



# Automatic image-based identification and biomass estimation of invertebrates

Johanna Ärje<sup>1,2,3</sup>  | Claus Melvad<sup>4</sup> | Mads Rosenhøj Jeppesen<sup>4</sup> |  
 Sigurd Agerskov Madsen<sup>4</sup> | Jenni Raitoharju<sup>5</sup> | Maria Strandgård Rasmussen<sup>1</sup> |  
 Alexandros Iosifidis<sup>6</sup> | Ville Tirronen<sup>7</sup> | Moncef Gabbouj<sup>2</sup> | Kristian Meissner<sup>5</sup> |  
 Toke Thomas Høye<sup>1</sup> 

<sup>1</sup>Department of Bioscience and Arctic Research Centre, Aarhus University, Ronde, Denmark; <sup>2</sup>Unit of Computing Sciences, Tampere University, Tampere, Finland; <sup>3</sup>Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland; <sup>4</sup>Aarhus School of Engineering and Arctic Research Centre, Aarhus University, Ronde, Denmark; <sup>5</sup>Finnish Environment Institute, Jyväskylä, Finland; <sup>6</sup>Department of Engineering, Aarhus University, Ronde, Denmark and <sup>7</sup>Department of Information Technology, University of Jyväskylä, Jyväskylä, Finland

## Correspondence

Johanna Ärje

Email: johanna.arje@gmail.com

## Funding information

Aarhus Universitet; Villum Fonden,  
 Grant/Award Number: 17523

**Handling Editor:** D. J. (Johan) Kotze

## Abstract

1. Understanding how biological communities respond to environmental changes is a key challenge in ecology and ecosystem management. The apparent decline of insect populations necessitates more biomonitoring but the time-consuming sorting and expert-based identification of taxa pose strong limitations on how many insect samples can be processed. In turn, this affects the scale of efforts to map and monitor invertebrate diversity altogether. Given recent advances in computer vision, we propose to enhance the standard human expert-based identification approach involving manual sorting and identification with an automatic image-based technology.
2. We describe a robot-enabled image-based identification machine, which can automate the process of invertebrate sample sorting, specimen identification and biomass estimation. We use the imaging device to generate a comprehensive image database of terrestrial arthropod species which is then used to test classification accuracy, that is, how well the species identity of a specimen can be predicted from images taken by the machine. We also test sensitivity of the classification accuracy to the camera settings (aperture and exposure time) to move forward with the best possible image quality. We use state-of-the-art Resnet-50 and InceptionV3 convolutional neural networks for the classification task.
3. The results for the initial dataset are very promising as we achieved an average classification accuracy of 0.980. While classification accuracy is high for most species, it is lower for species represented by less than 50 specimens. We found

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13428>

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

© 2020 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society

significant positive relationships between mean area of specimens derived from images and their dry weight for three species of *Diptera*.

4. The system is general and can easily be used for other groups of invertebrates as well. As such, our results pave the way for generating more data on spatial and temporal variation in invertebrate abundance, diversity and biomass.

#### KEYWORDS

biodiversity, classification, convolutional neural network, deep learning, insects, machine learning, spiders

## 1 | INTRODUCTION

The uncertainties around the state of global insect populations are largely due to data gaps and more efficient methods for quantifying abundance and identifying invertebrates are urgently needed (Seibold et al., 2019; Wagner, 2019). Commonly used passive traps, such as Malaise traps, produce samples, which are time-consuming to process. For this reason, samples are sometimes only weighed—as was the case in the study, which triggered the global attention around insect declines (Hallmann et al., 2017). In other studies, specimens are lumped into larger taxonomic groups (Høye & Forchhammer, 2008; Rich, Gough, & Boelman, 2013; Timms, Bowden, Summerville, & Buddle, 2012) or only specific taxa are identified (Hansen et al., 2016; Loboda, Savage, Buddle, Schmidt, & Høye, 2018). While such traps help standardize efforts across sampling events and are often preferred in long-term monitoring, the time and expertise needed to process (sort, identify, count and potentially weigh) samples of insects and other invertebrates from passive traps remains a key bottleneck in entomological research. In light of the apparent global decline of many invertebrate taxa and the Linnean biodiversity shortfall (i.e. only a small fraction of all species on Earth are described; Hortal et al., 2015), more efficient ways of processing invertebrate samples are in high demand. Such methods should ideally (a) not destroy specimens, which could be new to the study area or even new to science, (b) count the abundance of individual species and (c) estimate the biomass of such samples.

Reliable identification of species is pivotal but due to its inherent slowness and high costs, traditional expert identification has caused bottlenecks in the bioassessment process. As the demand for biological monitoring grows, and the number of taxonomic experts declines (Gaston & O'Neill, 2004), there is a need for alternatives to the manual processing and identification of monitoring samples (Borja & Elliott, 2013; Nygård et al., 2016). Genetic approaches are gaining popularity (Aylagas, Borja, Irigoien, & Rodríguez-Ezpeleta, 2016; Dunker et al., 2016; Elbrecht, Vamos, Meissner, Aroviita, & Leese, 2017; Kermarrec et al., 2014; Keskin, 2014; Raupach et al., 2010; Zimmermann, Glockner, Jahn, Enke, & Gemeinholzer, 2015). For cases in which only a few thousand specimens need to be identified to species, novel methods are considerably lowering the cost per barcode

(Srivathsan et al., 2018, 2019). However these approaches become too labour extensive in cases where large numbers of specimens need to be identified. Metabarcoding techniques such as, for example, Illumina paired-end sequencing of libraries generated with universal primer pairs work well but are only cheaper if the number of samples produced is sufficiently large (Aylagas, Borja, Muxika, & Rodríguez-Ezpeleta, 2018; Elbrecht et al., 2017). For small samples, metabarcoding is equally, or even more expensive as traditional taxonomy. The main caveat of metabarcoding is that no reliable abundance or biomass data can be obtained so far. Instead, machine learning methods could be used to semi-automate the task of manual species identification and specimen biomass estimation. Machine learning methods can also be used for pre-sorting samples to reduce barcoding efforts.

