleman et al. 2015). The four key ideas  
230 that characterize CNNs (local connections, shared weights, pooling, and the use of many

231 layers) facilitate minimal pre-processing and require no prior knowledge in designing  
232 features for classification; this represents a significant advance compared with traditional  
233 machine learning methods such as ANNs and SVMs (LeCun et al. 2015).

234       Spatially sparse convolutional neural networks (SparseConvNets) were initially  
235 designed for the recognition of Chinese handwriting. SparseConvNets recognize that the  
236 background of an image often occupies many pixels and not processing them allows the  
237 CNN approach to be applied more efficiently, with less computational cost (Graham  
238 2014). Plankton images can also be considered “sparse images,” as the majority of the  
239 image, even in segmented images, is background. The actual particle or organism  
240 occupies a relatively small percentage of the pixels in the image. Thus, not processing the  
241 background (white) pixels results in a much faster classification process.

242       An application of SparseConvNets with Fractional Max-Pooling (Graham 2015)  
243 was initially developed as part of the Poisson Process team for the 2015 National Data  
244 Science Bowl competition (3-mo machine learning competition to classify *ca.* 60,000  
245 ISIIS plankton images within 121 categories; dataset available at Cowen et al. 2015, see  
246 competition solution at: [www.kaggle.com/c/datasciencebowl/forums/t/13158/poisson-  
247 process-competition-report-and-code/](http://www.kaggle.com/c/datasciencebowl/forums/t/13158/poisson-process-competition-report-and-code/)). For the competition, a number of similar models  
248 were used to generate an ensemble solution. We chose the best single model from team  
249 Poisson Process and made small modifications to improve overall speed with little  
250 apparent changes in accuracy.

251       As opposed to artificial neural networks, which process images as vectors,  
252 convolutional neural networks process images as three-dimensional arrays. Images are  
253 represented in a computer as three-dimensional arrays with size  $N_x N_x C$ , where the first

254 two dimensions ( $N \times N$ ) are spatial dimensions, representing the number of pixels in the  
255 image, and the last dimension ( $C$ ) is the number of color-channels. In our case, the input  
256 image has dimensions of  $N \times N \times 1$ , as there is only one color channel in monochrome  
257 images (RGB color images have  $C=3$ ). However, the  $C$  dimension does not necessarily  
258 have to represent only true colors, but rather can be generalized and expanded to  
259 represent abstract “features” of an image. On a basic level, convolutional networks work  
260 by going through an iterative process of collecting features and appending them as 2-  
261 dimensional slices to the  $C$  dimension. These additional abstract color-channels are  
262 ‘value-added’ images, as they represent increasingly higher-level features, as the  
263 algorithm progresses from the bottom of the network to the top. Examples of features that  
264 would be detected at the bottom of the network include edges or combinations of edges,  
265 and at the top of the network, these features would be something biologically relevant,  
266 such as tails or antennae. These collections of features are constructed by a numerical  
267 optimization technique which involves iteratively showing the network training images  
268 from which it can learn discriminative features useful for classification.

269 Our network is constructed as a sequence of two alternating types of layers,  
270 termed convolutional and pooling layers. Convolutional layers form the main building  
271 block for CNNs, as they detect local combinations of features. Pooling layers operate by  
272 merging semantically similar features into one (for a general description, please see  
273 LeCun et al. 2015). The network has 13 convolutional layers in total, separated by 12  
274 pooling layers. The  $n$ -th convolutional layer looks at overlapping  $2 \times 2$  pixel regions of the  
275 image below, producing an output image with  $32 * n$  color-channels. The number of color-  
276 channels increases as we rise through the network in the expectation that we will produce

277 an increasingly rich description of the contents of the image. The interleaved pooling  
278 layers reduce the spatial size of the input image, but leave the number of color-channels  
279 unchanged. We use fractional max-pooling with scaling factor of  $1/\sqrt{2}$ . The scaling is  
280 multiplicative, so the image shrinks exponentially as we climb the network. This  
281 reduction in resolution offsets the increase in the number of color-channels, ensuring  
282 computational feasibility. After the last convolutional layer, we calculate the average of  
283 each color-channel over the spatial dimensions. We then perform multinomial logistic  
284 regression on the set of color-channel features to predict the class of the image. We used  
285 the SparseConvNet (<https://github.com/btgraham/SparseConvNet>) software package,  
286 which takes advantage of the sparsity of the images to reduce the computational burden,  
287 to train the convolutional neural network.

288

### 289 *Network training*

290 42,564 images were manually sorted in 108 classes to serve as a training set. We  
291 used 100+ classes to accurately represent the taxonomic diversity in the data (Fig 4).  
292 Initially, we started with a training set that was a subset of the data (not shown), which  
293 represented the actual proportions of objects in each class, but refined the training set by  
294 adding in rare classes. Since we were most interested in rarer groups (e.g., larval fish,  
295 jellies, etc.), they were inflated to provide a greater number of representative samples for  
296 the training set. The total number of images in each class of the training set is provided in  
297 Table S1.

298 In addition, at each `epoch` (i.e. training cycle), SparseConvNet picks examples  
299 from the training set and performs data augmentation (randomly rotates, skews, and

300 scales each image), hence creating subtle variations of the original shapes and simulating  
301 new training examples. This procedure is fairly common in CNNs and helps to generalize  
302 a model based on a limited set of examples.

303 We trained for 150 epochs, as this represented the point at which the error rate  
304 plateaued at a minimum value (14.9-15.1%). Using a g2.2xlarge instance on the Amazon  
305 elastic computing cloud (one NVIDIA GPU with 1536 CUDA cores), training 150  
306 epochs took *ca.* 24 hours.

### 307 *Model predictions*

308 To make the prediction robust, each of the 23.4M images in the full data set was  
309 passed through the fitted network 24 times, each time with different data augmentation  
310 parameters. The probabilities for each object to belong to each class, predicted by the  
311 model, were averaged over the 24 predictions, the maximum was found, and the  
312 corresponding class was considered as the predicted class. Total prediction time was 165  
313 machine hours (though < 36 actual hours, as we used five GPU instances in parallel).

### 314 *Classification groupings*

315  
316 The 108 original classes in the training set were mapped onto 37 broader groups,  
317 which represented taxonomic or functional groupings that were more relevant for  
318 ecological analyses (Fig. 4, Table S2). For example, many of the original classes were  
319 created for automated image classification purposes, with the distinctions between classes  
320 only morphological (e.g., straight vs. curved appendicularians) or due to an imaging or  
321 segmentation artifact (e.g., cropped bells and tentacles). Others were created to  
322

323 distinguish between different forms within a diverse class (e.g., detritus) that would  
324 otherwise pollute many other classes. Lastly, some taxonomic classes were grouped  
325 together for filtering purposes and further analyses (e.g., the fish groups).

326

### 327 *Model performance*

328 Confusion matrices (CMs) are a tool for quantifying classifier accuracy (e.g., Hu  
329 and Davis 2005, Bi et al. 2015). The calculated CM statistics included three values and  
330 three rates, calculated separately for each class  $i$ : values were numbers of true positives  
331 ( $TP_i$ ), false positives ( $FP_i$ , type I error), and false negatives ( $FN_i$ , type II error). The rates  
332 calculated were precision ( $P_i$ , eq. 3), recall ( $R_i$ , eq. 4), and the F1-score, which is the  
333 harmonic mean of the precision and recall rates ( $F1_i$ , eq. 5). For a given class, precision  
334 quantifies the “purity” of the prediction and recall quantifies the “completeness” of the  
335 prediction.

$$336 \quad P_i = TP_i / (TP_i + FP_i) \quad (\text{eq. 3})$$

$$337 \quad R_i = TP_i / (TP_i + FN_i) \quad (\text{eq. 4})$$

$$338 \quad F1_i = 2 \times P_i \times R_i / (P_i + R_i) \quad (\text{eq. 5})$$

339 A self-prediction of the training set and the associated CM represent the  
340 theoretical maximum of the classifier performance. It was computed to give a benchmark  
341 for determining which classes had naturally high variability and which ones were  
342 relatively homogenous (Table S1).

343 To evaluate the classification success on the full dataset, we performed spot-  
344 checks: 75,000 predicted images were picked randomly (0.30% of the total dataset) and

345 their identification was manually validated. The corresponding CM is shown in Table S2  
346 (see column “Without Probability Filtering”).

347

#### 348 *Probability filtering*

349 In the full dataset, images of organisms spanned the range in terms of quality:  
350 small (e.g., early life stage) to large (e.g., adult), blurry to clear, oriented towards, away,  
351 or to the side of the camera, etc. Thus, the images that were more difficult to predict, or  
352 less archetypal often were associated with a low prediction score. The prediction score is  
353 an output of any classification algorithm: for each candidate image, the algorithm  
354 computes a score (often a probability) associated with every category in the training set.  
355 Classification is then just a matter of picking the maximum score. However, for difficult  
356 to identify objects that could fit in many classes, even the maximum score can be low,  
357 reflecting a low confidence in the classification. Therefore, we used this score to filter  
358 classified images into “high” vs. “low” likelihood of correct classification using a  
359 threshold value set for each class. Faillettaz et al. (2016) first demonstrated this approach,  
360 showing that the removal of “low-confidence images” (in their case, over 70% of their  
361 original dataset) still allowed for the prediction of true spatial distributions of many taxa.

362 In the present study, we determined the appropriate threshold values for each  
363 class by predicting a new, independent, 43,000-member test set. All images in this test set  
364 were manually identified, which allowed us to detect prediction errors. For each class, we  
365 set the threshold value to be the classification score *above which* 95% of images were  
366 correctly classified into the corresponding *group* (as opposed to the *class* itself). Groups  
367 were used because many of the 108 classes were separated based on morphological

368 distinctions with little ecological relevance (e.g., “chaetognaths *curved*” vs “chaetognaths  
369 *straight*”). The 95% level, which resulted in 29.6% of images discarded (though  
370 individual classes varied, Table S3), was chosen as a compromise between improving  
371 classification accuracy and retaining enough images for ecological analyses. As a  
372 comparison, if the thresholds were set at the 90% level, then only 19% of images would  
373 be cut, but at 99% level, then 63% of all images would be discarded.

374         The discarded, “low confidence images”, were put into the “unknown” category.  
375 Since this affected some of the 75,000 randomly selected images used to compute the  
376 confusion matrix, a post-filtering confusion matrix was then recalculated. The differences  
377 between the CM stats before vs. after probability filtering are shown in Table S2, and the  
378 section “With Probability Filtering” gives the CM for the final processed dataset, which  
379 can be used for future ecological studies.

380

## 381 **RESULTS**

382 A total of 2.4 million raw ISIIS frames (nearly 40 hours of imaging, 10 TB data)  
383 were collected from eight transects in the northern Gulf of Mexico. The raw ISIIS frames  
384 were segmented into 23.4 million images (27 GB), and classified into all 108 categories  
385 (Table S2). After filtering, 64.3% of the images were retained, and 35.7% discarded.

386 Aside from the detritus and artifact images, there were 1.62 million images of  
387 phytoplankton and protists and 1.37 million images of mesozooplankton.

### 389 **Training set prediction: accuracy benchmarks**

390 Overall, the F1-score (the harmonic mean between precision and recall) for all  
391 classes was 88.1, with 67% with a F1-score over 90, and 83% with a F1 score over 80.  
392 The hardest classes to predict were the fish classes, with a mean F1 score of 70.7 (e.g.,  
393 myctophid fishes were often confused for other types of fishes), and the easiest classes to  
394 predict included the protists (mean F1 of 98.0), cyclopoid copepods (mean F1 of 96.1),  
395 and chaetognath classes (mean F1 of 93.4) (Table S1). At the group level, the F1-scores  
396 increase, such that the lowest was 85.6 (physonect siphonophores), and 70% had F1-  
397 scores of 95 or above (Table 1).

