Received: 4 June 2021 | Accepted: 25 August 2021

DOI: 10.1111/2041-210X.13712

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

APPLICATION

Sashimi: A toolkit for facilitating high-throughput organismal image segmentation using deep learning

Shawn T. Schwartz | Michael E. Alfaro

Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California, USA

Correspondence
Shawn T. Schwartz
Email: shawnschwartz@ucla.edu

Handling Editor: Samantha Price

Abstract

1. Digitized specimens are an indispensable resource for rapidly acquiring big datasets and typically must be pre-processed prior to conducting analyses. One crucial image pre-processing step in any image analysis workflow is image segmentation, or the ability to clearly contrast the foreground target from the background noise in an image. This procedure is typically done manually, creating a potential bottleneck for efforts to quantify biodiversity from image databases. Image segmentation meta-algorithms using deep learning provide an opportunity to relax this bottleneck. However, the most accessible pre-trained convolutional neural networks (CNNs) have been trained on a small fraction of biodiversity, thus limiting their utility.
2. We trained a deep learning model to automatically segment target fish from images with both standardized and complex, noisy backgrounds. We then assessed the performance of our deep learning model using qualitative visual inspection and quantitative image segmentation metrics of pixel overlap between reference segmentation masks generated manually by experts and those automatically predicted by our model.
3. Visual inspection revealed that our model segmented fishes with high precision and relatively few artifacts. These results suggest that the meta-algorithm (Mask R-CNN), in which our current fish segmentation model relies on, is well suited for generating high-fidelity segmented specimen images across a variety of background contexts at rapid pace.
4. We present *Sashimi*, a user-friendly command line toolkit to facilitate rapid, automated high-throughput image segmentation of digitized organisms. *Sashimi* is accessible to non-programmers and does not require experience with deep learning to use. The flexibility of Mask R-CNN allows users to generate a segmentation model for use on diverse animal and plant images using transfer learning with training datasets as small as a few hundred images. To help grow the taxonomic scope of images that can be recognized, *Sashimi* also includes a central database for sharing and distributing custom-trained segmentation models of other unrepresented organisms. Lastly, *Sashimi* includes both auxiliary image pre-processing functions useful for some popular downstream color pattern analysis workflows, as well as a simple script to aid users in qualitatively and quantitatively assessing

segmentation model performance for complementary sets of automatically and manually segmented images.

KEYWORDS
automation, big data, convolutional neural network, deep learning, high-throughput, image segmentation, mask R-CNN, reproducibility

1 | INTRODUCTION

Image pre-processing is a fundamental step for any image analysis workflow (Gonzalez & Woods, 2002; Pennekamp & Schtickzelle, 2013), including those for surveying disease in fisheries production, studying animal behaviour, and quantifying phenotypic and morphological diversity (Alfaro et al., 2019; Yao et al., 2013). Nearly every image analysis workflow requires image segmentation, the process of accurately distinguishing the foreground (target) of the image from the background environmental noise in the image (Gonzalez & Woods, 2002; Pennekamp & Schtickzelle, 2013). Even in the best cases where the target is surrounded by a near uniform background, automated algorithms for image segmentation may require manual adjustment, whereas noisy backgrounds with low signal-to-noise ratios may preclude the use of tools in commercial packages such as Adobe Photoshop. As comparative studies in biology move towards more comprehensive sampling of the tree of life, the demands for approaches yielding fast, reliable and reproducible data collection increase. Computer vision techniques provide one means for accurate and scalable image pre-processing in biology (Lürig et al., 2021; Muñoz & Price, 2019; Porto & Voje, 2020).

Machine learning, and especially deep learning algorithms, which permit the identification and classification of complex patterns in noisy environments (Carranza-Rojas et al., 2017; Cheng et al., 2017; Gomez Villa et al., 2017; Joly et al., 2016; Lee et al., 2018; Marques et al., 2018; Norouzzadeh et al., 2018; Qin et al., 2016; Raitoharju et al., 2016; Salman et al., 2016; Wäldchen & Mäder, 2018a, 2018b; Wäldchen et al., 2018; Weinstein, 2018; Willis et al., 2017), hold enormous promise for image processing in organismal biology (Christin et al., 2019; Muñoz & Price, 2019). Deep learning is a subset of machine learning using multilayered neural networks (Christin et al., 2019)—models inspired by biological nervous and visual systems (Cadieu et al., 2014; Felleman & Van Essen, 1991; Hubel & Wiesel, 1962; LeCun et al., 2015; Olden et al., 2008). Neural networks 'learn' via an iterative process of training and updating internal model parameters (weights) as a function of the magnitude of error between the expected output and the model's output. The overarching goal of training a neural network is to iteratively minimize the error between model output and expected output by optimally adjusting model weights and reaching model convergence, such that the trained neural network generalizes well to novel input data. Model weights are adjusted to minimize error on each subsequent run using an algorithm called stochastic gradient descent with backpropagation (LeCun et al., 1989; Rumelhart et al., 1995). This approach optimizes the magnitude of change for model weights between each iteration to prevent against rapid model divergence or delayed model convergence.

Convolutional neural networks (CNNs) are one class of neural networks generalizable to problems in computer vision (see Christin et al., 2019; LeCun et al., 2015 for a review). Deep learning with CNNs has become the dominant approach for mostly any computer vision task requiring object detection/recognition (LeCun et al., 2015). CNNs work by creating feature maps across multiple layers, with abstraction exponentially increasing across each subsequent layer in the network; the changes in abstraction reveal the unfolding of meaningful knowledge from input image data—which ultimately are transformed into high-level image information in fully connected layers—and produce classification labels as output (Kozma et al., 2018).

Convolutional neural networks yield impressive performance for tasks relying on biological vision and perceptual processing, such as image recognition and classification (Krizhevsky et al., 2017) and have revealed promising utility for automating data collection approaches for studies in ecology and evolutionary biology at substantially greater speeds than manual approaches (Christin et al., 2019; Lürig et al., 2021; Norouzzadeh et al., 2018; Schneider et al., 2019). To illustrate, deep learning with CNNs has been successful for ecological applications, including the automated identification of sea turtles (Gray et al., 2018) and segmentation of cetaceans (Gray et al., 2019) with field imaging from drones, in addition to acquiring location and behavioural data from camera-traps (Schneider et al., 2018). CNNs have also revealed performance similar to that of humans on visual image recognition tasks (He et al., 2015), such as automating disease screening and detection on chest radiographs (Lakhani & Sundaram, 2017) and traffic-sign detection for autonomous vehicles (Zhu et al., 2016). Despite the impressive performance and flexibility afforded by neural networks, CNNs are still prone to classification errors in scenarios that would be trivial for a human classifier and thus do not principally outperform human vision on image recognition tasks (Firestone, 2020; He et al., 2015).

Pre-trained CNNs provide enormous potential for users to implement deep learning applications out-of-the-box without additional training. The utility of pre-trained CNNs out-of-the-box is constrained by how relevant the novel input data are to the data the CNN was originally trained on. For instance, ImageNet comprises more than 14 million high-resolution images across nearly 22,000 categories and is often used as a starting point for recognition tasks with deep learning (Deng et al., 2009; Krizhevsky et al., 2017). Using a CNN pre-trained on ImageNet, interested users could implement a machine vision task without needing to train an entire neural network from the ground up. Although, model performance may be weak if the contents of input images are distantly related or completely underrepresented amongst training images contained within ImageNet. Out-of-the-box, the utility of popular approaches like ImageNet are limited for applications in ecology and evolutionary biology. Nearly 50% of categories in ImageNet are man-made objects, and whereas animals represent approximately 40% of ImageNet categories (Baker et al., 2018), the biodiversity represented is naturally imbalanced with respect to species abundance, availability of photographs, and a lack of representative contexts for rare occurrences (Beery et al., 2020; Chao, 1989; Krizhevsky et al., 2017; Schneider et al., 2020; Van Horn et al., 2018). Given that the hierarchical categorization of ImageNet is not systematically aligned to any sort of phylogenetic classification scheme, CNNs trained on ImageNet should not be expected to perform well in recognizing much of the biodiversity across the tree of life directly out-of-the-box without additional training.