Several computer-based identification systems for biological monitoring have been proposed and tested in the last two decades. While Potamis (2014) classified birds based on sound and Qian, HongBin, Zhen, and XiangBo (2011) used acoustic signals to identify bark beetles, most computer-based identification systems use morphological features and image data for species prediction. Schröder, Drescher, Steinhage, and Kastenholz (1995), Weeks, Gauld, Gaston, and O'Neill (1997), Liu, Shen, Zhang, and Yang (2008), LeQuing and Zhen (2012), Perre et al. (2016) and Feng, Bhanu, and Heraty (2016) classified bees, butterflies, fruit flies and wasps based on wing features. In aquatic research, automatic or semi-automatic systems have been developed to identify algae (e.g. Santhi, Pradeepa, Subashini, & Kalaiselvi, 2013), zooplankton (e.g. Bochinski et al., 2018; Dai, Wang, Zheng, Ju, & Qiao, 2016) and benthic macroinvertebrates (e.g. Ärje et al., submitted; Raitoharju & Meissner, 2019). Recently, deep learning methods such as convolutional neural networks (CNNs) have been found to provide the best classification results for general image data (Deng et al., 2009; He, Zhang, Ren, & Sun, 2016) as well as for invertebrates (e.g. Ärje et al., submitted; Ding & Taylor, 2016; Raitoharju et al., 2019; Valan, Makonyi, Maki, Vondráček, & Ronquist, 2019; Xia, Chen, Wang, Zhang, & Xie, 2018). In recent years, iNaturalist, a citizen-science application and community for recoding and sharing nature observations, has accumulated a notable database of taxa images for training state-of-the-art CNNs (Van Horn et al., 2018). However, such field photos will not provide the same accuracy as can be achieved in the laboratory under controlled light conditions.

Classification based on single 2D images can suffer from variations in the viewing angle causing certain morphological traits to remain concealed. To overcome those limitations, Zhang, Gao, and Caelli (2010) have proposed a method for structuring 3D insect models from 2D images. Raitoharju et al. (2018) have presented an imaging system producing multiple images from two different angles for benthic macroinvertebrates. Using this latter imaging device and deep CNNs, Ärje et al. (submitted) have achieved classification accuracy within the range of taxonomic experts.

Our aim for this work was (a) to create a reproducible imaging system, (b) test the importance of different camera settings, (c) evaluate overall classification accuracy and (d) test the possibility of deriving biomass straight from geometrical features in images. To reach these objectives, we rebuilt the imaging system presented in Raitoharju et al. (2018) using industry components to make it completely reproducible. It has been made light proof to prevent extraneous light from affecting the images. We also developed a flushing mechanism to pass specimens through the imaging device. This is a critical improvement for automation as explained below.

For classification, we used Resnet-50 (He et al., 2016) and InceptionV3 (Szegedy, Canhoucke, Ioffe, Shlens, & Wojna, 2016) CNNs. We tested different camera settings (exposure time and aperture) to find the optimal settings for species identification, and we explored the necessary number of images per specimen to achieve high classification accuracy. Finally, for a subset of species, we tested if the area of a specimen derived directly from images taken by the device could serve as a proxy for biomass of the specimen.

## 2 | MATERIALS AND METHODS

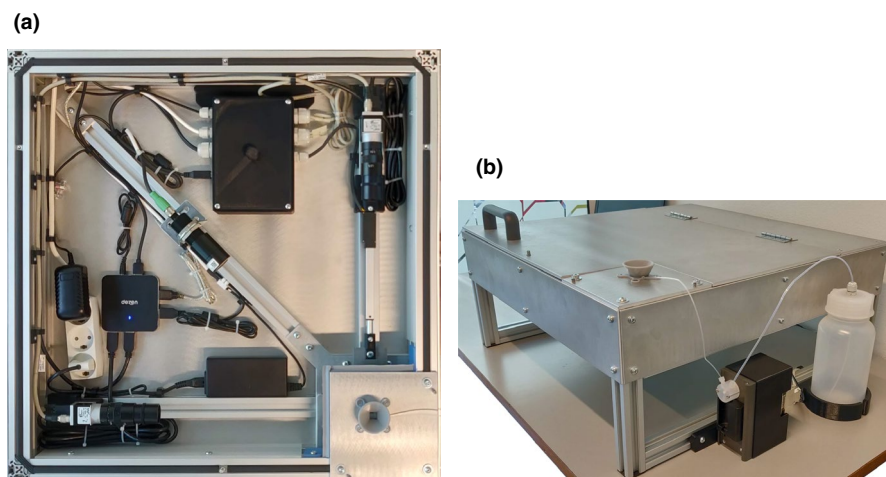
### 2.1 | The BIODISCOVER machine

To facilitate the automation of specimen identification, biomass estimation and sorting of invertebrate specimens, we improved the prototype imaging system developed for automatic identification of benthic macroinvertebrates (Raitoharju et al., 2018). We named the

new device BIODISCOVER machine, as an acronym for BIOlogical specimens Described, Identified, Sorted, Counted and Observed using Vision-Enabled Robotics. The system comprises an aluminium case with two Basler ACA1920-155UC cameras and LD75 lenses with  $\times 0.15$  to  $\times 0.35$  magnification and five aperture settings (maximum aperture ratio of 1:3.8). The cameras are placed at a  $90^\circ$  angle to each other in two corners of the case and in the other corners there are a high-power LED light (ODSX30-WHI Prox Light, which enables a maximum frame rate of 100 frames per second with an exposure time of  $1,000 \mu\text{s}$ ) and a rectangular cuvette made of optical glass and filled with ethanol. The inside of the case is depicted in Figure 1a. The case is rubber-sealed and has a lid to minimize extraneous light, shadows and other disturbances. The lid has an opening for the cuvette with a funnel for dropping specimens into the liquid. Figure 1b shows the new refill system, which pumps ethanol into the cuvette.