### 399 **Image filtering and confusion matrices**

400 In total, filtering removed 30% of all images, though this percentage differed by  
401 category. Out of 108 categories, 26 were well-predicted (over 60% retained after  
402 filtering), including the diatom chains, chaetognaths, dark detritus, protists, doliolids, and  
403 three calanoid copepod classes. Many of these 26 classes also fell within the top quarter

404 in terms of numerical abundance (10 classes, containing 18.1 out of 23.4 million images).  
405 In particular, the main artifact class (imaging artifacts, as opposed to detritus), which  
406 comprised over 8.2 million images, or 35% of the total, were well-predicted, and over  
407 93% were retained. However, the well-predicted classes were not only the common  
408 classes, since some of the rare but morphologically monotypic (e.g. the cestid  
409 ctenophores and goby fishes) also performed very well. In contrast, 22 classes were very  
410 heavily filtered, where less than 10% were retained. These classes included six (out of  
411 14) hydromedusae, five (out of 19) fish, two polychaete worms, two siphonophores, one  
412 copepod and one shrimp, and tended to be the less common but morphologically diverse  
413 classes (Table S3).

414 Results from the 75,000 random spot-checks showed that filtering improved the  
415 mean classification precision rate at the group level by 33 percentage points, from 53% to  
416 86% precision (Table S2). If only the biological groups were considered (thus excluding  
417 artifacts, detritus, and unknown, which was nearly 80% of the dataset), this increase was  
418 just slightly greater, from 51% to 87% precision. Twelve of the biological groups had less  
419 than 25 randomly drawn images (Table S2, also marked in grey in Table 1); these groups  
420 were very rare, each representing less than 0.12% of the total biological data. Excluding  
421 the rare biological groups, the precision rate after filtering was 89.7%.

422 Naturally, using the filtering thresholds decreased the total recall rate, by 19  
423 percentage points, from 63% to 44%. For the non-rare biological groups (n=23), the  
424 decrease was greater (22 points), but final recall rate was higher (47%). However, the F1-  
425 score, which is the harmonic mean of the precision rate and recall rate, only increased  
426 slightly, from 54% to 55%.

427           The final classification comparison was conducted between the (post-filtering)  
428 classifier and the training set, which represents the difference between a full dataset  
429 classification and the theoretical maximum for a classifier (Table 1). On the full dataset,  
430 classification precision was actually close to, or even exceeded that of the training set,  
431 which was possible because of the application of the filtering thresholds. Despite the  
432 corresponding decrease in the recall rate, a comparison of F1-scores showed that a few of  
433 the biological groups had a less than 10-point difference (*Oithona* copepods, sergestid  
434 shrimp, and ablyid siphonophores). Groups that were less common, or had a lot of natural  
435 variability, such as other shrimp, pteropods, and cydippid ctenophores, showed a much  
436 greater difference, of 70-80 points, which was largely due to low recall rates post-  
437 filtering. However, the average difference in F1-scores for the (non-rare) biological  
438 groups was 40%, representing a moderate difference between the final classifier and the  
439 training set.

440

441 **DISCUSSION**

442           We demonstrate the successful application of an image processing procedure,  
443 using a deep learning convolutional neural network, to classify a ~40 hour, 10 TB *in situ*  
444 plankton imaging dataset containing 25 million image segments into 108 classes. After  
445 applying a filtering threshold on the classification probabilities, and grouping the classes  
446 into 37 taxonomically and functionally meaningful groups, the average classifier  
447 precision on non-rare biological groups (n=23) was 93%, which is higher than any  
448 previous attempt on high sampling volume, *in situ* plankton images.

449           Since Culverhouse et al. (2003) published a finding that trained personnel are  
450 only able to achieve 67-83% self-consistency on an expert plankton classification task,  
451 that range has existed as a sort of *de facto* benchmark within the plankton imaging field  
452 in which computer classification can be considered to be as good as human classification  
453 (e.g. Hu and Davis 2005, 2006). In reality, Culverhouse et al. (2003)'s findings were  
454 specific to a *difficult* identification task, in which morphologically variable dinoflagellate  
455 species (genus *Dinophysis*) were being distinguished from each other. For *in situ*  
456 plankton images, it is not very difficult for a human to distinguish between broad  
457 plankton community-based groups (e.g. calanoid copepods, shrimps, and larval fish), but  
458 rather, the difficulty only lies when distinguishing within certain taxa (e.g. between larval  
459 mesopelagic fishes, or between small decapod shrimps). Of course, classification  
460 difficulty may vary due to environmental conditions and ecosystem composition.  
461 Nonetheless, we suggest that this benchmark should be revisited, and raised to at least  
462 90%. In manual sorting for the present dataset as well as others with 120+ classes (e.g.  
463 Cowen et al. 2015), the proportion of unknowns, in which an expert operator is unable to

464 sort the image, ranged from  $< 1\%$  for 30-35 classes to *ca.* 5% for 120-130 classes. In the  
465 present case, application of a method incorporating deep machine learning and filter  
466 thresholding resulted in nearly 90% precision on all the non-rare biological groups; this  
467 approaches the point in which we may consider an automated classifier to be as good as a  
468 human operator in sorting common plankton groups at higher taxonomic levels.

469       Application of the Faillettaz et al. (2016) filtering method gives us the ability to  
470 select for the highest probability images, and subsequently manipulate the precision  
471 levels (and by association, the recall rates) in the final classifier. This step allowed us to  
472 ensure the best description of biological patterns, which was important given the  
473 scientific goals of the image analysis procedure. Application of the filtering method  
474 would likely significantly increase the overall performance of other previously published  
475 classification schema, and would likely temper the difference between our results and  
476 those earlier studies. Secondly, we also note that it is not sufficient to judge a classifier  
477 by the class precision alone; the recall rate must also be incorporated. We therefore  
478 propose a more widespread adoption of the F1-score, which is the harmonic mean of the  
479 precision and recall.

480       Convolutional neural networks represent a significant advance over traditional  
481 machine learning methods, because they are designed to learn and automatically extract  
482 feature descriptors (LeCun et al. 2015). Aside from the construction of the neural  
483 network architecture, the single most important factor determining the success of the  
484 classifier was the training set. Fernandes et al. (2009) had proposed a computer-assisted  
485 method for determining the optimal number of classes (settled on 30), using a Tree-  
486 Augmented Naïve Bayes classifier. In our case, since deep learning methods are capable

487 of classifying many more classes, we manually defined 108 classes and then grouped  
488 them into 37 groups after classification, but future efforts with CNNs should utilize some  
489 amount of computer assistance in determining the identity and quantity of classes.

490         Deep learning methods require large amounts of training data, and our 42,000  
491 item training set for 108 categories was likely on the low end; significant amounts of data  
492 augmentation was necessary. However, this is still an order of magnitude greater than the  
493 training sets used by traditional machine learning plankton image classifiers: Hu and  
494 Davis (2006) used 200 images per class for seven classes, Bi et al. (2005) used 210  
495 images total for three classes, and Faillettaz et al. (2016) used 5,979 images for 14  
496 classes. Our choice of using a "natural" training set, where rare classes were augmented  
497 but not to the quantity of the most common classes, was a decision following our broader  
498 research objectives of describing mesozooplankton (including larval fish) distributions.  
499 These organisms are relatively rare, especially compared to protists and diatom chains,  
500 and thus needed special attention within the training set. Augmenting rare groups in the  
501 training set is naturally a time-consuming process. However, if the scientific objective of  
502 the image analysis system were to classify the detritus and common phytoplankton, then  
503 a more representative training set would achieve higher accuracies (Chang et al. 2012).  
504 Furthermore, to the extent possible, it was necessary to include the range of images, from  
505 the best (clearest, sharpest) image to the worst (most ambiguous, blurry) image, and to  
506 divide classes not only by taxonomy, but also morphological differences. Still, there were  
507 classes that did not perform very well (e.g., "shrimp other"), but were too difficult to  
508 separate further.

509           The development of our classifier (an application of the Spatially-Sparse  
510 Convolutional Neural Network, Graham 2014, 2015) was achieved following the 2015  
511 National Data Science Bowl, a [Kaggle.com](https://www.kaggle.com) machine learning competition. For the  
512 competition, we used the same ISIIS imaging system as in the present paper, but data  
513 from a different sampling region (Straits of Florida; competition data available at Cowen  
514 et al. 2015). While crowd-sourcing and machine learning competitions are not within the  
515 scope of the present paper (but discussed in Robinson et al. 2017), there were some key  
516 lessons we learned through the process that determined the successful application of the  
517 present classifier. First, in many competition settings, teams submit results that are an  
518 average of multiple models, also known as ensembles, which are computationally  
519 expensive and not necessarily the most realistic for real-world use. Therefore, it was  
520 critical to identify the single best model, which may or may not be part of the best  
521 ensemble (it was not in our case). Second, further development of the classification  
522 scheme was necessary after the competition ended. Essential to our success was the  
523 inclusion of a bio-computing specialist who could bridge the gap between the biologists  
524 and the computer scientists. Finally, the design of the competition dataset was also highly  
525 important, as it determined the types of solutions that emerged. We found that it was  
526 essential that the dataset had all the qualities of a good training set (ratio of images within  
527 rare vs. common classes, inclusion of high and low quality images, and separation of  
528 classes by taxonomy and morphology). These three key points facilitated the successful  
529 transfer of an image classifier between the competition and the present context.

530           As plankton datasets, both physical (e.g. the Continuous Plankton Recorder  
531 archive) and digital (the growing ISIIS collection), get larger and more comprehensive, it

532 is critical to note that the amount of samples to sort at a particular taxonomic resolution  
533 will always depend on the scientific question *and* the time available for analyses. For  
534 some questions, such as the spatio-temporal variability in ichthyoplankton distributions  
535 (Richardson et al. 2010) or the niche shift of sibling species (Beaugrand et al. 2002),  
536 manually sorting physical samples to the genus or species level is necessary, but in those  
537 cases, only a relatively small number of organisms can realistically be sorted. For other  
538 questions, such as the fine-scale distribution of broad taxonomic groups, the complete  
539 analyses of samples collected by high throughput imaging systems is most adapted. In  
540 that case, manual sorting would be time prohibitive, especially with increasing numbers  
541 of classes (we estimate that sorting into 40-50 classes, which can be done at 5,000 images  
542  $d^{-1}$ , occurs at roughly half the rate of sorting into 10-15 classes). Therefore, computer-  
543 assisted or fully automated classification becomes more expedient. Even the necessity of  
544 creating a training set, with associated independent test set, for each sampling region is  
545 time consuming (by our estimates, *ca.* 2-3 mo.). The future development of a master,  
546 global-level training set with regional filters could facilitate a more rapid image  
547 classification process. This could eventually lead to a minimal amount of manual  
548 identification work for each additional dataset (i.e. for spot-checks for the final confusion  
549 matrix, which would yield a class-specific correction factor for densities). The  
550 combination of the speed of classification, use of 100+ classes, high precision, and only  
551 needing to do small amounts of manual sorting would significantly increase the utility of  
552 plankton imaging systems, as we will be able to classify millions to billions of *in situ*  
553 plankton images quickly and accurately.