High-fidelity image segmentation is a unique problem in computer vision and is further constrained by the taxonomic imbalance of biodiversity in readily accessible, annotated datasets for training neural networks to perform segmentation. For image segmentation to work, the neural network must not only successfully recognize the presence or absence of a target in an image but also must isolate the specific pixels encapsulating that target (He et al., 2017). Region Based Convolutional Neural Networks (R-CNNs) assist in accomplishing this task by scanning across an image and extracting regions of interest to predict bounding boxes delineating the object from the background (Girshick et al., 2014). Mask R-CNN is one popular meta-algorithm, which extends the R-CNN to predict high-resolution segmentation masks (i.e. pixel boundaries indicating where the identified target object meets the background pixels), and can perform instance-level segmentation, or the ability to identify separate occurrences of a target object within an image (Abdulla, 2017; He et al., 2017). To accomplish this type of high-fidelity image segmentation, the R-CNN must be trained on carefully constructed segmentation masks, which are cartesian coordinates that form a polygonal contour mask around the region of interest across a training dataset. This makes using the widely implemented ImageNet database less effective for high-resolution image segmentation with Mask R-CNN, given that the annotations provided in ImageNet are strictly four-coordinate bounding box annotations, which reflect the approximate location of a target object within an image. Rather, the Microsoft Common Objects in Context (COCO) dataset provides detailed segmentation mask annotations for common objects in their natural contexts (Lin et al., 2014). Although, COCO out-of-the-box is still limited for segmentation tasks of unrepresented biodiversity given that COCO is highly anthropocentrically biased towards domesticated species and other terrestrial tetrapods (e.g. bird, cat, dog, cow, horse, mouse, sheep, zebra, giraffe, elephant, bear). The current lack of readily available image datasets with high-resolution image segmentation mask annotations limits how biologists can accessibly engage with deep learning applications to rapidly process images for downstream image analysis workflows without additional training. On the other hand, COCO can be used as a backbone for training new image segmentation models of underrepresented organisms provided its prior experience being trained to segment terrestrial tetrapods. This procedure is called transfer learning (Razavian et al., 2014), which is the process of using an already trained neural network as a starting point for training a model to be used on a new and/or unrelated task (Yosinski et al., 2014). With transfer learning, the R-CNN can use its learned features from prior training on COCO's nearly 328,000 image dataset and efficiently generalize to segment novel categories of stimuli (e.g. fish) with only a few hundred training examples. Without transfer learning, a significantly larger training dataset of many thousands of examples, at minimum, may be necessary as neural networks typically overfit on small training datasets (i.e. those comprised of only a few hundred examples) when no prior task experience is supplied (Gray et al., 2019). We use transfer learning with pre-trained COCO weights to successfully automate image segmentation of fishes currently unrepresented within COCO out-of-the-box.

Here, we present *Sashimi*, a user-friendly toolkit that facilitates the rapid execution of accurate, high-throughput image segmentation of digitized organisms—requiring no extensive programming nor deep learning implementation experience to use. Our software implements Matterport's (Abdulla, 2017) Mask R-CNN (He et al., 2017) implementation with a custom fish segmentation model trained using transfer learning (Razavian et al., 2014) against the 328,000 image COCO dataset (Lin et al., 2014). We focus on fish because they present a wide gamut of phenotypes and color and pattern diversity (Alfaro et al., 2019; Losey et al., 2003; Marshall, 2000; Marshall et al., 2003a, 2003b; Salis et al., 2018, 2019) and because machine learning approaches have recently been applied to the problem of identifying and measuring fishes (e.g. Baloch et al., 2017; Garcia et al., 2020; Qin et al., 2016; Yao et al., 2013; Yu et al., 2020). Our toolkit provides five key contributions: (a) 'plug-and-play' reproducible, automated segmentation of fish images in complex backgrounds directly out-of-the-box, (b) a central database for sharing and distributing custom-trained segmentation models of other underrepresented organisms to use within the toolkit, (c) the ability to quickly specify an organism of interest to segment from the animal classes already included in COCO without needing to modify the codebase, (d) additional image pre-processing tools for popular color pattern analysis workflows, such as *colordistance* (Weller & Westneat, 2019), *pavo* (Maia et al., 2013, 2019) or *patternize* (Van Belleghem et al., 2018) and (e) built-in qualitative and quantitative image segmentation accuracy and diagnostic tools directly compatible with the segmentation outputs from *Sashimi*. We assessed our approach by qualitatively and quantitatively comparing automatically and manually segmented fish images across a range of image backdrop complexity. We then discuss the strengths and limitations of our approach for processing fish images and consider how this approach can be extended to other branches of the tree of life.

2 | MATERIALS AND METHODS

The *Sashimi* toolkit is freely available via GitHub (https://github.com/ShawnTylerSchwartz/sashimi).

2.1 | Mask R-CNN architecture

Our software implements the Mask R-CNN architecture (Abdulla, 2017; He et al., 2017), an extension of the Faster R-CNN (Ren et al., 2017) algorithm for generating regions of interest. Mask R-CNN not only detects a target object in an image but also rapidly detects the pixel-level target region of interest, outputting a high-resolution segmentation contour reflecting the specific boundaries of the location of the target object within the image.

2.2 | Model training dataset acquisition

Our dataset comprises 910 images, sampled across seven phenotypically disparate reef fish families, randomly divided into training and validation sets (n_{train} = 720, $n_{\text{validation}}$ = 190; approximately 80% train, 20% validation). We acquired standardized digitized specimens from J.E. Randall's fish images (N = 747; n_{train} = 598, $n_{\text{validation}}$ = 149) distributed through the Bishop Museum (http://pbs.bishopmuseum.org/images/JER/) and more naturalistic images with noisy backgrounds (N = 163; n_{train} = 122, $n_{\text{validation}}$ = 41) from iNaturalist (https://www.inaturalist.org/). Examples of the types of images included in model training are shown in Figure 1.

2.3 | Model training procedure

We first used the VGG Image Annotator Version 1.0.6 (https://www.robots.ox.ac.uk/~vgg/software/via/via-1.0.6.html; Dutta et al., 2016) to manually annotate pixel coordinates to create precise polygonal mask contours directly around the fish body boundary (i.e. where the foreground pixels of the target fish body meet those of the background). We intentionally assigned all segmentation masks for each image a class label name corresponding to the general biological name of the organism (e.g. 'fish'). Given that our intention is to build broad, organism-specific models one-by-one, we suggest building organism-specific training sets where all segmentation contours across images are labelled the same name (i.e. 'whale'). We then used these coordinates to train a model using transfer learning (Razavian et al., 2014) with the COCO pre-trained weights (Lin et al., 2014), a ResNet-101 (a CNN with 101 layers; He et al., 2016) and a Feature Pyramid Network (a generic feature extractor for detecting objects across scales; Lin et al., 2017) backbone. Despite COCO not containing any images nor segmentation mask annotations of marine organisms, we opted to use the pre-trained COCO model weights to help make our custom fish segmentation model generalizable for broader recognition and segmentation of a phenotypically diverse gamut of fish images—similar to the Gray et al. (2019) implementation of Mask R-CNN for automating cetacean species identification and length estimation.

We based training on Matterport's open-source implementation of Mask R-CNN (Abdulla, 2017) using a desktop computer equipped with a GeForce RTX 2080 GPU. We trained our model for 160 epochs over three stages. Stage 1 (epochs 1–40) trained the network heads, stage 2 (epochs 41–120) fine-tuned ResNet-101 layers stage 4 and up, and stage 3 (epochs 121–160) fine-tuned all layers. Training stages 1 and 2 used a learning rate of 0.001, whereas stage 3 used a learning rate of 0.0001. All training stages had a weight decay of 0.0001, learning momentum of 0.9, and used image augmentation by flipping 50% of the images in the left-right orientation to increase the robustness of the neural network. Model training took approximately 8 hr to complete.

2.4 | Automated segmentation pipeline

The *Sashimi* command line interface allows users to automatically extract and segment target images in common image formats. *Sashimi* supports the extraction of multiple targets from a single image; however, the analysis pipeline described here focused on images of single specimens in lateral view, a common use case for color pattern analysis. Within *Sashimi*, users can specify the path to their image folder for batch processing, save images with a transparent background, assess segmentation accuracy and train new organism-specific segmentation models. The full instructions and options are provided on the GitHub repository.