The multiview imaging component is connected to a computer with an integrated software, which controls all parts of the machine. The program uses calibration images to detect objects differing from the background and triggers the light and cameras to take images as the specimen sinks in ethanol until it disappears from the assigned view point of the cameras. The program detects the specimen and crops the images to be 496 pixels wide (defined by the width of the cuvette) and 496 pixels high while keeping the specimen at the center of the image with regards to the height. If a specimen exceeds the height of 496 pixels, the resulting images will be higher. The images are stored onto the computer as PNG files.

The BIODISCOVER machine enables imaging multiple specimens before the cuvette needs to be emptied and refilled. This is accommodated by a small area at the bottom of the cuvette, where the specimens are outside of the field of view of the cameras. Once a sample containing multiple specimens is imaged, the software triggers the opening of a sliding plate, which acts as a valve and flushes the specimens into a container below the imaging device case. Several containers placed in a rack can be controlled by the software based on input from the classification algorithm used to identify species. This enables a sorting of specimens into predefined classes based on size or taxonomy. In this way, the system can, for

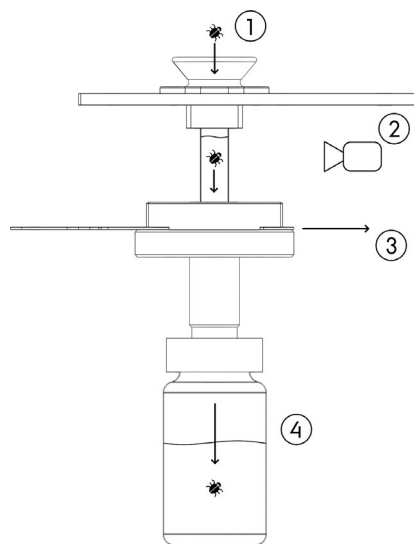


**FIGURE 1** The BIODISCOVER machine for imaging invertebrates with (a) depicting the inside of the case and (b) showing the new refill system which pumps ethanol into the cuvette used for imaging

instance, separate large and small specimens for further molecular study, separate insect orders, or separate common and rare species. The system is described in Figure 2. After the specimens have been flushed into the container for archiving, the pump in Figure 1b is used to refill the cuvette with ethanol.

## 2.2 | Classification experiments

Prior to large-scale imaging of reference collections of specimens of known identity, it is important to test the camera settings. As we plan to use the BIODISCOVER machine to create a large image database covering both terrestrial and aquatic invertebrates, it is important to optimize the different settings of the device to ensure the best possible image quality of the database with regard to classification accuracy. For this purpose, we imaged a pilot dataset with nine



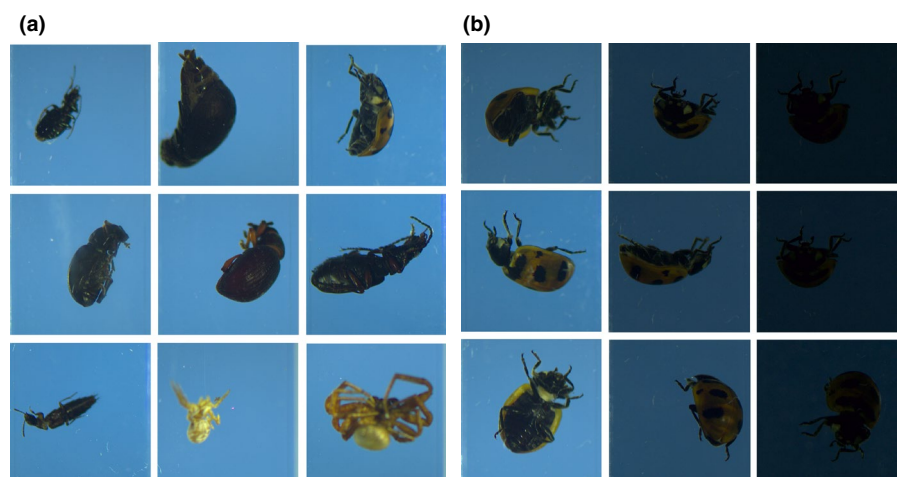
**FIGURE 2** The flush through system of the BIODISCOVER machine. (1) A funnel helps filling the cuvette with ethanol without airbubbles. (2) As the specimen floats in ethanol, two cameras capture images of it from two angles, (3) a valve is opened to flush the specimen through to (4) a container for further archiving

different combinations of camera settings. To study the importance of lighting, we explored the effect of varying exposure time values of (1,000, 1,500, 2,000  $\mu$ s) and to study the effect of the depth of field, we explored the effect of varying aperture values (1:3.8, 1:8, 1:16). Using the nine different combinations of camera settings, we imaged a dataset of nine terrestrial arthropod species, which were collected at Narsarsuaq, South Greenland and identified by morphology using Böcher, Kristensen, Pape, and Vilhelmsen (2015): *Bembidion grapii*, *Byrrhus fasciatus*, *Coccinella transversoguttata*, *Otiorhynchus arcticus*, *O. nodosus*, *Patrobis septentrionus*, *Quedius fellmanni*, *Xysticus deichmanni* and *X. durus* (see Figure 3a). For the pilot data, we wanted to include both species that have clear visual differences and should be easily identifiable and species from the same genera that have similar morphological features and are more difficult to tell apart.

The resulting nine datasets include the same specimens but the number of images varies depending on the camera settings since longer exposure time decreases the frame rate. Figure 3b shows the example images of the same *C. transversoguttata* specimen from each of the nine camera setting combinations. A few of the specimens were damaged during the imaging. Therefore to have comparable results, we removed any specimens that were not present in all nine datasets. In addition, we performed a crude, initial check for outliers by calculating the mean of blue, green and red pixel values per species and making a list of all specimens that had mean pixel values further than three standard deviations from the species average. We then manually checked the images of those listed specimens and removed images with only air bubbles or severed limbs. After this initial check, the number of images per specimen in the final data ranged from 1 to 376 (with 15 cases where a specimen had only 1 image). Table 1 gives the details on the final data.