554           It is evident that with the recent interest in plankton (e.g. from the *Tara* Oceans  
555 project, Bork et al. 2015) that there are many additional questions and areas for  
556 exploration regarding the base of the marine food chain. Imaging systems are inherently  
557 complementary to net-based sampling; physical samples are always going to be necessary  
558 for ecological questions requiring fine taxonomic resolution and the analysis of hard  
559 structures (e.g. otoliths), isotopes, or genetics. However, large-volume imaging systems  
560 can be particularly useful for addressing questions regarding rare, gelatinous, or large  
561 organisms in the context of predator-prey dynamics, horizontal and vertical aggregations,  
562 and fine-scale relationships to the environment. Plankton imaging systems can also  
563 provide important validation data for regional and global ocean ecosystem models, which  
564 suffer from insufficient data for constraining patterns and processes. The development of  
565 whole, integrated pipelines for plankton image analysis can enhance the utility of  
566 automated classification tools, and can eventually lead to the goal of real-time image  
567 processing done at sea. Combined with some net-sampling for taxonomic validation,  
568 plankton imaging systems can be an incredibly powerful tool, with applications in ocean  
569 monitoring and fisheries management, as well as in addressing many of the fundamental  
570 questions still existing within plankton ecology.

**571 ACKNOWLEDGEMENTS**

572 We thank Booz Allen Hamilton for sponsoring and for Kaggle.com for hosting the 2015  
573 National Data Science Bowl competition. Many thanks to Robin Faillettaz, Christian  
574 Briseño-Avena, and Kelly Robinson for helpful discussions, and to Chase Cobb and  
575 Clare Hansen for assistance in image analysis. Funding for this project came from NOAA  
576 NRDA Contract AB133C-11-CQ-0050 to RKC, and a National Science Foundation  
577 Graduate Research Fellowship and University of Miami Maytag Fellowship to JYL.  
578 RKC, JYL, and CM were also supported on NSF OCE Grant 1419987 (to RKC and Su  
579 Sponaugle) during the preparation of this manuscript.

580

**581 DATA AVAILABILITY**

582 All manually classified images from the full training set and test sets (43K probability  
583 filtering set and 75K random spot-check set), as well as text files containing predicted  
584 and validated classes for all test sets will be available on Zenodo.org (temporary version:  
585 <https://sandbox.zenodo.org/record/238694#.W3GreC2ZNTZ>, with permanent version at  
586 doi: 10.5281/zenodo.836492 upon publication). Executable files for the segmentation  
587 code will also be on Zenodo.org. Source code for SparseConvNet  
588 is available at <https://github.com/btgraham/SparseConvNet>, as well as on Zenodo.org.  
589

590 **REFERENCES**

- 591 Arnold, G. P., and P. B. N. Nuttall-Smith. 1974. Shadow cinematography of fish larvae.  
592 Mar. Biol. 28:51-53.
- 593 Ashjian, C. J., C. S. Davis, S. M. Gallagher, and P. Alatalo. 2001. Distribution of plankton,  
594 particles, and hydrographic features across Georges Bank described using the  
595 Video Plankton Recorder. Deep Sea Res. II 48:245-282
- 596 Beaugrand, G., P. C. Reid, F. Ibañez, J. A. Lindley, and M. Edwards. 2002.  
597 Reorganization of North Atlantic marine copepod biodiversity and climate.  
598 Science 296:1692-1694
- 599 Benfield, M. C., C. S. Davis, and S. M. Gallagher. 2000. Estimating the in-situ orientation  
600 of *Calanus finmarchicus* on Georges Bank using the Video Plankton Recorder.  
601 Plankton Biol. Ecol. 47(1):69-72
- 602 Benfield, M. C., C. J. Schwehm, R. G. Fredericks, G. Squyres, S. F. Keenan, and M.V.  
603 Trevorrow. 2003. Measurements of zooplankton distributions with a high-  
604 resolution digital camera system. In: L. Seuront, P.G. Strutton (Eds.) Handbook of  
605 Scaling Methods in Aquatic Ecology: Measurement, Analysis, Simulation. CRC  
606 Press, p. 17–30
- 607 Bi, H., Z. Guo, M. Benfield, C. Fan, M. Ford, S. Shahrestani, and J. Sieracki. 2015. A  
608 semi-automated image analysis procedure for in situ plankton imaging systems.  
609 PloS ONE 10:e0127121
- 610 Biard, T., L. Stemann, M. Picheral, N. Mayot, P. Vandromme, H. Hauss, G. Gorsky, L.  
611 Guidi, R. Kiko, and F. Not. 2016. *In situ* imaging reveals the biomass of giant  
612 protists in the global ocean. Nature 532:504-507

- 613 Bork, P., C. Bowler, C. de Vargas, G. Gorsky, E. Karsenti, and P. Wincker. 2015. *Tara*  
614 Oceans studies plankton at planetary scale. *Science* 348:873-873
- 615 Chang, C.-Y., P.-C. Ho, A. R. Sastri, Y.-C. Lee, G.-C. Gong, and C.-H. Hsieh. 2012.  
616 Methods of training set construction: Towards improving performance for  
617 automated mesozooplankton image classification systems. *Cont. Shelf Res.*  
618 36:19-28
- 619 Cowen, R. K., and C. M. Guigand 2008. *In Situ* Ichthyoplankton Imaging System (ISIS):  
620 System design and preliminary results. *Limnol. Oceanogr. Meth.* 6:126-132
- 621 Cowen, R. K., A. T. Greer, C. M. Guigand, J. A. Hare, D. E. Richardson, and H. J.  
622 Walsh. 2013. Evaluation of the *In Situ* Ichthyoplankton Imaging System (ISIS):  
623 Comparison with the traditional (bongo net) sampler. *Fish. Bull.* 111:1-12
- 624 Cowen, R. K., S. Sponaugle, K. L. Robinson, and J. Luo. 2015. PlanktonSet 1.0:  
625 Plankton imagery data collected from F.G. Walton Smith in Straits of Florida  
626 from 2014-06-03 to 2014-06-06 and used in the 2015 National Data Science Bowl  
627 (NODC Accession 0127422). NOAA National Centers for Environmental  
628 Information. Dataset. [doi:10.7289/V5D21VJD](https://doi.org/10.7289/V5D21VJD). [https://data.nodc.noaa.gov/cgi-](https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:0127422)  
629 [bin/iso?id=gov.noaa.nodc:0127422](https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:0127422)
- 630 Culverhouse, P. F., R. Williams, B. Reguera, R. E. Ellis, and T. Parisini. 1996. Automatic  
631 categorization of 23 species of dinoflagellate by artificial neural network. *Mar.*  
632 *Ecol. Prog. Ser.* 139:281-287
- 633 Culverhouse, P. F., R. Williams, B. Reguera, V. Herry, and S. González-Gil. 2003. Do  
634 experts make mistakes? A comparison of human and machine identification of  
635 dinoflagellates. *Mar. Ecol. Prog. Ser.* 312:297-309

- 636 C. S. Davis, S. M. Gallagher, M. S. Berman, L. R. Haury, and J. R. Strickler. 1992. The  
637 Video Plankton Recorder (VPR): Design and initial results. *Archiv für*  
638 *Hydrobiologie Beiheft, Ergebnisse der Limnologie*. 36:67–81
- 639 C. S. Davis, Q. Hu, S. M. Gallagher, X. Tang, and C. J. Ashjian. 2004. Real-time  
640 observation of taxa-specific plankton distributions: An optical sampling method.  
641 *Mar. Ecol. Prog. Ser.* 284:77-96
- 642 C. S. Davis, and D. J. McGillicuddy. 2006. Transatlantic abundance of the N<sub>2</sub>-fixing  
643 colonial cyanobacterium *Trichodesmium*. *Science* 312:1517-1520
- 644 Dieleman, S., K. W. Willett, and J. Dambre. 2015. Rotation-invariant convolutional  
645 neural networks for galaxy morphology prediction. *Mon. Not. R. Astron. Soc.*  
646 450:1441-1459
- 647 Faillettaz, R., M. Picheral, J. Y. Luo, C. Guigand, R. K. Cowen, and J.-O. Irisson. 2016.  
648 Imperfect automatic image classification successfully describes plankton  
649 distribution patterns. *Meth. Oceanogr.* 15-16:60-77
- 650 Fernandes, J. A., X. Irigoien, G. Boyra, J. A. Lozano, and I. Inza. 2009. Optimizing the  
651 number of classes in automated zooplankton classification. *J. Plankton. Res.*  
652 31:19-29
- 653 Gasparini, S., and E. Antajan. 2013. PLANKTON IDENTIFIER: a software for  
654 automatic recognition of planktonic organisms. [http://www.obs-](http://www.obs-vlfr.fr/~gaspari/Plankton_Identifier/index.php)  
655 [vlfr.fr/~gaspari/Plankton\\_Identifier/index.php](http://www.obs-vlfr.fr/~gaspari/Plankton_Identifier/index.php)
- 656 Gorsky, G., M. D. Ohman, M. Picheral, S. Gasparini, L. Stemann, J.-B. Romagnan, A.  
657 Cawood, S. Pesant, C. Garcia-Comas, and F. Prejger. 2010. Digital zooplankton  
658 image analysis using the ZooScan integrated system. *J. Plankton. Res.* 32:285-303

- 659 Graham, B. 2014. Spatially-sparse convolutional neural networks. arXiv:14096070
- 660 Graham, B. 2015. Fractional max-pooling. arXiv:1412.6071
- 661 Greer, A. T., R. K. Cowen, C. M. Guigand, M. A. McManus, J. C. Sevadjian, and A. H.  
662 V. Timmerman. 2013. Relationships between phytoplankton thin layers and the  
663 fine-scale vertical distributions of two trophic levels of zooplankton. *J. Plankton*  
664 *Res.* 35:939-956
- 665 Hinton, G., L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V.  
666 Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury. 2012. Deep neural  
667 networks for acoustic modeling in speech recognition: The shared views of four  
668 research groups. *IEEE Signal Process. Mag.* 29:82-97
- 669 Hu, Q., and C. Davis. 2005. Automatic plankton image recognition with co-occurrence  
670 matrices and support vector machine. *Mar. Ecol. Prog. Ser.* 295:21-31
- 671 Hu, Q., and C. Davis. 2006. Accurate automatic quantification of taxa-specific plankton  
672 abundance using dual classification with correction. *Mar. Ecol. Prog. Ser.* 306:51-  
673 61
- 674 Iyer, N. 2012. Machine vision assisted *In Situ* Ichthyoplankton Imaging System. M.S.  
675 Thesis, Purdue University, pp. 61
- 676 Jeffries, H. P., M. S. Berman, A. D. Poularikas, C. Katsinis, I. Melas, K. Sherman, and L.  
677 Bivins. 1984. Automated sizing, counting and identification of zooplankton by  
678 pattern recognition. *Mar. Biol.* 78:329-334
- 679 Jeffries, H. P., K. Sherman, R. Maurer, and C. Katsinis. 1980. Computer processing of  
680 zooplankton samples, pp 303-316. *In: Kennedy, V. (Ed.) Estuarine perspectives.*  
681 Academic Press, New York

- 682 Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. Imagenet classification with deep  
683 convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 1097-1105
- 684 Laney, S. R., and H. M. Sosik. 2014. Phytoplankton assemblage structure in and around a  
685 massive under-ice bloom in the Chukchi Sea. *Deep Sea Res. II* 105:30-41.
- 686 Lawrence, S., C. L. Giles, A. C. Tsoi, and A. D. Back. 1997. Face recognition: A  
687 convolutional neural-network approach. *IEEE Trans. Neural Netw.* 8:98-113
- 688 LeCun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. *Nature* 521:436-444
- 689 Luo, J. Y., B. Grassian, D. Tang, J.-O. Irisson, A. T. Greer, C. M. Guigand, S.  
690 McClatchie, and R. K. Cowen. 2014. Environmental drivers of the fine-scale  
691 distribution of a gelatinous zooplankton community across a mesoscale front.  
692 *Mar. Ecol. Prog. Ser.* 510:129-149
- 693 Luo, T., K. Kramer, D. B. Goldgof, L. O. Hall, S. Samson, A. Remsen, and T. Hopkins.  
694 2005. Active learning to recognize multiple types of plankton. *J. Mach. Learn.*  
695 *Res.* 6:589-613
- 696 McClatchie, S., R. Cowen, K. Nieto, A. Greer, J. Y. Luo, C. M. Guigand, D. Demer, D.  
697 Griffith, and D. Rudnick. 2012. Resolution of fine biological structure including  
698 small narcomedusae across a front in the Southern California Bight. *J. Geophys.*  
699 *Res.-Oceans* 117:C04020. doi:[10.1029/2011JC007565](https://doi.org/10.1029/2011JC007565).
- 700 Ohman, M. D., J. R. Powell, M. Picheral, and D. W. Jensen. 2012. Mesozooplankton and  
701 particulate matter responses to a deep-water frontal system in the southern  
702 California Current System. *J. Plankton Res.* 34:815-827
- 703 Ortner, P. B., S. R. Cummings, and R. P. Astring. 1979. Silhouette photography of  
704 oceanic zooplankton. *Nature* 277:50-51.