2.5 | *Sashimi* online model repository

We constructed a website to serve as a repository for the fish segmentation model (presented here) and future, community generated organismal segmentation models (https://sashimi.shawntylerschwartz.com). We aim to inspire other biologists interested in automated segmentation to create pre-trained models for their organism(s) of interest and share them to the *Sashimi* online database for the rest of the community to use and build upon. All models will be open-source and available to download, and users can submit requests to share new models, which will be evaluated before becoming publicly available.

2.6 | Evaluating fish segmentation model efficacy

2.6.1 | Qualitative image segmentation evaluation

We qualitatively assessed the performance of the current fish segmentation model by visually inspecting segmented outputs and reporting the visible strong and weak characteristics of these outputs.

FIGURE 1 Examples of wide visual phenotypic and morphological diversity, as well as variation in image background complexity, which constitute a subset of the fish images used to train the fish segmentation model presented here in the current study

2.6.2 | Quantitative image segmentation evaluation metrics

We evaluated the performance of our fish image segmentation model using four common metrics for assessing semantic segmentation accuracy: pixel accuracy, Equation 1; mean accuracy, Equation 2; mean intersection over union (IoU), Equation 3; and frequency-weighted IoU, Equation 4 (Long et al., 2015). The IoU approach is commonly used for instance segmentation tasks, with values >50% generally indicative of good detection (Gray et al., 2019; He et al., 2017). Here, we let n_{ij} be the number of pixels of class i predicted to belong to class j, $t_i = \sum_j n_{ij}$ be the total number of pixels of class i, and n_{cl} be the number of different classes. We computed each metric using the reference ('ground truth') segmentation contours (images we manually annotated with high precision) and the predicted segmentation masks from our custom-trained model (Figure 2) on our randomly selected validation image dataset, which included 41 images of fish in naturalistic and noisy backgrounds from iNaturalist and 149 standardized fish images from J.E. Randall's collection. We also report segmentation metric results for a test dataset of 60 novel images in the online Supporting Information. Additionally, we compared the results of a color pattern analysis workflow from an earlier study (Alfaro

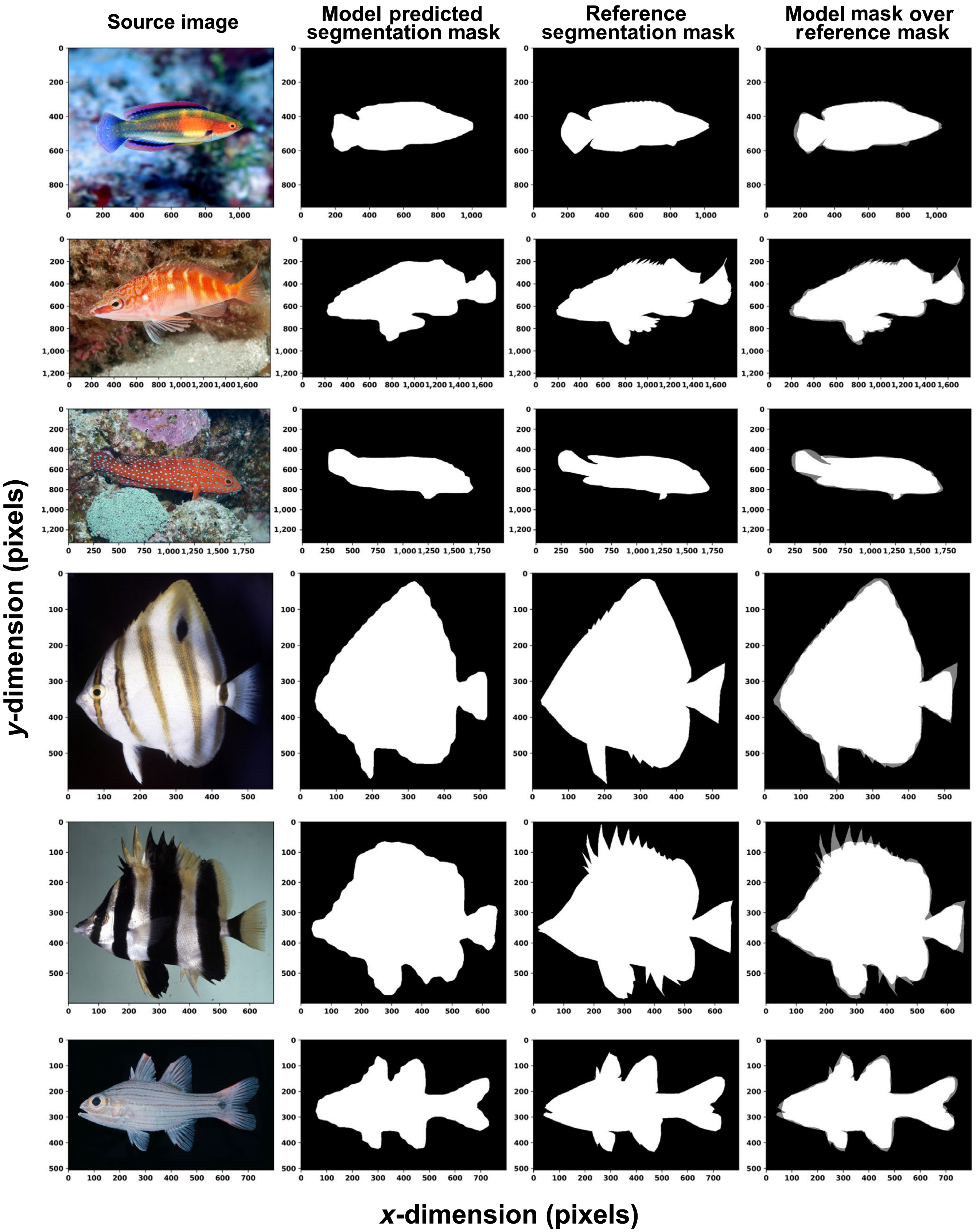

FIGURE 2 Examples of the predicted segmentation masks from our custom-trained fish segmentation model, reference segmentation contours (expert, high-precision manual segmentation mask annotations) and the superimposed image of the predicted and ground truth reference masks (dark grey regions indicate regions of the reference fish mask that were not captured by the model) for each respective source image

et al., 2019) using manual and automatically segmented images (see Supporting Information for Background, Methods, and Results).

$$\frac{\sum_i n_{ii}}{\sum_i t_i}, \tag{1}$$

$$\left(\frac{1}{n_{cl}}\right) \sum_i \frac{n_{ii}}{t_i}, \tag{2}$$

$$\left(\frac{1}{n_{cl}}\right) \sum_i \frac{n_{ii}}{(t_i + \sum_j n_{ji} - n_{ii})}, \tag{3}$$

$$\left(\sum_k t_k\right)^{-1} \sum_i \frac{t_i n_{ii}}{(t_i + \sum_j n_{ji} - n_{ii})}. \tag{4}$$

2.7 | Statistics

All statistics were performed using *JASP* (version 0.11.1; Love et al., 2019). We ran a 2 (Source: *iNaturalist*, *Randall*) × 4 (Metric: *pixel accuracy*, *mean accuracy*, *mean IoU*, *frequency-weighted IoU*) repeated-measures analysis of variance (ANOVA) to test for differences in image segmentation accuracy between manually annotated (reference) and Mask R-CNN generated segmentation mask contours for validation images from iNaturalist (complex backgrounds) and J.E. Randall's collection (relatively uniform backgrounds).