We split the data into training (70%), validation (10%) and test (20%) observations. As difficult specimens can introduce variation to the results, we performed the tests on 10 different random data divisions. If a specimen was selected for training, all the images of that specimen were used for training. To keep the results comparable between the different camera settings, we used the exact same training specimens for all camera settings. Correspondingly, the exact

**FIGURE 3** (a) The nine species included in the image dataset. From top left: *Bembidion grapii*, *Byrrhus fasciatus*, *Coccinella transversoguttata*, *Otiorhynchus arcticus*, *O. nodosus*, *Patrobis septentrionus*, *Quedius fellmanni*, *Xysticus deichmanni* and *X. durus*. (b) Example images of a *C. transversoguttata* specimen with different camera settings. The exposure setting goes from top to bottom [1,000, 1,500, 2,000] and the aperture from left to right [3.8, 8, 16]



**TABLE 1** Image data details stating the number of images in each dataset imaged with different camera setting (exposure = [1,000, 1,500, 2,000] and aperture = [1:3.8, 1:8, 1:16]). The number of specimens is the same for all datasets

Species	#Specimens	#Images								
		1,000 1:3.8	1,000 1:8	1,000 1:16	1,500 1:3.8	1,500 1:8	1,500 1:16	2,000 1:3.8	2,000 1:8	2,000 1:16
<i>Bembidion grapii</i>	17	2,274	2,554	2,619	1,677	1,625	1,741	1,268	1,270	1,266
<i>Byrrhus fasciatus</i>	52	4,344	4,778	5,157	3,222	3,402	3,054	2,371	2,262	2,278
<i>Coccinella transversoguttata</i>	57	5,705	5,607	5,770	3,958	3,962	4,183	2,776	2,770	2,748
<i>Otiorhynchus arcticus</i>	50	3,197	3,318	3,488	2,556	2,220	2,225	1,898	1,614	1,700
<i>Otiorhynchus nodosus</i>	139	9,166	10,010	9,864	6,818	6,524	6,571	4,796	4,563	4,690
<i>Patrobis septentrionus</i>	108	11,056	11,583	11,738	8,383	8,311	8,028	6,004	5,808	6,148
<i>Quedius fellmanni</i>	42	5,749	6,438	6,363	4,577	4,461	4,393	3,708	3,270	3,318
<i>Xysticus deichmanni</i>	25	2,434	2,709	2,611	1,800	1,890	1,802	1,680	1,363	1,364
<i>Xysticus durus</i>	43	3,997	4,113	4,043	3,036	2,922	2,841	2,212	2,119	2,110

same validation and testing specimens were used for each camera setting combination. The number of images, exposure and aperture differed for the camera setting combinations, but the specimens remained the same, that is, if a difficult, atypical specimen of a certain species was selected for testing that same specimen was used for testing all the camera setting combinations, making the identification task equally difficult for all the settings.

To examine whether the BIODISCOVER machine benefits from having two cameras shooting from different angles, we performed a test where, for each specimen, we counted the numbers of images captured by each of the cameras. To compare the two camera angles, we selected an equal amount of images from both cameras. For each specimen, we checked which camera had captured less images and randomly sampled the same amount of images from the other camera as well. Finally, we randomly sampled the exact same amount of images for each specimen, this time including images from both cameras. Thus, we obtained three datasets, each with the same total amount of images. To account for variation in a single data split, we ran the test again on 10 data divisions into training, validation and test observations.

For the classification task, we tested two widely used deep CNN architectures, namely Resnet-50 (He et al., 2016) and InceptionV3 (Szegedy et al., 2016), both pretrained with the Imagenet database (Deng et al., 2009). For each data division, we used the training observations to fine-tune the weights of the pre-trained CNN. To feed the images to the network, we scaled them all to 128 × 128 pixels. This caused slight distortion to specimens taller than 496 pixels but the majority of the images (86%) are square-shaped and thus remained undistorted. We used batch normalization, a batch size of 128 and a decaying learning rate (0.001, 0.0001, 0.00001, 0.000001), training the network for 50 epochs with each learning rate. The validation images were used to select optimal weights for the network by comparing the validation accuracy after each epoch. Finally, the test observations were used to compute the final classification accuracy.

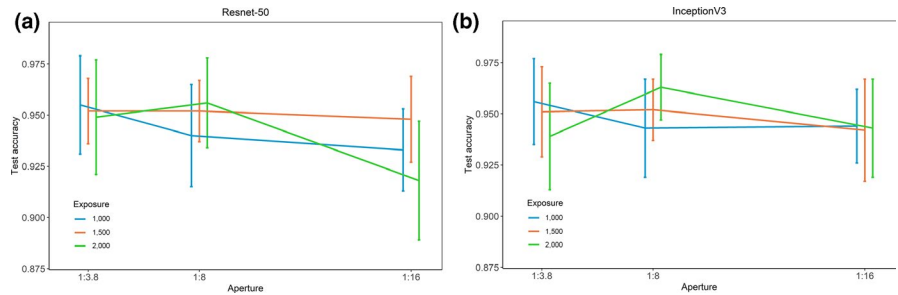
As we used multiple images per observation, we used a decision rule to determine the final species of the observation based on the predictions for all the images. The simplest option was to use majority vote, that is, the species that was predicted most often among the images of the specimen was chosen as the final prediction. The classification accuracy is reported as proportion of correctly predicted specimens based on the majority vote rule.

The BIODISCOVER machine derives geometric features from each image taken of each individual. These features include the area of the specimen in the image, which can be used for biomass prediction. For this purpose, we imaged three species of *Diptera* with the optimal camera settings and measured dry weight for a subset of this data ( $n = 65$ ). The species included in this dataset were *Dolichopus groenlandicus* ( $n = 25$ ,  $n_{\text{img}} = 1,788$ ), *D. plumipes* ( $n = 20$ ,  $n_{\text{img}} = 1,646$ ) and *Tachina ampliforceps* ( $n = 20$ ,  $n_{\text{img}} = 547$ ). The area was calculated from images as average per specimen. After imaging, each specimen was dried at 70°C for 48 hr and weighed on a scale to the nearest 0.0001 g to quantify dry weight. For biomass prediction, we performed a logarithmic transformation on the data and fitted a linear mixed model to examine the relationship between the average area and dry weight, using the species as a random factor. However, the model assumptions could not be met with the data; hence, we fitted separate generalized linear models for each species.