- 705 Ortner, P.B., L. C. Hill, and H.E. Edgerton. 1981. In-situ silhouette photography of Gulf  
706 Stream zooplankton. *Deep-Sea Res. I* 28A:1569-1576.
- 707 Petrik, C. M., G. A. Jackson, and D. M. Checkley, Jr. 2013. Aggregates and their  
708 distributions determined from LOPC observations made using an autonomous  
709 profiling float. *Deep Sea Res. I* 74:64-81
- 710 Picheral, M., L. Guidi, L. Stemmann, D. Karl, G. Iddaoud, and G. Gorsky. 2010. The  
711 Underwater Vision Profiler 5: An advanced instrument for high spatial resolution  
712 studies of particle size spectra and zooplankton. *Limnol. Oceanogr. Meth.* 8:462-  
713 473
- 714 Richardson, D. E., J. K. Llopiz, C. M. Guigand, and R. K. Cowen. 2010. Larval  
715 assemblages of large and medium-sized pelagic species in the Straits of Florida.  
716 *Prog. Oceanogr.* 86:8-20
- 717 Robinson, K. L., J. Y. Luo, S. Sponaugle, C. Guigand, and R. K. Cowen, Tale of Two  
718 Crowds. *In review.*
- 719 Samson, S., T. Hopkins, A. Remsen, L. Langebrake, T. Sutton, and J. Patten. 2001. A  
720 system for high resolution zooplankton imaging. *IEEE J. Ocean. Eng.* 26(4):671-  
721 676
- 722 Schmid, M. S., C. Aubry, J. Grigor, and L. Fortier. 2016. The LOKI underwater imaging  
723 system and an automatic identification model for the detection of zooplankton  
724 taxa in the Arctic Ocean. *Meth. Oceanogr.* 15-16:129-160.
- 725 Simpson, R., R. Williams, R. Ellis, and P. F. Culverhouse. 1992. Biological pattern  
726 recognition by neural networks. *Mar. Ecol. Prog. Ser.* 79:303-308

- 727 Tsechpenakis, G., C. Guigand, and R. K. Cowen. 2007. Image analysis techniques to  
728 accompany a new *In Situ* Ichthyoplankton Imaging System OCEANS 2007, p 1-6
- 729 Tsechpenakis, G., C. Guigand, and R. K. Cowen. 2008. Machine vision-assisted *In Situ*  
730 Ichthyoplankton Imaging System. *Sea Technology* 49:15-20

For Review Only

## TABLES

**Table 1.** Comparison of precision, recall, and F1-ratio for the prediction of the training set compared to the full dataset (from the random 75,000 spot-checks, after applying the filtering thresholds), calculated at the group level. Gray rows indicate rare groups with < 25 images in the spot-checked set.

Class	Training Set			Full Dataset		
	Precision	Recall	F1	Precision	Recall	F1
appendicularian	97.5	99.2	98.3	90	39.1	54.5
artifact	94.6	96.8	95.7	96.5	90.5	93.4
chaetognath	98.7	99	98.8	92.7	46.9	62.3
copepod_calanoid	98.3	99.2	98.7	98.8	62.4	76.5
copepod_copilia	97.7	97.7	97.7	60	75	66.7
copepod_oithona	97.5	99.8	98.6	100	57.6	73.1
ctenophore_beroida	98.8	89.4	93.9	100	20	33.3
ctenophore_cestida	100	99.3	99.6	100	33.3	50
ctenophore_cydippid	100	83.3	90.9	100	3.1	6
ctenophore_lobata	99.3	98.7	99	100	14.3	25
detritus	97.4	92.6	94.9	98.2	55.3	70.8
diatom_chain	97.6	98	97.8	92.3	78.6	84.9
echinoderm	98.2	98.6	98.4	100	16.5	28.3
ephyra	100	100	100	52.8	47.5	50
fish	99.5	99.8	99.6	76.3	38.5	51.2
hydro_liriope	93.5	96.1	94.8	100	21.4	35.3
hydro_narcomedusae	97.4	95.5	96.4	98.8	23	37.3
hydro_other	88.8	92.3	90.5	79.7	15	25.2
hydro_rhopalonematidae	95.4	97	96.2	96.7	33.1	49.3
medusa_pelagia	98.8	95.3	97	NA	0	NA
polychaete_worm	99	99	99	90	45.8	60.7
protist	99.1	99.6	99.3	93.9	53.7	68.3
pteropod	99	94.6	96.8	55.6	14.7	23.3
radiolarian_chain	100	100	100	100	77.8	87.5
shrimp_euphausiid	95.8	98.8	97.3	94.1	69.6	80
shrimp_lucifer	93.1	94.7	93.9	100	32.5	49.1
shrimp_other	97.4	91.7	94.5	86.2	12.3	21.5
shrimp_sergestidae	92.1	89.4	90.7	90	50	64.3
siph_ablyidae	96.7	97.8	97.2	87.5	21.9	35
siph_calycophoran	95.9	95.6	95.7	98.6	24.6	39.4
siph_lilyopsis_rosea	92	89.6	90.8	100	44.4	61.5
siph_physonect	95	80	86.9	NA	0	NA
stomatopods	95.3	95.3	95.3	13.3	66.7	22.2
tornaria	98.4	98.4	98.4	50	25	33.3
trichodesmium	92.3	95.5	93.9	93.2	29.5	44.8
tunicates	97.8	99.2	98.5	95.5	50.9	66.4
zoa	97.9	100	98.9	1.5	92.7	3

## FIGURES

**Figure 1.** Sampling sites in the northern Gulf of Mexico (GOM), spanning two months (July and August 2011), in nearshore and offshore sites, occurring during evening (solid lines) or morning (dashed lines) times. Each transect was sampled over 6 hours, and consisted of tow-yo undulations, from the surface to a maximum of 50 m for the inshore sites and 130 m for the offshore sites.

**Figure 2.** Flowchart overview of the processing steps. Raw frames were first flat-fielded and corrected, then segmented into smaller image segments. A training set was generated to train the image classifier, which was a convolutional neural net (CNN). The full dataset of ~24M images were classified. Afterwards, a random subset of 75k images were spot-checked (manually validated) to estimate the accuracy of the classifier. Separately, a 43k-test set was also validated and used to set probability thresholds, which separated the classified dataset into low probability (discarded into “unknown”) and high probability images (retained). Confusion matrices (CMs) were generated to evaluate classifier performance at each step.

**Figure 3.** An example of the flat-fielding process, showing the raw image (left), the flat-field frame that was removed from the raw image (center), and the corrected frame (right).

**Figure 4 (A,B).** Example images from each of the 108 classes within the learning set. Classes [and corresponding groups] are: 1) acantharia protist 1, 2) acantharia protist 2 [*protist*]; 3) appendicularian sinusoidal tail, 4) appendicularian slight curve, 5) appendicularian straight [*appendicularian*]; 6) artifact 1, 7) artifact 2, 8) bubbles [*artifact*]; 9) chaetognath c-curved, 10)

chaetognath curved, 11) chaetognath dark, 12) chaetognath straight [*chaetognath*]; 13) copepod calanoid, 14) copepod calanoid eggs, 15) copepod calanoid eucalanus, 16) copepod calanoid flatheads, 17) copepod calanoid frilly antennae, 18) copepod calanoid large, 19) copepod calanoid small long-antennae [*copepod calanoid*]; 20) copepod cyclopoid copilia [*copepod copilia*]; 21) copepod cyclopoid oithona, 22) copepod cyclopoid oithona eggs [*copepod oithona*]; 23) copepod escape [*copepod calanoid*]; 24) ctenophore beroida [*ctenophore beroida*]; 25) ctenophore cestida [*ctenophore cestida*]; 26) ctenophore cydippid [*ctenophore cydippid*]; 27) ctenophore lobata mnemiopsis, 28) ctenophore lobata ocyropsis, 29) ctenophore lobata type 1 [*ctenophore lobata*]; 30) detritus sparse blob, 31) detritus casings, 32) detritus dark, 33) detritus filamentous [*detritus*]; 34) diatom chain string, 35) diatom chain tube [*diatom chain*]; 36) echinoderm brachiolaria, 37) echinoderm pluteus [*echinoderm*]; 38) ephyra [*ephyra*]; 39) fecal pellets [*detritus*]; 40) fish bregmacerotidae, 41) fish carangidae, 42) fish ceratioidei, 43) fish echeneidae, 44) fish engraulidae, 45) fish gobiidae, 46) fish gonostomatidae, 47) fish labroidei, 48) fish leptocephali, 49) fish microdesmidae, 50) fish myctophidae, 51) fish ophidiidae, 52) fish phosichthyidae, 53) fish pleuronectiformes, 54) fish scombridae, 55) fish serranidae, 56) fish synodontidae, 57) fish trichiuridae, 58) fish xyrichtys [*fish*]; 59) hydromedusae rhopalonematidae [*hydro rhopalonematidae*], 60) hydromedusae eucheilota, 61) hydromedusae haliscera [*hydro other*]; 62) hydromedusa liriopae tetraphylla, 63) hydromedusa liriopae cut-off-bell [*hydro liriopae*]; 64) hydromedusae narcomedusae other [*hydro narcomedusae*]; 65) hydromedusae rhopalonema 2 [*hydro rhopalonematidae*], 66) hydromedusae solmaris rhodoloma, 67) hydromedusae solmaris spp, 68) hydromedusae solmundella, 69) hydromedusae tiny solmaris [*hydro narcomedusae*]; 70) hydromedusae type 1 small bell, 71) hydromedusae type 2, 72) hydromedusae type 3 [*hydro other*]; 73) medusa pelagia noctiluca [*medusa pelagia*];

74) medusa tentacles [*hydro other*]; 75) polychaete type 1, 76) polychaete type 2, 77) polychaete type 3 [*polychaete worm*]; 78) protist noctiluca, 79) protist radiolarian, 80) protist radiolarian clear [*protist*]; 81) pteropod type 1, 82) pteropod type 2, 83) pteropod type 3 [*pteropod*]; 84) radiolarian chain [*radiolarian chain*]; 85) shrimp caridean, 86) shrimp caridean small [*shrimp other*]; 87) shrimp euphausiid, 88) shrimp euphausiid escape posture [*shrimp euphausiid*]; 89) shrimp lucifer [*shrimp lucifer*]; 90) shrimp mysid [*shrimp other*]; 91) shrimp sergestidae [*shrimp sergestidae*]; 92) siphonophore ablyidae [*siph ablyidae*]; 93) siphonophore calycophoran pointy head no-stem, 94) siphonophore calycophoran pointy head with-stem, 95) siphonophore calycophoran round head [*siph calycophoran*]; 96) siphonophore lilyopsis rosea [*siph lilyopsis rosea*]; 97) siphonophore physonect [*siph physonect*]; 98) stomatopods [*stomatopods*]; 99) tornaria acorn worm larvae [*tornaria*]; 100) trichodesmium bow-tie, 101) trichodesmium tuft, 102) trichodesmium puff [*trichodesmium*]; 103) tunicate doliolid, 104) tunicate doliolid budding, 105) tunicate salp, 106) tunicate salp chains [*tunicates*]; 107) unknown dark blob [*detritus*]; 108) zoea [*zoea*].