3 | RESULTS

3.1 | Qualitative image segmentation evaluation

Fishes segmented with our custom Mask R-CNN fish segmentation model were generally similar to manually segmented fishes. In most cases, we observed high performance for images automatically segmented by deep learning, regardless of whether individuals of a particular genera were included in the training dataset (Figure 3), suggesting that the model generalizes well to novel fish species. We found a small number of cases where the model performed poorly, particularly when presented with elongate parts of the body. For

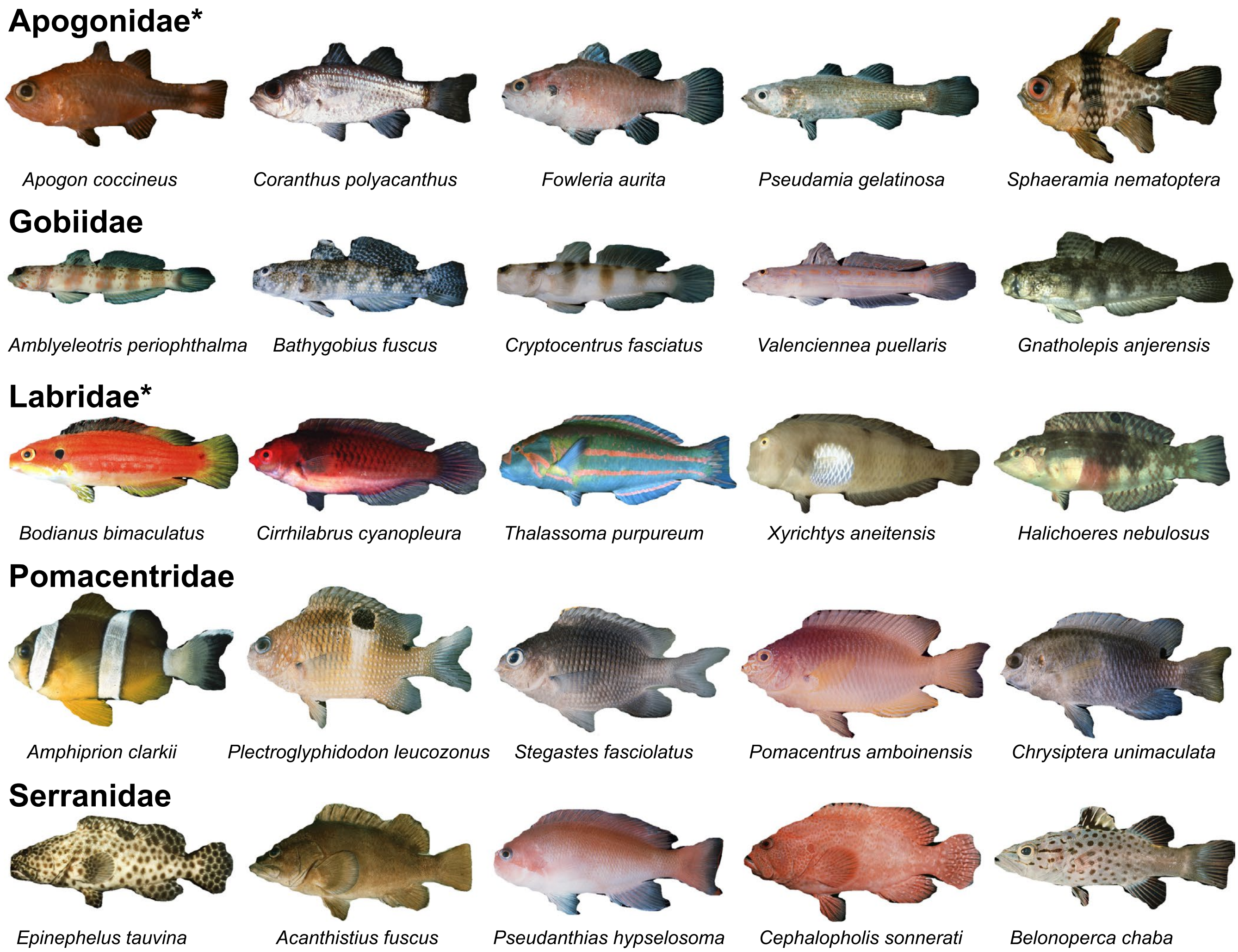

FIGURE 3 Examples of fish images that were automatically segmented using *Sashimi*. Segmented organisms presented here are from the following families: Apogonidae, Gobiidae, Labridae, Pomacentridae and Serranidae. The model was trained on taxa sampled from families labelled with an asterisk (*)

example, long dorsal spines, anal flags and long rostra were sometimes clipped (Figures 4 and 5a). We also observed wavy patterns for pixels along the boundaries of some fishes and small patches of stray background pixels (Figures 5b and 6).

3.2 | Quantitative image segmentation evaluation metrics

All 190 validation images had mean IoU scores >50% (M = 93.8%, SD = 1.4%, minimum = 87.5%, maximum = 96.9%), indicating excellent model-predicted segmentation masks compared to manually drawn reference masks. Comparing across image segmentation metrics and image sources, we found a significant main effect of evaluation metric on accuracy, $F(1.79, 336.75) = 739.53$, $p_{adj.} < 0.001$, suggesting that independent of image source (iNaturalist, J.E. Randall), accuracy metrics varied significantly from one another (Figure 7). We also found a significant main effect of image source, $F(1, 188) = 60.70$, $p < 0.001$, such that regardless of accuracy metric, images from iNaturalist were generally segmented with higher accuracy (M = 96.5%, SD = 1.1%) than were J.E. Randall's images (M = 95.2%, SD = 0.8%), $t(188) = 7.79$, Cohen's d = 0.57, $p_{adj.} < 0.001$. Lastly, we uncovered a