### 3 | RESULTS

Our first objective was to find optimal camera settings for the imaging device for species identification. The average classification accuracy across 10 test sets is presented in Figure 4 (and Table 3, Appendix S1). Based on the results for our pilot data, the optimal camera settings for both CNN were exposure = 2,000 μs and aperture = 1:8. InceptionV3 network produced the highest classification accuracy with these camera settings. For this network, the best camera settings also yielded the second lowest standard deviation. The differences between the

**FIGURE 4** Average test classification accuracies ± standard deviation for different camera setting combinations for (a) Resnet-50 network and (b) InceptionV3 network



**TABLE 2** Normalized average confusion matrix over 10 random data splits for data imaged with exposure = 2,000 μs and aperture = 1:8, classified with InceptionV3 network. The rows of the table represent the true species while the columns represent the predicted species and the cells give the average percentage over the 10 test data. The species-wise classification accuracy is marked in bold

	<i>Be_gr</i>	<i>By_fa</i>	<i>Co_tr</i>	<i>Do_gr</i>	<i>Do_pl</i>	<i>Ot_ar</i>	<i>Ot_no</i>	<i>Pa_se</i>	<i>Qu_fel</i>	<i>Ta_am</i>	<i>Xy_de</i>	<i>Xy_du</i>
<i>Be_gr</i>	<b>0.756</b>	0	0	0	0	0	0	0.122	0.122	0	0	0
<i>By_fa</i>	0	<b>0.992</b>	0	0	0	0	0.008	0	0	0	0	0
<i>Co_tr</i>	0	0	<b>1.0</b>	0	0	0	0	0	0	0	0	0
<i>Do_gr</i>	0	0	0	<b>1.0</b>	0	0	0	0	0	0	0	0
<i>Do_pl</i>	0	0	0	0.015	<b>0.985</b>	0	0	0	0	0	0	0
<i>Ot_ar</i>	0	0	0	0	0	<b>0.910</b>	0.090	0	0	0	0	0
<i>Ot_no</i>	0	0.004	0	0	0	0.019	<b>0.977</b>	0	0	0	0	0
<i>Pa_se</i>	0	0	0	0	0	0.004	0.004	<b>0.991</b>	0	0	0	0
<i>Qu_fel</i>	0	0	0.010	0	0.010	0	0	0	<b>0.980</b>	0	0	0
<i>Ta_am</i>	0	0	0	0.005	0	0	0	0.010	0	<b>0.985</b>	0	0
<i>Xy_de</i>	0	0	0	0	0	0	0	0	0	0	<b>0.941</b>	0.059
<i>Xy_du</i>	0.029	0	0	0	0	0	0	0	0.010	0	0.029	<b>0.933</b>

settings were small but we observed that decreasing aperture to 1:16 decreased the classification accuracy. For higher exposure, an initial decrease in aperture enhanced the results while decreasing aperture to 1:16 led to decreased classification accuracy. For exposure = 1,000 μs, even increasing aperture to 8 decreased classification accuracy. The optimal camera settings are intuitive as they provide sharp images while having as much light as possible.

In addition to the majority vote decision rule, we also calculated classification accuracy using the weighted sum rule (Raitoharju & Meissner, 2019). The results were similar but led to slightly lower overall accuracy (Table 4, Appendix S1).

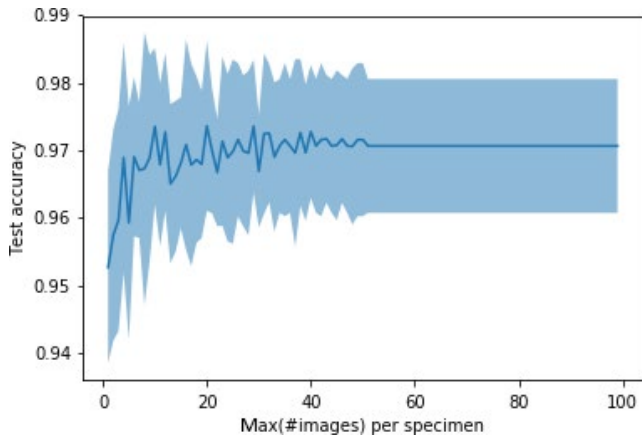
To test whether the BIODISCOVER machine benefits from having two cameras shooting from different angles, we performed a test on the data imaged with the optimal camera settings. The classification accuracy was higher when using images from both cameras (Table 5, Appendix S1). In addition, for Resnet-50, the standard deviation was lower meaning there is less variation in the classification accuracy due to choice of test specimens. The classification accuracies in this test are slightly lower than in Table 3 as for this particular test we are using less images per specimen (approximately 50%).

Once we optimized the camera settings, we re-trained the InceptionV3 network with the data including also the three *Diptera* species. The average classification accuracy over 10 test sets was 0.980. The information of individual classification decisions is