**Figure 5.** Confusion matrix on the 75,000 random images, classified into 108 classes, and then grouped into 38 groups (including unknowns). Low-probability images were moved into Unknowns. Rows show computer-predicted classes, and columns show human-validated classes. Color indicates proportion of images sorted from computer-predicted classes into manually verified classes, scaled by row. Grey rows indicate rare categories where no (high probability) images were randomly selected for validation.

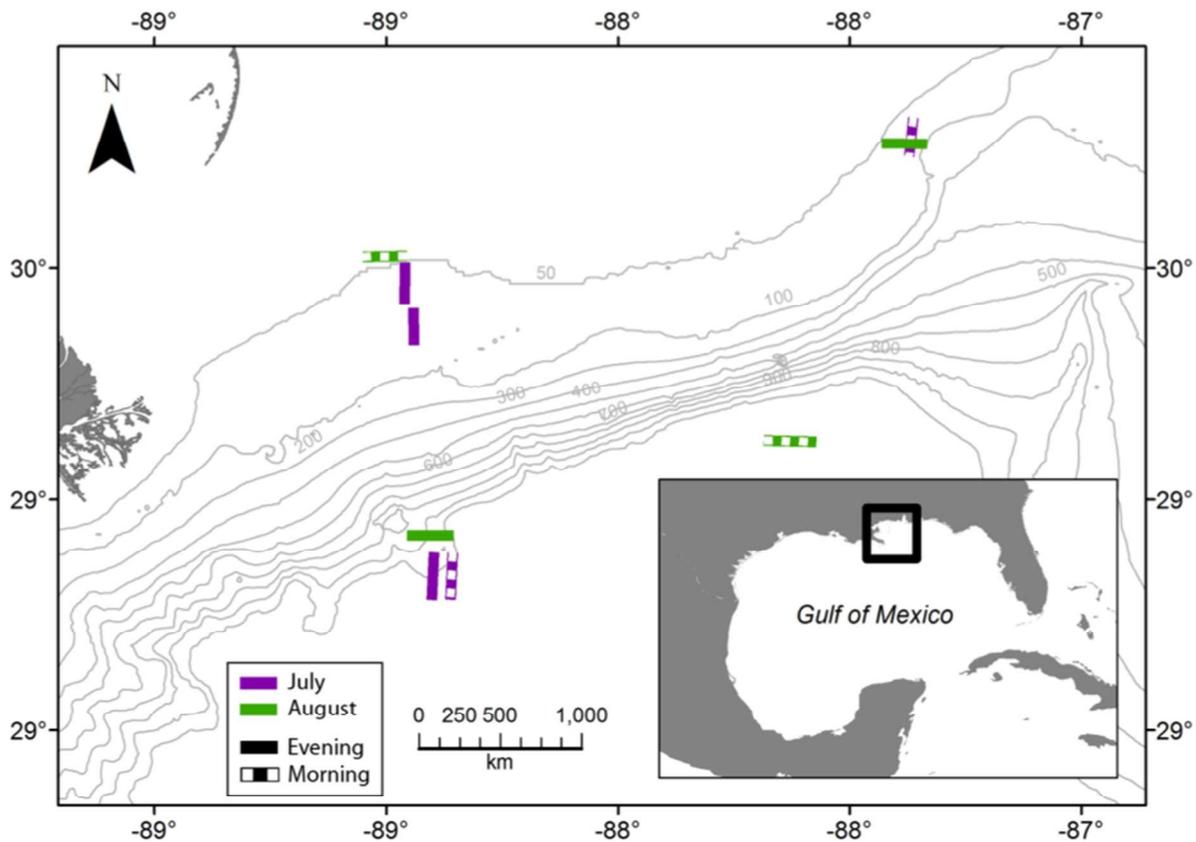


Figure 1.

www Only

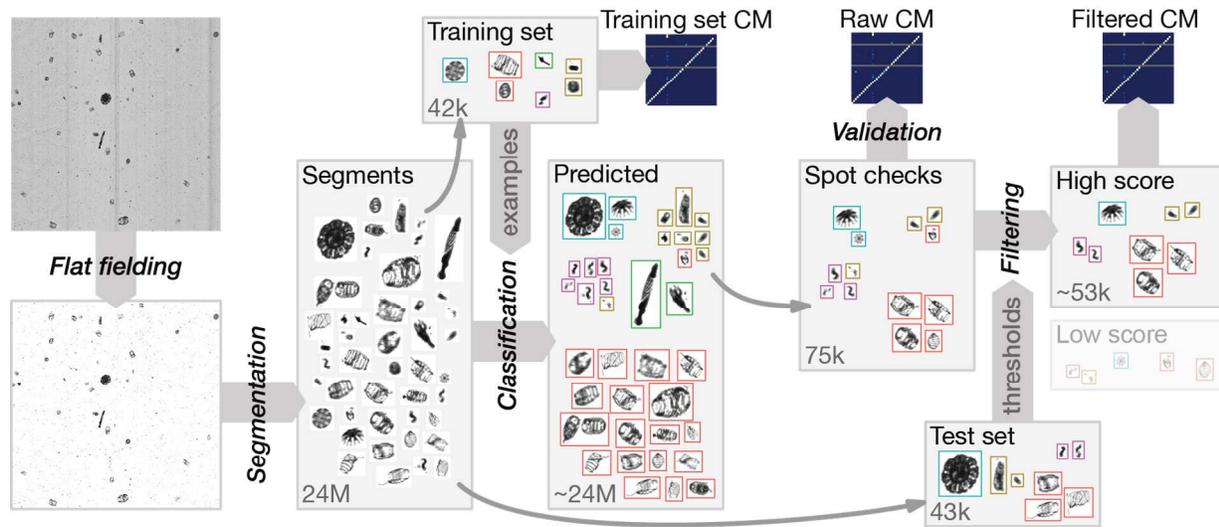


Figure 2.

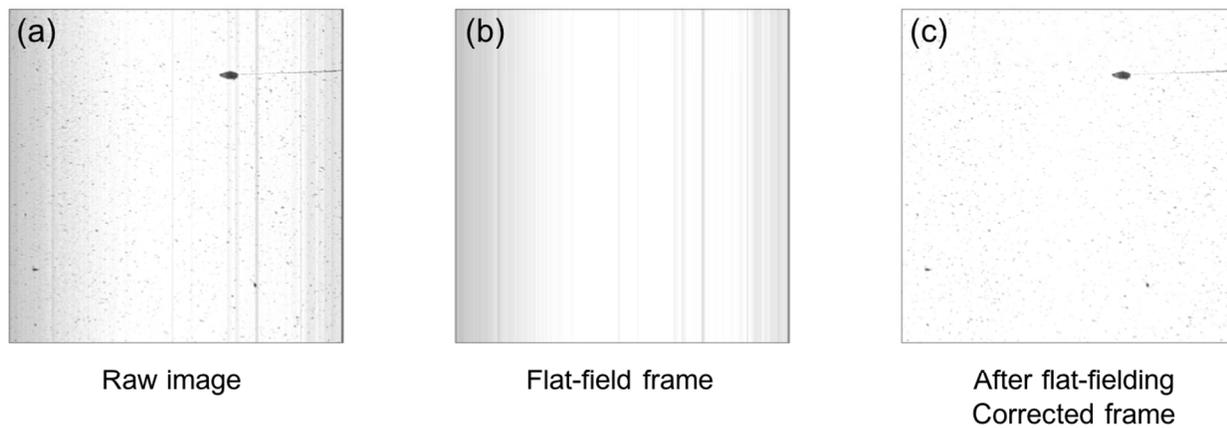
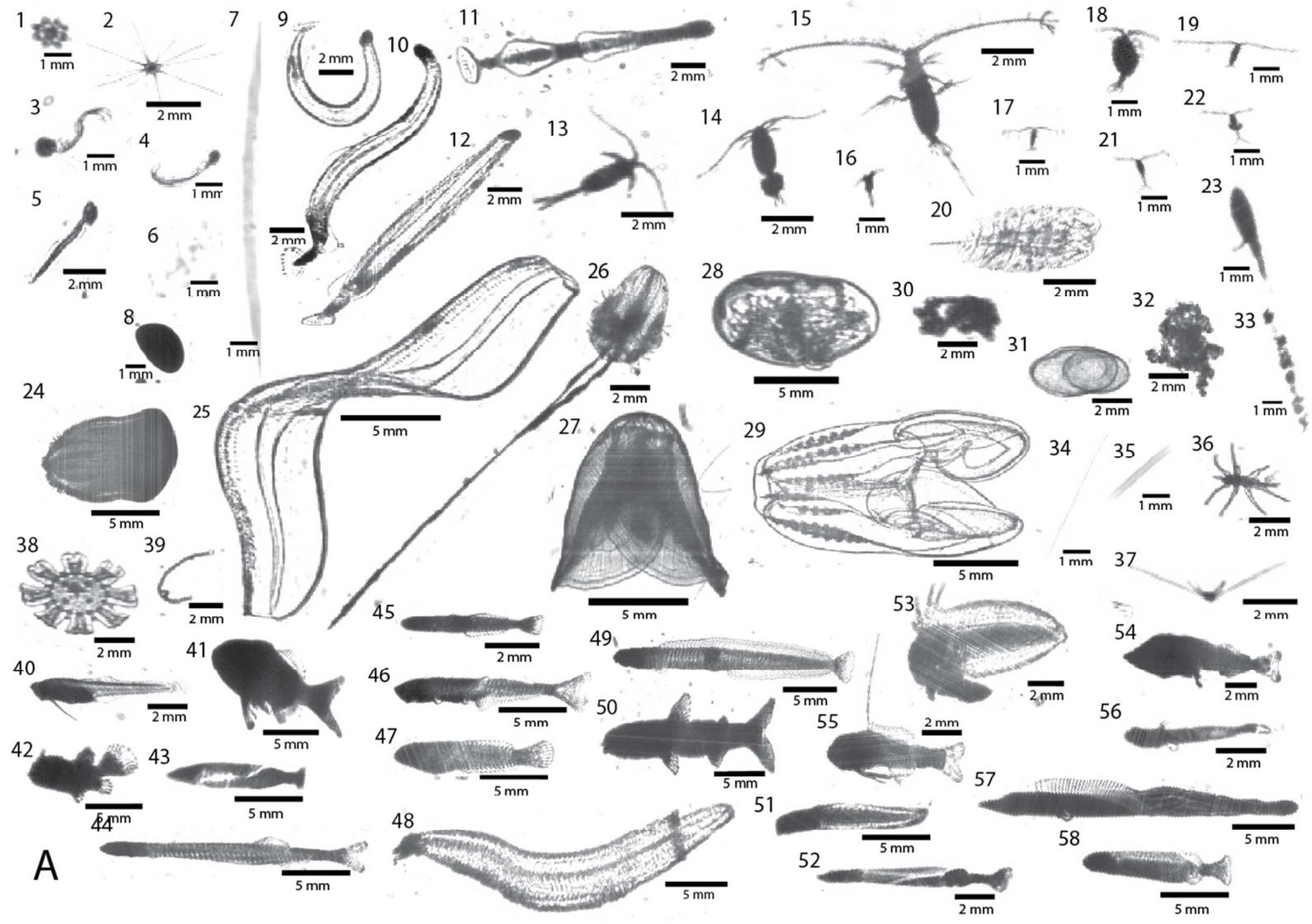


Figure 3.