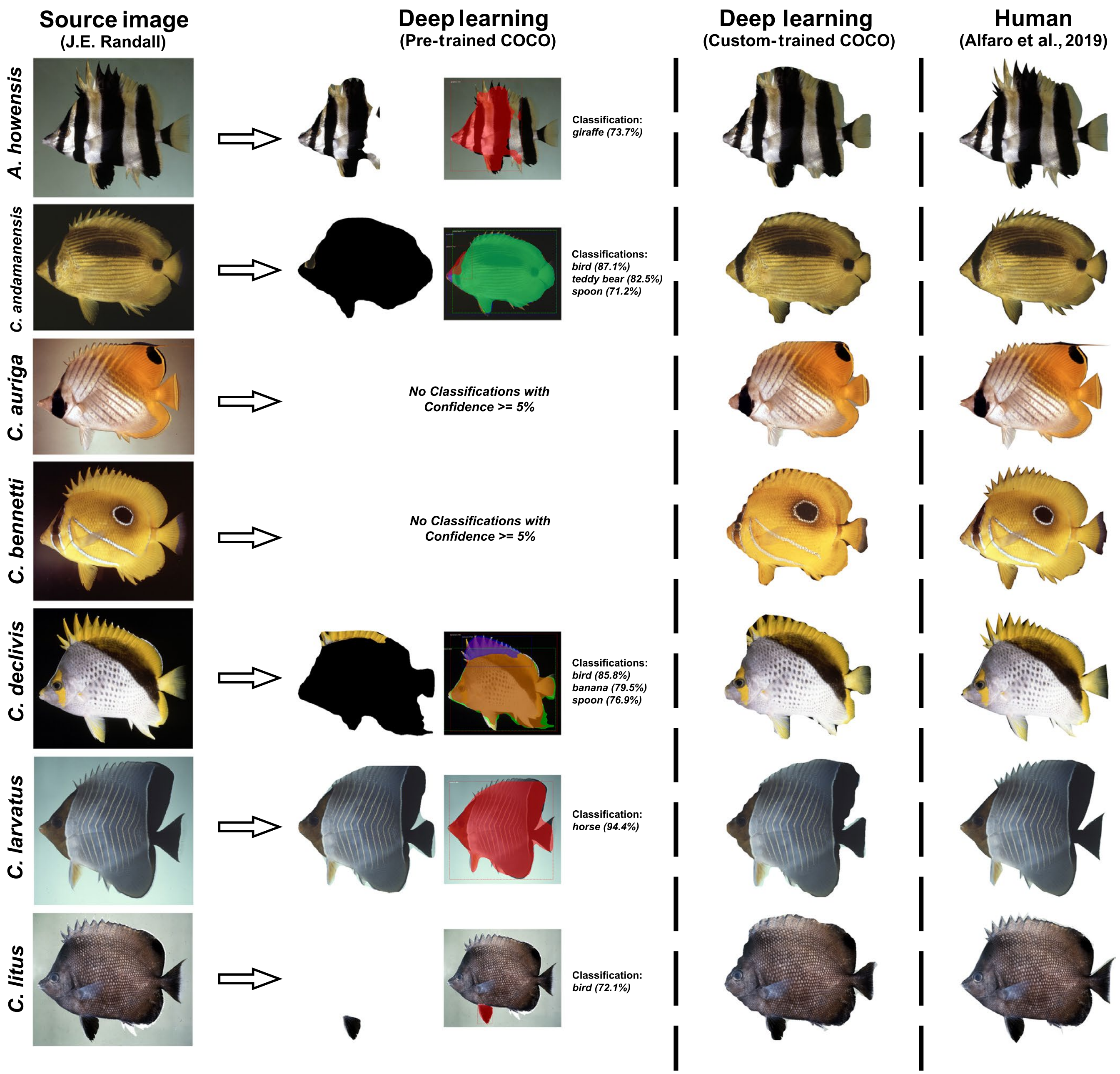

FIGURE 4 Butterflyfishes ('Source image') that were digitized by manual segmentation ('Human') from Alfaro et al. (2019) and by deep learning (with both the pre-trained COCO weights only and our custom-trained model built upon the pre-trained COCO weights). Prediction confidence for classification labels from the pre-trained COCO weights only model is presented in parentheses following the classification labels (which were only displayed when model prediction confidence was ≥5%)

significant interaction between accuracy metric and image source, $F(1.79, 336.75) = 118.24$, $p_{adj.} < 0.001$. Bonferroni-corrected post hoc paired-sample *t*-tests for the significant main effect of metric revealed frequency-weighted IoU to be significantly less than pixel and mean accuracy, but significantly higher than mean IoU. Additionally, pixel accuracy was significantly higher than both mean accuracy and mean IoU, and mean accuracy was significantly higher than mean IoU (Table 1; post hoc comparisons for the significant interaction are presented in Supporting Information Table S1). Segmentation metrics for the novel test dataset revealed the same overall pattern of results reported here for the validation dataset (see Supporting Information Tables S2 and S3). *Sashimi* and manually segmented images also yielded statistically similar inferences of color pattern (Supporting Information Tables S4 and S5).

4 | DISCUSSION

Our custom-trained model exhibited strong performance in segmenting fish images in standardized and natural settings (Figures 4 and 7). The high segmentation accuracy we obtained results from the Mask R-CNN meta-algorithm's ability to successfully adapt to most computer vision tasks (He et al., 2017). In general, segmentation of standardized images compares favourably to manual segmentation, preserving gross and fine morphological features necessary for many kinds of morphometric analyses. Some body shapes did challenge the model, possibly rendering these automatically segmented images as unsuitable for measurement of the most elongate fin spines or of species with extreme rostral elongation, such as *Forcipiger* butterflyfishes (Figure 6a). We suspect that additional training datasets comprised of fishes with extreme morphologies would improve fidelity of edge contour predictions. Although, images segmented under the current model yield similar results to a manual workflow for color pattern analysis (Supporting Information) and are likely to have sufficient fidelity for a wide range of applications in ecology and evolution. These results highlight the potential of R-CNNs for biological applications (He et al., 2017) and extend the range of this approach for identifying fishes across lab and field conditions (Garcia et al., 2020; Qin et al., 2016; Salman et al., 2016; Yu et al., 2020).

Despite the potential of deep learning, existing barriers to implementing these tools are substantial for non-specialists. A review

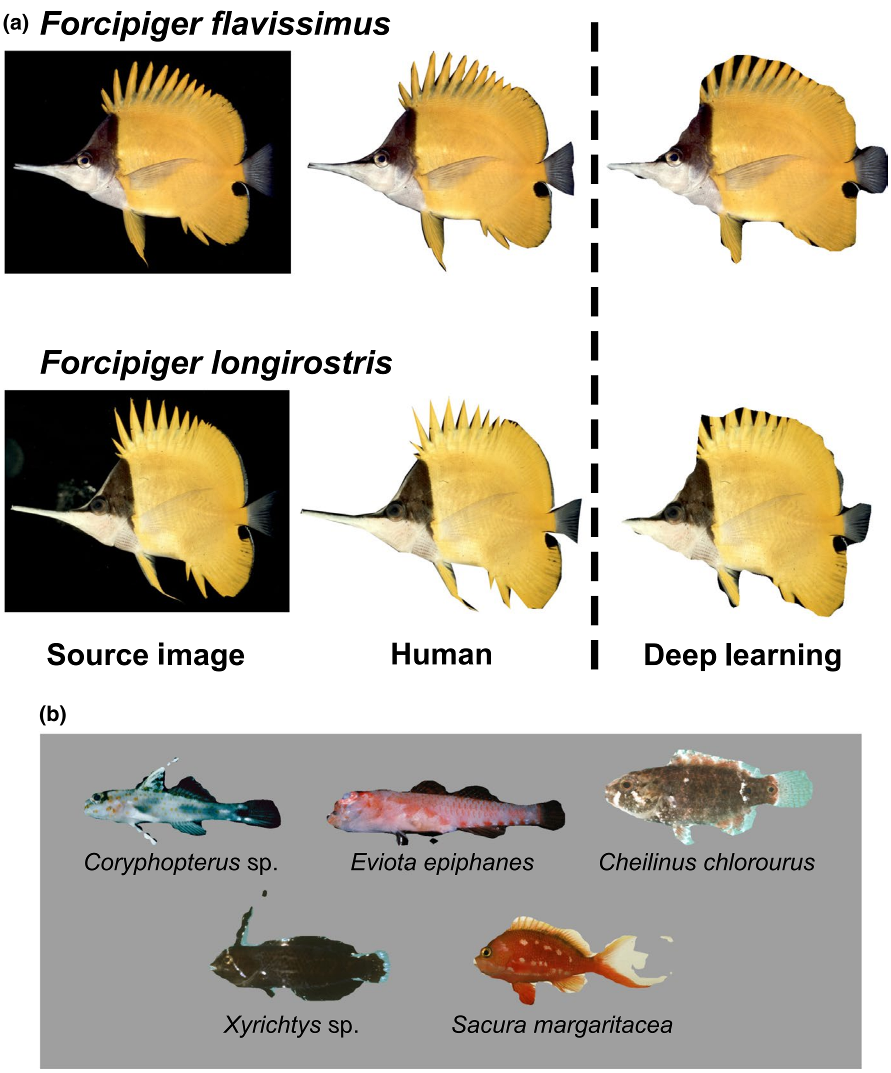

FIGURE 5 (a) Butterflyfishes of the genus *Forcipiger*, which are characterized by distinct dorsal spines and long snouts (rostrums), that were digitized either by manual segmentation or by our custom-trained fish segmentation model. (b) Examples of automatically segmented fish images with small patches of stray background pixels that were not completely captured by our current model