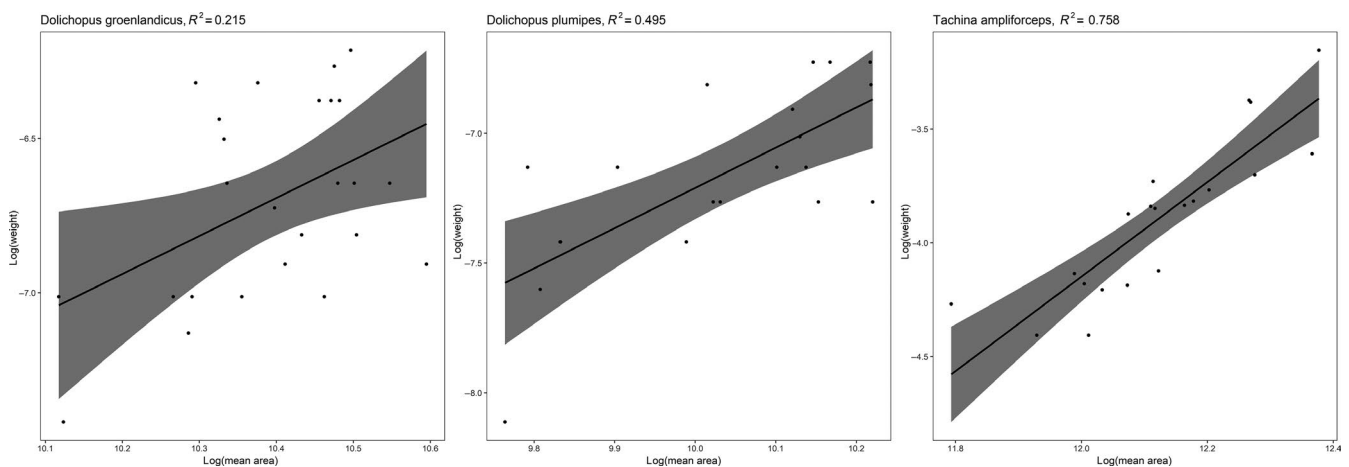
shown in a confusion matrix with the true species on the rows and the predicted species on the columns. Table 2 displays the normalized average confusion matrix over the 10 random data splits for InceptionV3 CNN with the optimal camera settings. As for individual species, *B. grapii* was the hardest to identify. Some of the specimens were misclassified as *P. septentrionus* and *Q. fellmanni*. In addition, *Otiorhynchus arcticus* and *O. nodosus* were often confused, as well as *X. deichmanni* and *X. durus*. Other common classification errors were misclassifying *B. fasciatus* as *O. nodosus* and misclassifying *X. durus* as *B. grapii*. The species that performed poorly compared to the others were species with the lowest number of images in the data. The accuracy could be improved by collecting more data on these species or using data augmentation techniques.

When considering automated biomonitoring, one key factor is the time it takes to automatically identify the taxonomic identity of a specimen. In taxa identification scenarios, optimizing the time used for testing is more important than optimizing the time used for training. The training time for the network using the optimal camera settings was 13 hr and 55 min (see Table 6, Appendix S1, for other camera settings). However, training of the network needs to be done only once. The number of images per specimen affects the total time of identification as each image needs a prediction. To optimize the number of images per specimen, we tested how this affects the classification accuracy. As the specimens had varying number of images,

we tested with the maximum number of images per specimen,  $N_{\max}$ . If a specimen had less images, we used all of them. If a specimen had more images, we randomly sampled  $N_{\max}$  of them. Again, we ran this test on the 10 data splits imaged with the optimal camera settings. The results are shown in Figure 5. The average number of images per specimen is 47 so while some specimens had over 100 images, the test accuracy stabilized at approximately 50 images. The same accuracy of approximately 96% could already be achieved with 20 images per specimen but lower numbers of images increased the variation in the classification accuracy. While increasing the maximum number of images per specimen does increase the time for taxa predictions, testing time is not an issue. Even with a maximum 100 images per specimen, the time taken to predict taxa for the entire beetle and spider test data of 110 specimens with a total 5,300 images was on average 40 s. However, fixing the maximum number of images per specimen would mean less images for the BIODISCOVER device to record and store onto the computer, enabling faster imaging process and saving computational resources.



**FIGURE 5** Classification accuracy of the test specimens plotted against the maximum number of images per specimen. The solid line shows the average over 10 data divisions and the light blue area represents the average  $\pm$  standard deviation



**FIGURE 6** Generalized linear models for biomass of *Dolichopus groenlandicus*, *D. plumipes* and *Tachina ampliforceps* with average area predicting dry weight

Figure 6 shows the results of the biomass prediction. The logarithm-transformed average area was found to be statistically significant predictor of dry weight for all three *Diptera* species (see Table 7 in Appendix S1). However, considering the  $R$ -squares of the different models, the average area is a good predictor only for the largest species, *T. ampliforceps* ( $r^2 = 0.758$ ). For the two small *Dolichopus* species, relationships were weaker.

## 4 | DISCUSSION

We have presented an image-based identification system (i.e. the BIODISCOVER machine) for insects and other invertebrates as an alternative to manual identification. We demonstrated a very high classification accuracy on a test set of images of 249 specimens of known identity belonging to one of 12 insect and spider species. We were also able to show that biomass of individual specimens could be predicted straight from information in the images. Together, our results pave the way for future non-destructive, automatic, image-based identification and biomass estimation of bulk invertebrate samples.

We imaged specimens of seven beetle, two spider and three fly species with the BIODISCOVER machine with different values for exposure time and aperture settings and found that the best classification accuracy was obtained with an exposure time 2,000  $\mu$ s and an aperture 1:8. With these settings, we obtained a high test classification accuracy of 98.0%, demonstrating the great potential of the BIODISCOVER machine for the use in species identification. In Ärje et al. (submitted), for example, taxonomic experts achieved an accuracy of 93.9% with a dataset of 39 taxonomic groups analyzing physical specimens, not images. Compared to human experts or genetic methods, a key challenge for the machine learning approach is to identify species which are not already known to the reference database. While this is an area of active research, it currently imposes a strong demand for comprehensive reference databases of species of known identity. While such databases are being

constructed, classification can be carried out at a higher taxonomic level. While adding more species to the data will increase the difficulty of the classification task (Ärje et al., submitted) as well as the training time, data augmentation and adding more specimens of each species can be used to improve the results for rare species (Raitoharju et al., 2016; Sohrab & Raitoharju, submitted). Possibly data augmentation techniques can be also developed for predicting the species for fragmented specimens often found in samples. If the fragment is small compared to the original specimen, it will likely be difficult to assign to a specific class, but most likely the confidence value will be low for all classes, which will allow the user to recognize it is a fragment. The same would go for other organic or inorganic debris.

Regarding training time, Imagenet (Deng et al., 2009) includes a total 21,841 subcategories with over 500 images per subcategory. Training time of deep learning models on such big datasets is long.<sup>1</sup> Recent research in parallel computing (Goyal et al., 2017) has shown that distributed synchronous Stochastic Gradient Descent may be the way to overcome this issue. To estimate testing time, we must consider the entire process involving entering the metadata related to the sample into the software, moving each individual specimen to the cuvette and predicting the species. The imaging is done as soon as the specimen reaches the bottom of the cuvette and the trained algorithm can run on the same computer, which runs the imaging software. The classification task with the current database takes <1 s per specimen and if specimens from a sample can be placed together in the same vial after processing, the whole process can be done in about 10 min for a sample of 100 specimens.