For Review Only



A

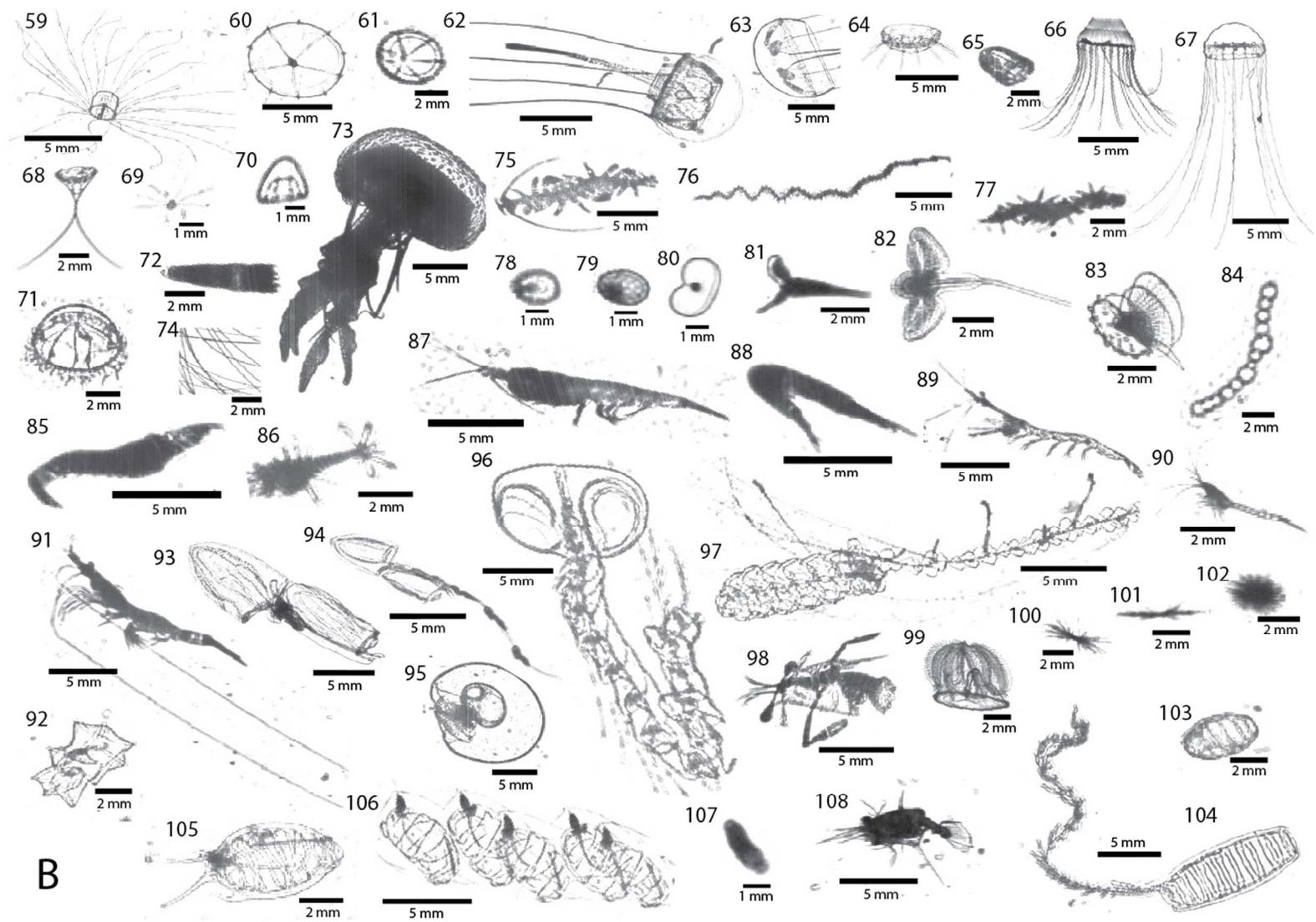


Figure 4 (A,B).



## SUPPLEMENTAL TABLES AND FIGURES

**Table S1.** Confusion matrix statistics of a model prediction on the original 108 class training set, showing number of true positives (TP), false positives (FP), false negatives (FN), total number of objects in each class, precision ( $P = TP / (TP + FP)$ ), recall ( $R = TP / (TP + FN)$ ), and F1-score (harmonic mean of precision and recall,  $F1 = 2 * P * R / (P + R)$ ).

#	Class	TP	FP	FN	Total	Prec.	Recall	F1-score
1	acantharia_protist_type1	73	2	3	<b>76</b>	97.3	96.1	<b>96.7</b>
2	acantharia_protist_type2	504	2	0	<b>504</b>	99.6	100	<b>99.8</b>
3	appendicularian_sinusoidal_tail	1392	175	96	<b>1488</b>	88.8	93.5	<b>91.1</b>
4	appendicularian_slight_curve	956	173	215	<b>1171</b>	84.7	81.6	<b>83.1</b>
5	appendicularian_straight	510	86	67	<b>577</b>	85.6	88.4	<b>87</b>
6	artifact_type1	211	13	4	<b>215</b>	94.2	98.1	<b>96.1</b>
7	artifact_type2	96	3	11	<b>107</b>	97	89.7	<b>93.2</b>
8	bubbles	205	13	2	<b>207</b>	94	99	<b>96.4</b>
9	chaetognath_c_curved	66	2	8	<b>74</b>	97.1	89.2	<b>93</b>
10	chaetognath_curved	65	9	9	<b>74</b>	87.8	87.8	<b>87.8</b>
11	chaetognath_dark	108	5	4	<b>112</b>	95.6	96.4	<b>96</b>
12	chaetognath_straight	120	7	1	<b>121</b>	94.5	99.2	<b>96.8</b>
13	copepod_calanoid	65	8	41	<b>106</b>	89	61.3	<b>72.6</b>
14	copepod_calanoid_eggs	304	29	21	<b>325</b>	91.3	93.5	<b>92.4</b>
15	copepod_calanoid_eucalanus	199	3	9	<b>208</b>	98.5	95.7	<b>97.1</b>
16	copepod_calanoid_flatheads	303	77	58	<b>361</b>	79.7	83.9	<b>81.7</b>
17	copepod_calanoid_fringed_antennae	90	14	46	<b>136</b>	86.5	66.2	<b>75</b>
18	copepod_calanoid_large	705	88	34	<b>739</b>	88.9	95.4	<b>92</b>
19	copepod_calanoid_small_longantennae	173	7	2	<b>175</b>	96.1	98.9	<b>97.5</b>
20	copepod_cyclopoid_copilia	253	6	6	<b>259</b>	97.7	97.7	<b>97.7</b>
21	copepod_cyclopoid_oithona	1685	78	130	<b>1815</b>	95.6	92.8	<b>94.2</b>
22	copepod_cyclopoid_oithona_eggs	2427	170	19	<b>2446</b>	93.5	99.2	<b>96.3</b>
23	copepod_escape	94	6	2	<b>96</b>	94	97.9	<b>95.9</b>
24	ctenophore_beroida	84	1	10	<b>94</b>	98.8	89.4	<b>93.9</b>
25	ctenophore_cestida	133	0	1	<b>134</b>	100	99.3	<b>99.6</b>
26	ctenophore_cydippid	10	0	2	<b>12</b>	100	83.3	<b>90.9</b>
27	ctenophore_lobata_mnemiopsis	656	13	4	<b>660</b>	98.1	99.4	<b>98.7</b>
28	ctenophore_lobata_ocyropsis	12	0	12	<b>24</b>	100	50	<b>66.7</b>
29	ctenophore_lobata_type1	23	0	1	<b>24</b>	100	95.8	<b>97.9</b>
30	detritus_blob_sparse	2315	300	321	<b>2636</b>	88.5	87.8	<b>88.1</b>
31	detritus_casing	182	2	1	<b>183</b>	98.9	99.5	<b>99.2</b>
32	detritus_dark	350	62	58	<b>408</b>	85	85.8	<b>85.4</b>
33	detritus_filamentous	511	93	270	<b>781</b>	84.6	65.4	<b>73.8</b>

34	diatom_chain_string	1101	46	80	<b>1181</b>	96	93.2	<b>94.6</b>
35	diatom_chain_tube	1012	102	58	<b>1070</b>	90.8	94.6	<b>92.7</b>
36	echinoderm_brachiolaria	263	8	2	<b>265</b>	97	99.2	<b>98.1</b>
37	echinoderm_pluteus	288	2	6	<b>294</b>	99.3	98	<b>98.6</b>
38	ephyra	144	0	0	<b>144</b>	100	100	<b>100</b>
39	fecal_pellets	177	12	12	<b>189</b>	93.7	93.7	<b>93.7</b>
40	fish_bregmacerotidae	1756	230	34	<b>1790</b>	88.4	98.1	<b>93</b>
41	fish_carangidae	222	108	74	<b>296</b>	67.3	75	<b>70.9</b>
42	fish_ceratioidei	4	1	16	<b>20</b>	80	20	<b>32</b>
43	fish_echeneidae	63	36	27	<b>90</b>	63.6	70	<b>66.6</b>
44	fish_engraulidae	784	83	107	<b>891</b>	90.4	88	<b>89.2</b>
45	fish_gobiidae	911	118	130	<b>1041</b>	88.5	87.5	<b>88</b>
46	fish_gonostomatidae	38	7	125	<b>163</b>	84.4	23.3	<b>36.5</b>
47	fish_labroidei	825	72	90	<b>915</b>	92	90.2	<b>91.1</b>
48	fish_leptocephali	125	10	14	<b>139</b>	92.6	89.9	<b>91.2</b>
49	fish_microdesmidae	37	2	4	<b>41</b>	94.9	90.2	<b>92.5</b>
50	fish_myctophidae	1187	513	245	<b>1432</b>	69.8	82.9	<b>75.8</b>
51	fish_ophidiidae	75	12	33	<b>108</b>	86.2	69.4	<b>76.9</b>
52	fish_phosichthyidae	73	30	140	<b>213</b>	70.9	34.3	<b>46.2</b>
53	fish_pleuronectiformes	95	7	7	<b>102</b>	93.1	93.1	<b>93.1</b>
54	fish_scombridae	323	91	70	<b>393</b>	78	82.2	<b>80</b>
55	fish_serranidae	6	0	23	<b>29</b>	100	20.7	<b>34.3</b>
56	fish_synodontidae	1	0	172	<b>173</b>	100	0.6	<b>1.2</b>
57	fish_trichiuridae	75	10	3	<b>78</b>	88.2	96.2	<b>92</b>
58	fish_xyrichtys	1057	91	86	<b>1143</b>	92.1	92.5	<b>92.3</b>
59	hydro_rhopalonematidae	675	31	19	<b>694</b>	95.6	97.3	<b>96.4</b>
60	hydro_euceilota_spp	65	0	6	<b>71</b>	100	91.5	<b>95.6</b>
61	hydro_haliscera_spp	42	17	10	<b>52</b>	71.2	80.8	<b>75.7</b>
62	hydro_liriope	118	8	9	<b>127</b>	93.7	92.9	<b>93.3</b>
63	hydro_liriope_cutoffbell	51	8	2	<b>53</b>	86.4	96.2	<b>91</b>
64	hydro_narco_other	170	24	16	<b>186</b>	87.6	91.4	<b>89.5</b>
65	hydro_rhopalonema2	58	11	10	<b>68</b>	84.1	85.3	<b>84.7</b>
66	hydro_solmaris_rhodoloma	401	20	54	<b>455</b>	95.2	88.1	<b>91.5</b>
67	hydro_solmaris_spp	191	5	12	<b>203</b>	97.4	94.1	<b>95.7</b>
68	hydro_solmundella	302	2	2	<b>304</b>	99.3	99.3	<b>99.3</b>
69	hydro_tinysolmaris	52	28	19	<b>71</b>	65	73.2	<b>68.9</b>
70	hydro_type1_smallbell	105	35	22	<b>127</b>	75	82.7	<b>78.7</b>
71	hydro_type2	112	9	17	<b>129</b>	92.6	86.8	<b>89.6</b>
72	hydro_type3	95	8	2	<b>97</b>	92.2	97.9	<b>95</b>
73	medusa_Pelagia	82	1	4	<b>86</b>	98.8	95.3	<b>97</b>
74	medusa_tentacles	65	17	7	<b>72</b>	79.3	90.3	<b>84.4</b>
75	polychaete_type1	27	2	6	<b>33</b>	93.1	81.8	<b>87.1</b>
76	polychaete_type2	97	0	3	<b>100</b>	100	97	<b>98.5</b>
77	polychaete_type3	163	10	3	<b>166</b>	94.2	98.2	<b>96.2</b>