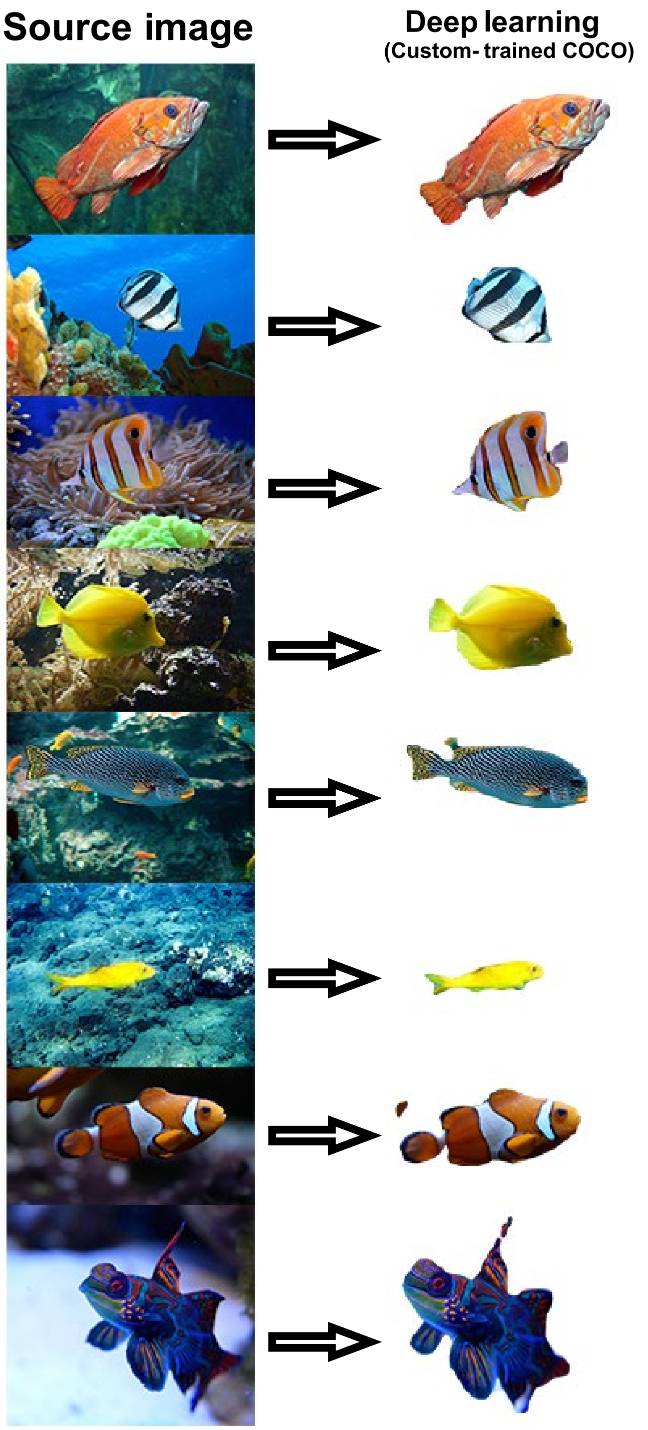

FIGURE 6 Image segmentation of novel fish images in their natural contexts with noisy backdrops using our custom-trained fish segmentation model

reflecting the lack of accessible tools for non-programmers recommends that ecologists and evolutionary biologists should consult with computer scientists before adopting these approaches, highlighting the lack of user-friendly tools for deep learning (Christin et al., 2019). The lack of user-friendly tools for bridging the 'last mile' connecting the enormous power of R-CNNs to biologists working across the diversity of the tree of life likely explains the limited use of deep learning algorithms within ecology and evolution. We believe that *Sashimi* can help bridge this gap and allow non-specialists to develop powerful pipelines for image analysis.

Some limitations may arise when attempting to utilize Mask R-CNN for a novel group of organisms. Specifically, one might observe poor model performance for novel input images visually deviating from those primarily comprising the original training dataset. Generalizability in model performance will ultimately depend on the variation of the examples supplied during model training. To remedy this problem, ecologists and evolutionary biologists should carefully select both common and rarer examples of digitized organismal images reflecting a diverse set of appearances, backdrops, and contexts. Gathering images representing high phenotypic and contextual visual diversity should help enhance model generalizability and performance in most cases. Over and above training dataset construction is considering the iterative nature of model training required to achieve performance suitable for one's specific needs. For instance, if an ecologist aims to segment the bodies of organisms for a color pattern analysis, images with small visual artifacts along the boundaries of the body should not expectedly impact downstream analytical goals. However, a morphometric analysis aiming to measure landmarks on regions at the edge of the body may require more fine-grained model tuning such that predicted segmentation masks more carefully extract the foreground pixels from the background pixels at the boundaries of the target. Such model training may require hundreds or even thousands of relevant example images and may possibly require additional generations of training. Users should also consider the quality of their supplied mask annotations for training dataset images. Care should be taken during manual annotation to ensure coordinates reflect a smooth boundary delineating the background pixels from the foreground pixels, rather than a more jagged, rough approximation of the target's location within the image. In sum, users interested in refining the model for different use cases should anticipate iteratively training models with different sized training datasets and parameters until suitable performance is achieved.

Overall, *Sashimi* provides an extensible toolkit for automating and evaluating image segmentation performance using the powerful deep learning meta-algorithm, Mask R-CNN (He et al., 2017). As studies in ecology and evolutionary biology continue to move towards analyses of phenotype at massive phylogenetic scales (e.g. Baliga & Mehta, 2019; Chang & Alfaro, 2016; Price et al., 2019; Rabosky et al., 2018), having a toolkit which aims to simplify and streamline the image segmentation procedure from start-to-finish will help to eliminate the bottleneck between the rapid acquisition and slow extraction of meaningful data by (a) facilitating high-throughput image segmentation, (b) easing the foregoing technical barriers potentially prohibiting biologists from taking advantage of the power of deep learning image pre-processing for their studies, (c) making model diagnostic metrics and visualization accessible with the 'click-of-a-button' and (d) promoting open-access sharing of clade-specific models to facilitate reproducible and efficient

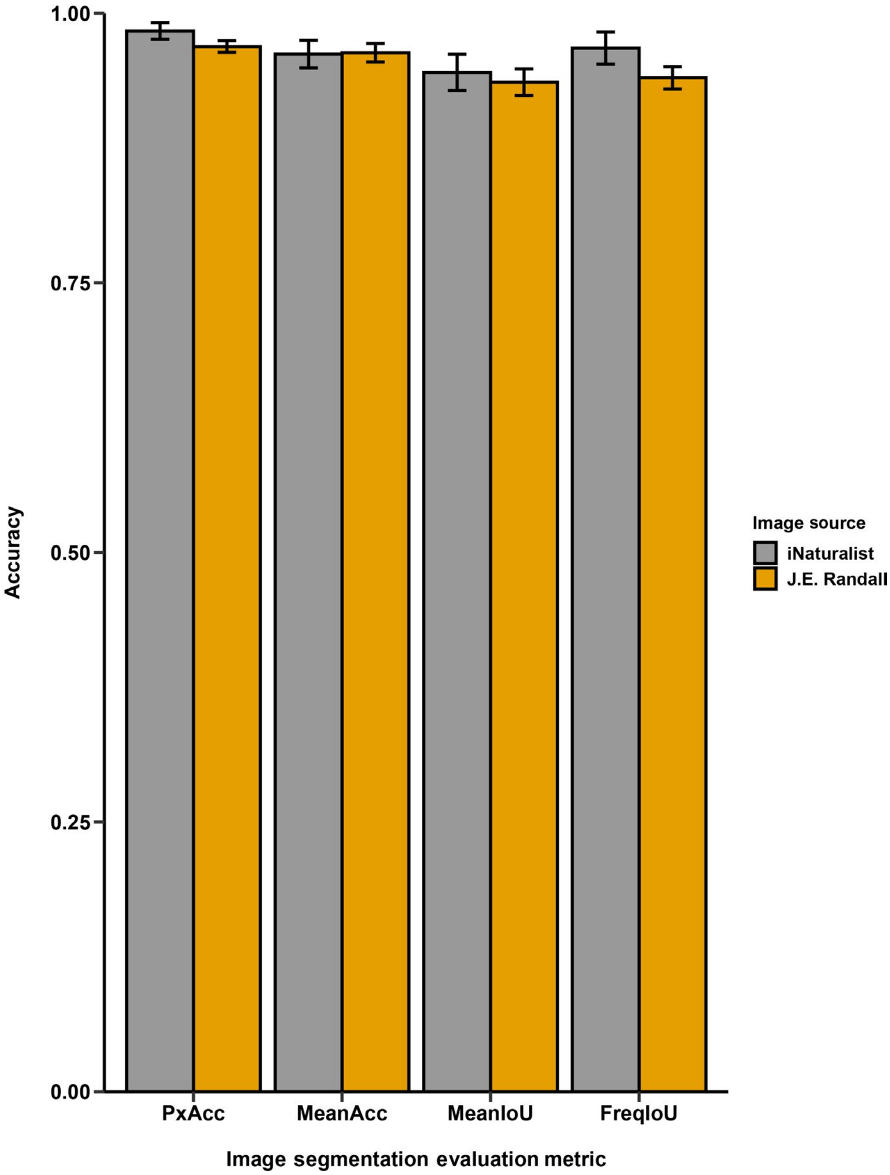

FIGURE 7 Bar plot of four common image segmentation evaluation metrics: pixel accuracy (PxAcc, Equation 1), mean accuracy (MeanAcc, Equation 2), mean intersection over union (MeanIoU, Equation 3) and frequency-weighted intersection over union (FreqIoU, Equation 4) by image source (iNaturalist, J.E. Randall). Error bars represent ±1 *SE* of the mean