We tested predicting biomass from images on a subset of three fly species. We explored a joined mixed model for all species but the small data restricted our final analysis to three species-wise generalized linear models. The average area of the specimen was a good predictor for dry weight for the largest species, *T. ampliforceps*, but the two smaller species would require more data for better results. For instance, by weighing more species of different sizes, it would be possible to quantify the uncertainty associated with using general relationships between area and dry weight constructed from multiple, related species (e.g. species belonging to the same family). The BIODISCOVER machine can easily be used with any animal small enough to fit into the cuvette. Since the imaging device comprises standard industry components with a total cost of approximately €5,000, we have made it possible to build more copies of the BIODISCOVER machine. The flow-through and refill systems facilitate easy archiving of samples. Furthermore, the BIODISCOVER machine also saves metadata from the images, for example, geometric features that can be used in automatic biomass predictions.

The imaging device is one of three components for automatic image-based species identification. We are currently working on implementing (a) a computer-vision-enabled robotic arm to

automatically detect insects from a bulk sample in a tray, choose among different tools to move individual specimens to the imaging device and (b) a sorting rack to place specimens in the preferred container after imaging based on, for example, taxonomic identity, size or rarity. With these additions, the BIODISCOVER machine offers high-throughput, non-destructive taxonomic identification, size/biomass estimation, counting and further morphological data, while keeping the specimens intact. Given that the robotic arm is standard industry equipment, we are on the verge of producing a truly automated species identification system for invertebrates, both aquatic and terrestrial.

## ACKNOWLEDGEMENTS

We would like to thank CSC for computational resources. T.T.H. acknowledges funding from his VILLUM Experiment project 'Automatic Insect Detection' (grant 17523) and an Aarhus University synergy grant. The authors have no conflicts of interest to declare.

## AUTHORS' CONTRIBUTIONS

J.Ä. drafted the paper with contributions from T.T.H. with the other authors providing feedback and approving the final manuscript; C.M., T.T.H., M.R.J. and S.A.M. designed and built the BIODISCOVER machine with inputs from K.M. and V.T. who had designed the prototype described in Raitoharju et al. (2018); M.S.R. imaged the arthropod specimens and J.Ä., T.T.H., A.I.; M.G. and J.R. designed the classification experiments for the data.

## DATA AVAILABILITY STATEMENT

The image data (Ärje et al., 2020) has been made publicly available on Zenodo.org.

## ORCID

Johanna Ärje  <https://orcid.org/0000-0003-0710-9044>

Toke Thomas Høye  <https://orcid.org/0000-0001-5387-3284>

## REFERENCES

- Ärje, J., Melvad, C., Jeppesen, M. R., Madsen, S. A., Raitoharju, J., Rasmussen, M. S., ... Høye, T. T. (2020). BIODISCOVER data I. *Zenodo*. <https://doi.org/10.5281/zenodo.3826426>
- Ärje, J., Raitoharju, J., Iosifidis, A., Tirronen, V., Meissner, K., Gabbouj, M., ... Kärkkäinen, S. (submitted). Human experts vs. machines in taxa recognition. *arXiv Preprint*: 1708.06899v4.
- Aylagas, E., Borja, A., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. *Frontiers in Marine Science*, 3, 1809–1812. <https://doi.org/10.3389/fmars.2016.00096>
- Aylagas, E., Borja, A., Muxika, I., & Rodríguez-Ezpeleta, N. (2018). Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. *Ecological Indicators*, 95, 194–202. <https://doi.org/10.1016/j.ecolind.2018.07.044>
- Böcher, J., Kristensen, N. P., Pape, T., & Vilhelmsen, L. (Eds.). (2015). *The Greenland entomofauna: An identification manual of insects, spiders and their allies*. Boston, MA: Brill.
- Bochinski, E., Bacha, G., Eiselein, V., Walles, T. J. W., Nejstgaard, J. C., & Sikora, T. (2018). Deep active learning for in situ plankton classification. In *International Conference on Pattern Recognition* (pp. 5–15). Cham, Switzerland: Springer.

<sup>1</sup>As an example, training a rather small network topology ResNet18 on ImageNet requires 40 hr using 2 V100 GPUs, each having 32GB of RAM.