78	protist_noctiluca	100	4	3	<b>103</b>	96.2	97.1	<b>96.6</b>
79	protist_rad	108	3	0	<b>108</b>	97.3	100	<b>98.6</b>
80	protist_rad_clear	103	1	2	<b>105</b>	99	98.1	<b>98.5</b>
81	pteropod_type1	164	6	12	<b>176</b>	96.5	93.2	<b>94.8</b>
82	pteropod_type2	64	0	5	<b>69</b>	100	92.8	<b>96.3</b>
83	pteropod_type3	67	1	4	<b>71</b>	98.5	94.4	<b>96.4</b>
84	radiolarian_chain	113	0	0	<b>113</b>	100	100	<b>100</b>
85	shrimp_caridae	71	14	38	<b>109</b>	83.5	65.1	<b>73.2</b>
86	shrimp_caridae_sm	277	10	12	<b>289</b>	96.5	95.8	<b>96.1</b>
87	shrimp_euphausiid	921	38	3	<b>924</b>	96	99.7	<b>97.8</b>
88	shrimp_euphausiid_escape	47	6	10	<b>57</b>	88.7	82.5	<b>85.5</b>
89	shrimp_lucifer	2544	189	143	<b>2687</b>	93.1	94.7	<b>93.9</b>
90	shrimp_mysid	81	0	2	<b>83</b>	100	97.6	<b>98.8</b>
91	shrimp_sergestidae	1693	146	200	<b>1893</b>	92.1	89.4	<b>90.7</b>
92	siph_ablyidae	753	26	17	<b>770</b>	96.7	97.8	<b>97.2</b>
93	siph_calycophoran_pointy_notail	235	23	13	<b>248</b>	91.1	94.8	<b>92.9</b>
94	siph_calycophoran_pointy_tail	104	9	8	<b>112</b>	92	92.9	<b>92.4</b>
95	siph_calycophoran_round	915	44	58	<b>973</b>	95.4	94	<b>94.7</b>
96	siph_lilyopsis_rosea	103	9	12	<b>115</b>	92	89.6	<b>90.8</b>
97	siph_physonect	76	4	19	<b>95</b>	95	80	<b>86.9</b>
98	stomatopods	101	5	5	<b>106</b>	95.3	95.3	<b>95.3</b>
99	tornaria_acorn_worm_larvae	63	1	1	<b>64</b>	98.4	98.4	<b>98.4</b>
100	tricho_bow_tie	171	15	13	<b>184</b>	91.9	92.9	<b>92.4</b>
101	tricho_tuft	95	28	19	<b>114</b>	77.2	83.3	<b>80.1</b>
102	tricho_puff	103	3	0	<b>103</b>	97.2	100	<b>98.6</b>
103	tunicate_doliolid	467	19	8	<b>475</b>	96.1	98.3	<b>97.2</b>
104	tunicate_doliolid_budding	266	2	2	<b>268</b>	99.3	99.3	<b>99.3</b>
105	tunicate_salp	92	8	21	<b>113</b>	92	81.4	<b>86.4</b>
106	tunicate_salp_chains	92	16	1	<b>93</b>	85.2	98.9	<b>91.5</b>
107	unknown_dark_blob	177	54	79	<b>256</b>	76.6	69.1	<b>72.7</b>
108	zoaea	46	1	0	<b>46</b>	97.9	100	<b>98.9</b>
	<b>Total</b>	<b><math>\Sigma</math>TP</b>	<b><math>\Sigma</math>FP</b>	<b><math>\Sigma</math>FN</b>	<b><math>\Sigma</math>Tot.</b>	<b><math>\overline{Prec.}</math></b>	<b><math>\overline{Recall}</math></b>	<b><math>\overline{F1}</math></b>
		<b>38605</b>	<b>3959</b>	<b>3959</b>	<b>42564</b>	<b>91.5</b>	<b>87.3</b>	<b>88.1</b>

**Table S2.** Confusion matrix on spot-checks of 75,000 random images, showing values with and without probability filtering. Images were pulled from a random subset of the whole dataset, so proportions of images in each category reflects the proportions in the data. The gray rows indicates rare classes, where less than 25 images were randomly pulled out for spot-checks. Type B = biological, and NB = non-biological. TP = True Positives, FP = False Positives, and FN = False Negatives. Note that for each group, TP + FN = Total count. F1 ratio is the harmonic mean between precision and recall.

Class	Type	Total #	Without probability filtering						With probability filtering					
			TP	FP	FN	Precision	Recall	F1	TP	FP	FN	Precision	Recall	F1
<b>All groups</b>		<b>75000</b>	<b>65515</b>	<b>9485</b>	<b>9485</b>	<b>53.2</b>	<b>63.3</b>	<b>54.0</b>	<b>51285</b>	<b>23715</b>	<b>23715</b>	<b>84.2</b>	<b>39.9</b>	<b>49.3</b>
<b>Biological groups, non-rare</b>		<b>15286</b>	<b>12316</b>	<b>6296</b>	<b>2970</b>	<b>54.6</b>	<b>69.6</b>	<b>57.3</b>	<b>8949</b>	<b>631</b>	<b>6337</b>	<b>90.7</b>	<b>38.9</b>	<b>51.1</b>
appendicularian	B	1218	915	947	303	49.1	75.1	59.4	476	53	742	90	39.1	54.5
artifact	NB	26112	24955	1490	1157	94.4	95.6	95	23632	850	2480	96.5	90.5	93.4
chaetognath	B	1608	909	115	699	88.8	56.5	69.1	754	59	854	92.7	46.9	62.3
copepod_calanoid	B	2067	1703	152	364	91.8	82.4	86.8	1290	16	777	98.8	62.4	76.5
copepod_copilia	B	12	12	117	0	9.3	100	17	9	6	3	60	75	66.7
copepod_oithona	B	663	647	493	16	56.8	97.6	71.8	382	0	281	100	57.6	73.1
ctenophore_beroida	B	10	4	3	6	57.1	40	47	2	0	8	100	20	33.3
ctenophore_cestida	B	6	3	0	3	100	50	66.7	2	0	4	100	33.3	50
ctenophore_cydippid	B	32	7	10	25	41.2	21.9	28.6	1	0	31	100	3.1	6
ctenophore_lobata	B	7	3	2	4	60	42.9	50	1	0	6	100	14.3	25
detritus	NB	33141	28182	1501	4959	94.9	85	89.7	18342	327	14799	98.2	55.3	70.8
diatom_chain	B	5817	5569	2321	248	70.6	95.7	81.3	4574	382	1243	92.3	78.6	84.9
echinoderm	B	91	45	129	46	25.9	49.5	34	15	0	76	100	16.5	28.3
ephyra	B	40	23	51	17	31.1	57.5	40.4	19	17	21	52.8	47.5	50
fish	B	117	90	258	27	25.9	76.9	38.7	45	14	72	76.3	38.5	51.2
hydro_liriope	B	14	12	11	2	52.2	85.7	64.9	3	0	11	100	21.4	35.3
hydro_narcomedusae	B	369	330	303	39	52.1	89.4	65.8	85	1	284	98.8	23	37.3
hydro_other	B	341	177	437	164	28.8	51.9	37	51	13	290	79.7	15	25.2
hydro_rhopalonematidae	B	178	173	125	5	58.1	97.2	72.7	59	2	119	96.7	33.1	49.3
medusa_pelagia	B	1	0	0	1	NA	0	NA	0	0	1	NA	0	NA

polychaete_worm	B	59	47	203	12	18.8	79.7	30.4	27	3	32	90	45.8	60.7
protist	B	1140	669	87	471	88.5	58.7	70.6	612	40	528	93.9	53.7	68.3
pteropod	B	34	10	55	24	15.4	29.4	20.2	5	4	29	55.6	14.7	23.3
radiolarian_chain	B	45	35	0	10	100	77.8	87.5	35	0	10	100	77.8	87.5
shrimp_euphausiid	B	46	38	14	8	73.1	82.6	77.6	32	2	14	94.1	69.6	80
shrimp_lucifer	B	77	70	38	7	64.8	90.9	75.7	25	0	52	100	32.5	49.1
shrimp_other	B	203	75	91	128	45.2	36.9	40.6	25	4	178	86.2	12.3	21.5
shrimp_sergestidae	B	18	17	25	1	40.5	94.4	56.7	9	1	9	90	50	64.3
siph_ablyidae	B	32	31	129	1	19.4	96.9	32.3	7	1	25	87.5	21.9	35
siph_calycophoran	B	285	213	164	72	56.5	74.7	64.3	70	1	215	98.6	24.6	39.4
siph_lilyopsis_rosea	B	9	4	1	5	80	44.4	57.1	4	0	5	100	44.4	61.5
siph_physonect	B	20	2	6	18	25	10	14.3	0	0	20	NA	0	NA
stomatopods	B	3	2	19	1	9.5	66.7	16.6	2	13	1	13.3	66.7	22.2
tornaria	B	4	1	8	3	11.1	25	15.4	1	1	3	50	25	33.3
trichodesmium	B	278	126	30	152	80.8	45.3	58.1	82	6	196	93.2	29.5	44.8
tunicates	B	546	414	144	132	74.2	75.8	75	278	13	268	95.5	50.9	66.4
unknown	NB	354	0	0	354	NA	0	NA	328	21885	26	1.5	92.7	3
zoea	B	3	2	6	1	25	66.7	36.4	1	1	2	50	33.3	40

**Table S3.** Probability filtering of the 108 classes, showing the original classes, groupings, total original counts, values of the probability thresholds per class, percent retained after filtering, and percent discarded after filtering.