TABLE 1 Post hoc comparisons for the main effect of evaluation metric

Comparison	t(188)	$p_{adj.}$	Cohen's d
FreqIoU < MeanAcc	−14.39	<0.001	−1.04
FreqIoU > MeanIoU	9.03	<0.001	0.66
FreqIoU < PxAcc	−46.87	<0.001	−3.40
MeanAcc > MeanIoU	43.92	<0.001	3.19
MeanAcc < PxAcc	−11.49	<0.001	−0.83
MeanIoU < PxAcc	−46.04	<0.001	−3.34

Note: Student's *t* test. Bonferroni-corrected (for multiple comparisons) post hoc paired-sample *t*-tests comparing the four image segmentation validation metrics; pixel accuracy (PxAcc, Equation 1), mean accuracy (MeanAcc, Equation 2), mean intersection over union (MeanIoU, Equation 3) and frequency-weighted intersection over union (FreqIoU, Equation 4). Cohen's *d* does not correct for multiple comparisons

organismal image pre-processing workflows in a future of big data research in integrative biology (Muñoz & Price, 2019).

ACKNOWLEDGEMENTS

The authors would like to thank W. Tsai, M. Juhn, E. Karan, R. Franco and J. Chang for helpful comments on the manuscript, A. Schwartz, M. Perillo, M. Chari and A. Butler for their assistance in generating reference segmentation mask contours used to train our fish segmentation model, and H. Weller and T. McCraney for their advice, guidance and support throughout this project.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

S.T.S. and M.E.A. conceived of the study; S.T.S. wrote the software. Both authors contributed to the writing of the manuscript.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13712.

DATA AVAILABILITY STATEMENT

The *Sashimi* toolkit is freely accessible from GitHub (https://github.com/ShawnTylerSchwartz/sashimi). The source code is also archived on Zenodo at https://doi.org/10.5281/zenodo.5353751 (Schwartz & Alfaro, 2021a). Data are archived on Dryad at https://doi.org/10.5068/D16M4N (Schwartz & Alfaro, 2021b). The *Sashimi* website provides an open-access repository containing the fish segmentation model used in the current study and will house future models as they become available (https://sashimi.shawntylerschwartz.com).

ORCID

Shawn T. Schwartz https://orcid.org/0000-0001-6444-8451
Michael E. Alfaro https://orcid.org/0000-0002-8898-8230

REFERENCES

Abdulla, W. (2017). Mask r-cnn for object detection and instance segmentation on keras and tensorflow.

Alfaro, M. E., Karan, E. A., Schwartz, S. T., & Shultz, A. J. (2019). The evolution of color pattern in butterflyfishes (Chaetodontidae). *Integrative and Comparative Biology*, *59*, 604–615. https://doi.org/10.1093/icb/icz119

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLOS Computational Biology*, *14*, e1006613. https://doi.org/10.1371/journal.pcbi.1006613

Baliga, V. B., & Mehta, R. S. (2019). Morphology, ecology, and biogeography of independent origins of cleaning behavior around the world. *Integrative and Comparative Biology*, *59*, 625–637. https://doi.org/10.1093/icb/icz030

Baloch, A., Ali, M., Gul, F., Basir, S., & Afzal, I. (2017). Fish Image Segmentation Algorithm (FISA) for improving the performance of image retrieval system. *International Journal of Advanced Computer Science and Applications*, *8*. https://doi.org/10.14569/IJACSA.2017.081252

Beery, S., Liu, Y., Morris, D., Piavis, J., Kapoor, A., Meister, M., Joshi, N., & Perona, P. (2020). Synthetic examples improve generalization for rare classes. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 852–862).

Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., & DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLOS Computational Biology*, *10*, e1003963. https://doi.org/10.1371/journal.pcbi.1003963

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., & Joly, A. (2017). Going deeper in the automated identification of Herbarium specimens. *BMC Evolutionary Biology*, *17*. https://doi.org/10.1186/s12862-017-1014-z

Chang, J., & Alfaro, M. E. (2016). Crowdsourced geometric morphometrics enable rapid large-scale collection and analysis of phenotypic data. *Methods in Ecology and Evolution*, *7*, 472–482. https://doi.org/10.1111/2041-210X.12508

Chao, A. (1989). Estimating population size for sparse data in capture-recapture experiments. *Biometrics*, 427–438. https://doi.org/10.2307/2531487

Cheng, X., Zhang, Y., Chen, Y., Wu, Y., & Yue, Y. (2017). Pest identification via deep residual learning in complex background. *Computers and Electronics in Agriculture*, *141*, 351–356. https://doi.org/10.1016/j.compag.2017.08.005

Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, *10*, 1632–1644. https://doi.org/10.1111/2041-210X.13256

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). IEEE.

Dutta, A., Gupta, A., & Zissermann, A. (2016). *VGG image annotator (VIA)*. Retrieved from http://www.robots.ox.ac.uk/~vgg/software/via

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, NY: 1991)*, *1*(1), 1–47.

Firestone, C. (2020). Performance vs. competence in human–machine comparisons. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 26562–26571. https://doi.org/10.1073/pnas.1905334117

Garcia, R., Prados, R., Quintana, J., Tempelaar, A., Gracias, N., Rosen, S., Vågstøl, H., & Løvall, K. (2020). Automatic segmentation of fish using deep learning with application to fish size measurement. *ICES Journal of Marine Science*, *77*, 1354–1366. https://doi.org/10.1093/icesjms/fsz186

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580–587).

Gomez Villa, A., Salazar, A., & Vargas, F. (2017). Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics*, *41*, 24–32. https://doi.org/10.1016/j.ecoinf.2017.07.004

Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing*. Prentice Hall.

Gray, P. C., Bierlich, K. C., Mantell, S. A., Friedlaender, A. S., Goldbogen, J. A., & Johnston, D. W. (2019). Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. *Methods in Ecology and Evolution*, *10*, 1490–1500. https://doi.org/10.1111/2041-210X.13246

Gray, P. C., Fleishman, A. B., Klein, D. J., Mckown, M. W., Bézy, V. S., Lohmann, K. J., & Johnston, D. W. (2018). A convolutional neural network for detecting sea turtles in drone imagery. *Methods in Ecology and Evolution*, *10*(3), 345–355. https://doi.org/10.1111/2041-210X.13132

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961–2969).

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*, 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Joly, A., Bonnet, P., Goëau, H., Barbe, J., Selmi, S., Champ, J., Dufour-Kowalski, S., Affouard, A., Carré, J., Molino, J.-F., Boujemaa, N., & Barthélémy, D. (2016). A look inside the Pl@ntNet experience. *Multimedia Systems*, *22*, 751–766. https://doi.org/10.1007/s00530-015-0462-9

Kozma, R., Ilin, R., & Siegelmann, H. T. (2018). Evolution of abstraction across layers in deep learning neural networks. *Procedia Computer Science*, *144*, 203–213. https://doi.org/10.1016/j.procs.2018.10.520

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*, 84–90. https://doi.org/10.1145/3065386

Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, *284*, 574–582. https://doi.org/10.1148/radiol.2017162326

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

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in Neural Information Processing Systems*, *2*, 396–404.

Lee, S. H., Chan, C. S., & Remagnino, P. (2018). Multi-organ plant classification based on convolutional and recurrent neural networks. *IEEE Transactions on Image Processing*, *27*, 4287–4301. https://doi.org/10.1109/TIP.2018.2836321

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117–2125).

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In *Computer Vision – ECCV 2014* (pp. 740–755). Springer International Publishing.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431–3440).

Losey, G., McFarland, W., Loew, E., Zamzow, J., Nelson, P., & Marshall, N. (2003). Visual biology of Hawaiian coral reef fishes. I. Ocular transmission and visual pigments. *Copeia*, *2003*, 433–454. https://doi.org/10.1643/01-053

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., & Epskamp, S. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*, 1–17.