#	Class	Group	Total	Prob. filter	% Kept	% Discard
1	acantharia_protist_type1	protist	25886	0.804	17.5	82.5
2	acantharia_protist_type2	protist	78880	0.4161	93.4	6.6
3	appendicularian_sinusoidal_tail	appendicularian	208569	0.8189	51.91	48.09
4	appendicularian_slight_curve	appendicularian	228665	0.9191	11.98	88.02
5	appendicularian_straight	appendicularian	137453	0.7857	21.25	78.75
6	artifact_type1	artifact	8201765	0.7199	92.63	7.37
7	artifact_type2	artifact	11118	0.996	1.78	98.22
8	bubbles	artifact	10017	0.9898	25.13	74.87
9	chaetognath_c_curved	chaetognath	1545	0.4244	80.32	19.68
10	chaetognath_curved	chaetognath	3701	0.3247	90.73	9.27
11	chaetognath_dark	chaetognath	259481	0.5445	77.42	22.58
12	chaetognath_straight	chaetognath	60903	0.1848	99.78	0.22
13	copepod_calanoid	copepod_calanoid	67528	0.1621	99.94	0.06
14	copepod_calanoid_eggs	copepod_calanoid	35059	0.6059	59.6	40.4
15	copepod_calanoid_eucalanus	copepod_calanoid	37834	0.1859	99.36	0.64
16	copepod_calanoid_flatheads	copepod_calanoid	133661	0.2157	99.75	0.25
17	copepod_calanoid_frillyAntennae	copepod_calanoid	35423	0.4735	55.81	44.19
18	copepod_calanoid_large	copepod_calanoid	163326	0.6264	57.64	42.36
19	copepod_calanoid_small_longantennae	copepod_calanoid	18037	0.7493	31.25	68.75
20	copepod_cyclopoid_copilia	copepod_copilia	40258	0.9438	8.38	91.62
21	copepod_cyclopoid_oithona	copepod_oithona	228400	0.9265	31.81	68.19
22	copepod_cyclopoid_oithona_eggs	copepod_oithona	143712	0.9404	34.12	65.88
23	copepod_escape	copepod_calanoid	73310	0.8732	22.01	77.99
24	ctenophore_beroida	ctenophore_beroida	1033	0.8787	26.82	73.18
25	ctenophore_cestida	ctenophore_cestida	714	0.308	92.72	7.28
26	ctenophore_cydippid	ctenophore_cydippid	4481	0.9262	7.5	92.5
27	ctenophore_lobata_mnemiopsis	ctenophore_lobata	2110	0.9941	30.95	69.05
28	ctenophore_lobata_ocyropsis	ctenophore_lobata	363	0.795	9.37	90.63
29	ctenophore_lobata_type1	ctenophore_lobata	271	0.7904	41.7	58.3
30	detritus_blob_sparse	detritus	6795341	0.6826	65.37	34.63
31	detritus_casing	detritus	10535	0.9792	10.87	89.13
32	detritus_dark	detritus	879666	0.1896	99.98	0.02

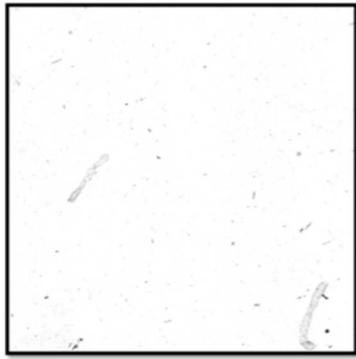
33	detritus_filamentous	detritus	1104433	0.5946	41.31	58.69
34	diatom_chain_string	diatom_chain	1599323	0.5908	83.07	16.93
35	diatom_chain_tube	diatom_chain	861349	0.9475	24.98	75.02
36	echinoderm_brachiolaria	echinoderm	19850	0.8697	14.34	85.66
37	echinoderm_pluteus	echinoderm	39949	0.935	10.89	89.11
38	ephyrae	ephyra	22441	0.62	41.57	58.43
39	fecal_pellets	detritus	337405	0.8071	19.58	80.42
40	fish_bregmacerotidae	fish_bregmacerotidae	23696	0.8179	20.02	79.98
41	fish_carangidae	fish_other	2922	0.6751	4.24	95.76
42	fish_ceratioidei	fish_other	18	0.4	16.67	83.33
43	fish_echeneidae	fish_other	506	0.6373	6.32	93.68
44	fish_engraulidae	fish_engraulidae	17051	0.8873	7.72	92.28
45	fish_gobiidae	fish_gobiidae	1738	0.4473	65.94	34.06
46	fish_gonostomatidae	fish_other	27	0.4	44.44	55.56
47	fish_labroidei	fish_other	1607	0.5223	50.16	49.84
48	fish_leptocephali	fish_other	8720	0.8931	9.45	90.55
49	fish_microdesmidae	fish_other	27	0.293	77.78	22.22
50	fish_myctophidae	fish_myctophidae	34205	0.5174	28.93	71.07
51	fish_ophidiidae	fish_other	352	0.2741	72.44	27.56
52	fish_phosichthyidae	fish_other	3590	0.4406	16.69	83.31
53	fish_pleuronectiformes	fish_other	5622	0.4677	26.08	73.92
54	fish_scombridae	fish_other	6305	0.7189	3.19	96.81
55	fish_serranidae	fish_other	5	0.4	20	80
56	fish_synodontidae	fish_other	401	0.1974	52.12	47.88
57	fish_trichiuridae	fish_other	290	0.441	45.86	54.14
58	fish_xyrichtys	fish_xyrichtys	3949	0.6252	38.72	61.28
59	hydro_rhopalonematidae	hydro_rhopalonematidae	39185	0.9852	32.46	67.54
60	hydro_euchelota_spp	hydro_other	839	0.9346	33.97	66.03
61	hydro_haliscera_spp	hydro_other	2121	0.9059	7.87	92.13
62	hydro_liriope	hydro_liriope	5247	0.9686	19.5	80.5
63	hydro_liriope_cutoffbell	hydro_liriope	1427	0.9894	7.22	92.78
64	hydro_narco_other	hydro_narcomedusae	16836	0.9913	4.35	95.65
65	hydro_rhopalonema2	hydro_rhopalonematidae	42323	0.9673	4.4	95.6
66	hydro_solmaris_rhodoloma	hydro_narcomedusae	50965	0.7825	20.12	79.88
67	hydro_solmaris_spp	hydro_narcomedusae	2923	0.9069	36.02	63.98
68	hydro_solmundella	hydro_narcomedusae	26406	0.9221	36.36	63.64
69	hydro_tinysolmaris	hydro_narcomedusae	100296	0.7853	6.02	93.98
70	hydro_type1_smallbell	hydro_other	30137	0.7628	7.91	92.09
71	hydro_type2	hydro_other	7359	0.6295	38.71	61.29
72	hydro_type3	hydro_other	3639	0.8075	15.58	84.42
73	medusa_Pelagia	medusa_pelagia	447	0.8918	34	66
74	medusa_tentacles	hydro_other	143834	0.8972	11.42	88.58
75	polychaete_type1	polychaete_worm	3916	0.6359	9.22	90.78

76	polychaete_type2	polychaete_worm	1827	0.2887	96.5	3.5
77	polychaete_type3	polychaete_worm	76895	0.9183	9.37	90.63
78	protist_noctiluca	protist	30117	0.2528	99.32	0.68
79	protist_rad	protist	52375	0.2288	99.36	0.64
80	protist_rad_clear	protist	23841	0.2615	99.88	0.12
81	pteropod_type1	pteropod	8391	0.665	15.58	84.42
82	pteropod_type2	pteropod	5806	0.6687	18.72	81.28
83	pteropod_type3	pteropod	5762	0.7679	14.7	85.3
84	radiolarian_chain	radiolarian_chain	10579	0.2932	98.15	1.85
85	shrimp_caridae	shrimp_other	4617	0.8125	8.34	91.66
86	shrimp_caridae_sm	shrimp_other	40329	0.7715	19.08	80.92
87	shrimp_euphausiid	shrimp_euphausiid	15380	0.5932	61.58	38.42
88	shrimp_euphausiid_escape	shrimp_euphausiid	2209	0.4812	46.27	53.73
89	shrimp_lucifer	shrimp_lucifer	31633	0.8455	35.25	64.75
90	shrimp_mysid	shrimp_other	5148	0.8495	18.51	81.49
91	shrimp_sergestidae	shrimp_sergestidae	15238	0.9732	15.75	84.25
92	siph_ablyidae	siph_ablyidae	50518	0.9952	6.45	93.55
93	siph_calycophoran_pointy_notail	siph_calycophoran	41298	0.8834	27.31	72.69
94	siph_calycophoran_pointy_tail	siph_calycophoran	14650	0.6591	47.55	52.45
95	siph_calycophoran_round	siph_calycophoran	59727	0.9841	6.63	93.37
96	siph_lilyopsis_rosea	siph_lilyopsis_rosea	1843	0.5354	72.82	27.18
97	siph_physonect	siph_physonect	4695	0.883	17.98	82.02
98	stomatopods	stomatopods	6015	0.3433	62.39	37.61
99	tornaria_acorn_worm_larvae	tornaria	3096	0.7542	9.46	90.54
100	tricho_bow_tie	trichodesmium	10731	0.5208	72.07	27.93
101	tricho_tuft	trichodesmium	26651	0.6139	46.92	53.08
102	tricho_puff	trichodesmium	11933	0.7488	47.36	52.64
103	tunicate_doliolid	tunicates	111780	0.8433	48.02	51.98
104	tunicate_doliolid_budding	tunicates	25639	0.2869	95.25	4.75
105	tunicate_salp	tunicates	2338	0.7394	12.36	87.64
106	tunicate_salp_chains	tunicates	25450	0.9096	4.46	95.54
107	unknown_dark_blob	detritus	180127	0.7406	14.34	85.66
108	zoa	zoa	1415	0.688	23.53	76.47
		<b>Total</b>	<b>23,380,779</b>		<b>70.36</b>	<b>29.64</b>

## Figure captions

**Figure S1.** Example of the histogram equalization procedure on a flat-fielded frame. After flat-fielding, some images are very white, resulting in insufficient contrast between the target organisms and the background. This causes the segmentation algorithm to not work well and at times miss (not segment) those organisms. In order to increase contrast, we applied histogram equalization, which is a form of histogram modification that redistributes the gray-level values of pixels within an image so that the number of pixels at all gray levels is nearly the same.

**Figure S2.** Signal to Noise Ratios (SNR) on a representative sample of 62 images, half chosen from clear water and the other half chosen from turbid water. The segmentation algorithm often over-segments frames from turbid waters. We used a standard 3x3 median filter to estimate the signal in the frames (after histogram equalization), and then subtracted the signal from the histogram equalized frames for the noise-frames. There is a clear separation in the SNR between noisy and non-noisy images, so a SNR cutoff of 25 was chosen.



Flat-fielded frame

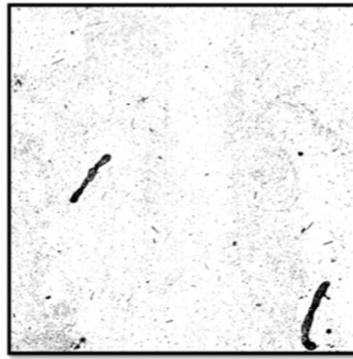
Histogram  
equalized frame

Figure S1.

For Review Only

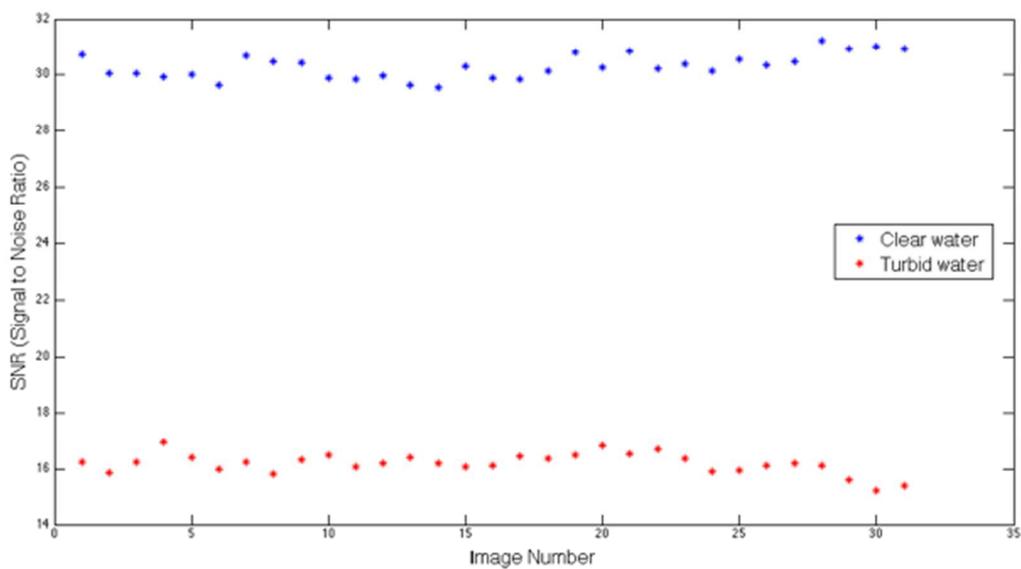


Figure S2.

Review Only