Lürig, M. D., Donoughe, S., Svensson, E. I., Porto, A., & Tsuboi, M. (2021). Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. *Frontiers in Ecology and Evolution*, *9*. https://doi.org/10.3389/fevo.2021.642774

Maia, R., Eliason, C. M., Bitton, P.-P., Doucet, S. M., & Shawkey, M. D. (2013). pavo: An R package for the analysis, visualization and organization of spectral data. *Methods in Ecology and Evolution*.

Maia, R., Gruson, H., Endler, J. A., & White, T. E. (2019). pavo 2: New tools for the spectral and spatial analysis of colour in R. *Methods in Ecology and Evolution*, *10*, 1097–1107.

Marques, A. C. R., M. Raimundo, M., B. Cavalheiro, E. M., F. P. Salles, L., Lyra, C., & J. Von Zuben, F. (2018). Ant genera identification using an ensemble of convolutional neural networks. *PLoS ONE*, *13*, e0192011. https://doi.org/10.1371/journal.pone.0192011

Marshall, N. J. (2000). Communication and camouflage with the same 'bright' colours in reef fishes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *355*, 1243–1248. https://doi.org/10.1098/rstb.2000.0676

Marshall, N., Jennings, K., McFarland, W., Loew, E., & Losey, G. (2003a). Visual biology of Hawaiian coral reef fishes. II. Colors of Hawaiian coral reef fish. *Copeia*, *2003*, 455–466. https://doi.org/10.1643/01-055

Marshall, N., Jennings, K., McFarland, W., Loew, E., & Losey, G. (2003b). Visual biology of Hawaiian coral reef fishes. III. Environmental light and an integrated approach to the ecology of reef fish vision. *Copeia*, *2003*, 467–480. https://doi.org/10.1643/01-056

Muñoz, M. M., & Price, S. A. (2019). The future is bright for evolutionary morphology and biomechanics in the era of big data. *Integrative and Comparative Biology*, *59*, 599–603. https://doi.org/10.1093/icb/icz121

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United Sates of America*, *115*, E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without tears: A primer for ecologists. *The Quarterly Review of Biology*, *83*, 171–193. https://doi.org/10.1086/587826

Pennekamp, F., & Schtickzelle, N. (2013). Implementing image analysis in laboratory-based experimental systems for ecology and evolution: A hands-on guide. *Methods in Ecology and Evolution*, *4*, 483–492. https://doi.org/10.1111/2041-210X.12036

Porto, A., & Voje, K. L. (2020). ML-morph: A fast, accurate and general approach for automated detection and landmarking of biological structures in images. *Methods in Ecology and Evolution*, *11*, 500–512. https://doi.org/10.1111/2041-210X.13373

Price, S. A., Friedman, S. T., Corn, K. A., Martinez, C. M., Larouche, O., & Wainwright, P. C. (2019). Building a body shape morphospace of teleostean fishes. *Integrative and Comparative Biology*, *59*, 716–730. https://doi.org/10.1093/icb/icz115

Qin, H., Li, X., Liang, J., Peng, Y., & Zhang, C. (2016). DeepFish: Accurate underwater live fish recognition with a deep architecture. *Neurocomputing*, *187*, 49–58. https://doi.org/10.1016/j.neucom.2015.10.122

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M., & Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, *559*, 392–395. https://doi.org/10.1038/s41586-018-0273-1

Raitoharju, J., Riabchenko, E., Meissner, K., Ahmad, I., Iosifidis, A., Gabbouj, M., & Kiranyaz, S. (2016). Data enrichment in fine-grained classification of aquatic macroinvertebrates. In *2016 ICPR 2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI)* (pp. 43–48). IEEE.

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 806–813).

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

Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995). Backpropagation: The basic theory. In *Backpropagation: Theory, architectures and applications*. (pp. 1–34).

Salis, P., Lorin, T., Laudet, V., & Frédérich, B. (2019). Magic traits in magic fish: Understanding color pattern evolution using reef fish. *Trends in Genetics*, *35*, 265–278. https://doi.org/10.1016/j.tig.2019.01.006

Salis, P., Roux, N., Soulat, O., Lecchini, D., Laudet, V., & Frédérich, B. (2018). Ontogenetic and phylogenetic simplification during white stripe evolution in clownfishes. *BMC Biology*, *16*. https://doi.org/10.1186/s12915-018-0559-7

Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., & Harvey, E. (2016). Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography: Methods*, *14*, 570–585.

Schneider, S., Greenberg, S., Taylor, G. W., & Kremer, S. C. (2020). Three critical factors affecting automated image species recognition performance for camera traps. *Ecology and Evolution*, *10*, 3503–3517. https://doi.org/10.1002/ece3.6147

Schneider, S., Taylor, G. W., & Kremer, S. (2018). Deep learning object detection methods for ecological camera trap data. In *2018 15th Conference on computer and robot vision (CRV)* (pp. 321–328). IEEE.

Schneider, S., Taylor, G. W., Linquist, S., & Kremer, S. C. (2019). Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution*, *10*, 461–470. https://doi.org/10.1111/2041-210X.13133

Schwartz, S. T., & Alfaro, M. E. (2021a). Sashimi: A toolkit for facilitating high-throughput organismal image segmentation using deep learning. In *Methods in Ecology and Evolution* (v2.0.0). Zenodo. https://doi.org/10.5281/zenodo.5353752

Schwartz, S. T., & Alfaro, M. E. (2021b). Data from: *Sashimi*: A toolkit for facilitating high-throughput organismal image segmentation using deep learning. *Methods in Ecology and Evolution*. https://doi.org/10.5068/D16M4N

Van Belleghem, S. M., Papa, R., Ortiz-Zuazaga, H., Hendrickx, F., Jiggins, C. D., Owen Mcmillan, W., & Counterman, B. A. (2018). patternize: An R package for quantifying colour pattern variation. *Methods in Ecology and Evolution*, *9*, 390–398. https://doi.org/10.1111/2041-210X.12853

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., & Belongie, S. (2018). The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8769–8778).

Wäldchen, J., & Mäder, P. (2018a). Machine learning for image based species identification. *Methods in Ecology and Evolution*, *9*, 2216–2225. https://doi.org/10.1111/2041-210X.13075

Wäldchen, J., & Mäder, P. (2018b). Plant species identification using computer vision techniques: A systematic literature review. *Archives of Computational Methods in Engineering*, *25*, 507–543. https://doi.org/10.1007/s11831-016-9206-z

Wäldchen, J., Rzanny, M., Seeland, M., & Mäder, P. (2018). Automated plant species identification—Trends and future directions. *PLOS Computational Biology*, *14*, e1005993. https://doi.org/10.1371/journal.pcbi.1005993

Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, *87*, 533–545. https://doi.org/10.1111/1365-2656.12780

Weller, H. I., & Westneat, M. W. (2019). Quantitative color profiling of digital images with earth mover's distance using the R package colordistance. *PeerJ*, *7*, e6398. https://doi.org/10.7717/peerj.6398

Willis, C. G., Ellwood, E. R., Primack, R. B., Davis, C. C., Pearson, K. D., Gallinat, A. S., Yost, J. M., Nelson, G., Mazer, S. J., Rossington, N. L., Sparks, T. H., & Soltis, P. S. (2017). Old plants, new tricks: Phenological research using herbarium specimens. *Trends in Ecology & Evolution*, *32*, 531–546. https://doi.org/10.1016/j.tree.2017.03.015

Yao, H., Duan, Q., Li, D., & Wang, J. (2013). An improved K-means clustering algorithm for fish image segmentation. *Mathematical and Computer Modelling*, *58*, 790–798.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *arXiv Preprint arXiv:1411.1792*.

Yu, C., Fan, X., Hu, Z., Xia, X., Zhao, Y., Li, R., & Bai, Y. (2020). Segmentation and measurement scheme for fish morphological features based on Mask R-CNN. *Information Processing in Agriculture*, *7*, 523–534. https://doi.org/10.1016/j.inpa.2020.01.002

Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., & Hu, S. (2016). Traffic-sign detection and classification in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2110–2118).

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Schwartz, S. T., & Alfaro, M. E. (2021). *Sashimi*: A toolkit for facilitating high-throughput organismal image segmentation using deep learning. *Methods in Ecology and Evolution*, 00, 1–14. https://doi.org/10.1111/2041-210X.13712