- Borja, A., & Elliott, M. (2013). Marine monitoring during an economic crisis: The cure is worse than the disease. *Marine Pollution Bulletin*, 68, 1–3. <https://doi.org/10.1016/j.marpolbul.2013.01.041>
- Dai, J., Wang, R., Zheng, H., Ju, G., & Qiao, X. (2016). ZooplanktonNet: Deep convolutional network for zooplankton classification. In *OCEANS 2016* (pp. 1–6). Shanghai, China: IEEE.
- Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *IEEE Computer Vision and Pattern Recognition (CVPR)*. (pp. 248–255). IEEE.
- Ding, W., & Taylor, G. (2016). Automatic moth detection from trap images for pest management. *Computers and Electronics in Agriculture*, 123, 17–28. <https://doi.org/10.1016/j.compag.2016.02.003>
- Dunker, K. J., Spulveda, A. J., Massengill, R. L., Olsen, J. B., Russ, O. L., Wenburg, J. K., & Antonovich, A. (2016). Potential of environmental DNA to evaluate northern pike (*Esox lucius*) eradication efforts: An experimental test and case study. *PLoS ONE*, 11(9), e0162277. <https://doi.org/10.1371/journal.pone.0162277>
- Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, 8(10), 1265–1275. <https://doi.org/10.1111/2041-210X.12789>
- Feng, L., Bhanu, B., & Heraty, J. (2016). A software system for automated identification and retrieval of moth images based on wing attributes. *Pattern Recognition*, 51, 225–241. <https://doi.org/10.1016/j.patcog.2015.09.012>
- Gaston, K. J., & O'Neill, M. A. (2004). Automates species identification: Why not? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1444), 655–667.
- Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., ... He, K. (2017). Accurate, large minibatch SGD: Training ImageNet in 1 hour. *arXiv Preprint: 1706.02677*.
- Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., ... de Kroon, H. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS ONE*, 12(10), e0185809. <https://doi.org/10.1371/journal.pone.0185809>
- Hansen, R. R., Hansen, O. L. P., Bowden, J. J., Treier, U. A., Normand, S., & Høye, T. T. (2016). Meter scale variation in shrub dominance and soil moisture structure Arctic arthropod communities. *PeerJ*, 4, e2224. <https://doi.org/10.7717/peerj.2224>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). IEEE.
- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Logo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Høye, T. T., & Forchhammer, M. C. (2008). Phenology of high-arctic arthropods: Effects of climate on spatial, seasonal and inter-annual variation. *Advances in Ecological Research*, 40, 299–324.
- Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Frigerio, J., Humbert, J., & Bouchez, A. (2014). A next-generation sequencing approach to river biomonitoring using benthic diatoms. *Freshwater Science*, 33, 349–363. <https://doi.org/10.1086/675079>
- Keskin, E. (2014). Detection of invasive freshwater fish species using environmental DNA survey. *Biochemical Systematics and Ecology*, 56, 68–74.
- LeQuing, Z., & Zhen, Z. (2012). Automatic insect classification based on local mean colour feature and supported vector machines. *Oriental Insects*, 46(3/4), 260–269. <https://doi.org/10.1080/00305316.2012.738142>
- Liu, F., Shen, Z.-R., Zhang, J.-W., & Yang, H.-Z. (2008). Automatic insect identification based on color characters. *Chinese Bulletin of Entomology*, 45, 150–153.
- Loboda, S., Savage, J., Buddle, C. M., Schmidt, N. M., & Høye, T. T. (2018). Declining diversity and abundance of High Arctic fly assemblages over two decades of rapid climate warming. *Ecography*, 41, 265–277. <https://doi.org/10.1111/ecog.02747>
- Nygård, H., Oinonen, S., Lehtiniemi, M., Hällfors, H., Rantajarvi, E., & Uusitalo, L. (2016). Price versus value of marine monitoring. *Frontiers in Marine Science*, 3, 205.
- Perre, P., Faria, F. A., Jorge, L. R., Rocha, A., Torres, R. S., Souza-Filho, M. F., ... Zucchi, R. A. (2016). Toward and automated identification of *Anastrepha* fruit flies in the *fraterculus* group (Diptera, Tephritidae). *Neotropical Entomology*, 45(5), 554–558.
- Potamis, I. (2014). Automatic classification of a taxon-rich community recorded in the wild. *PLoS ONE*, 9(5), e96936.
- Qian, L., HongBin, W., Zhen, Z., & XiangBo, K. (2011). Automatic stridulation identification of bark beetles based on MFCC and BP network. *Jourlan of Beijing Forestry University*, 33(5), 81–85.
- Raitoharju, J., Årje, J., Iosifidis, A., Tirronen, V., Meissner, K., Gabbouj, M., ... Kärkkäinen, S. (2019). *FIN-Benthic 2*. Etsin. Retrieved from <http://urn:nbn:fi:csc-kata20181023162421949328>
- Raitoharju, J., & Meissner, K. (2019). On confidences and their use in (semi-)automatic multi-image taxa identification. In *Proceeding of IEEE Symposium Series on Computational Intelligence* (pp. 1338–1343). IEEE.
- Raitoharju, J., Riabchenko, E., Ahmad, I., Iosifidis, A., Gabbouj, M., Kiranyaz, S., ... Meissner, K. (2018). Benchmark database for fine-grained image classification of benthic macroinvertebrates. *Image and Vision Computing*, 78, 73–83. <https://doi.org/10.1016/j.imavis.2018.06.005>
- 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 *ICPR 2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI)*. (pp. 43–48). IEEE.
- Raupach, M. J., Astrin, J. J., Hannig, K., Peters, M. K., Stoeckle, M. Y., & Wägele, J.-W. (2010). Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology*, 7, 26. <https://doi.org/10.1186/1742-9994-7-26>
- Rich, M. E., Gough, L., & Boelman, N. T. (2013). Arctic arthropod assemblages in habitats of differing shrub dominance. *Ecography*, 36, 994–1003. <https://doi.org/10.1111/j.1600-0587.2012.00078.x>
- Santhi, N., Pradeepa, C., Subashini, P., & Kalaiselvi, S. (2013). Automatic identification of algal community from microscopic images. *Bioinformatics and Biology Insights*, 7, 3. <https://doi.org/10.4137/BBI.S12844>
- Schröder, S., Drescher, W., Steinhage, V., & Kastenholz, B. (1995). An automated method for the identification of bee species (Hymenoptera: Apoidea). In *Proceedings of the International Symposium on Conserving Europe's Bees*. London, UK: International Bee Research Association & Linnean Society.
- Seibold, S., Gossner, M. M., Simons, N. K., Blütgen, N., Müller, J., Ambarli, D., ... Weisser, W. W. (2019). Arthropod decline in grasslands and forests is associated with landscape-level drivers. *Nature*, 574, 671–674. <https://doi.org/10.1038/s41586-019-1684-3>
- Sohrab, F., & Raitoharju, J. (submitted). Boosting rare benthic macroinvertebrate taxa identification with one-class classification. *arXiv Preprint: 2002.10420*.
- Srivathsan, A., Baloglu, B., Wang, W., Tan, W., Bertrand, D., Ng, A., ... Meier, R. (2018). A MinION™-based pipeline for fast and cost-effective DNA barcoding. *Molecular Ecology Resources*, 18(5), 1035–1049. <https://doi.org/10.1111/1755-0998.12890>
- Srivathsan, A., Hartop, E., Puniamoorthy, J., Lee, W., Kutty, S., Kurina, O., & Meier, R. (2019). Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biology*, 17(96), 1–20. <https://doi.org/10.1186/s12915-019-0706-9>
- Szegedy, C., Canhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethink the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2818–2826). IEEE.

- Timms, L. L., Bowden, J. J., Summerville, K. S., & Buddle, C. M. (2012). Does species-level resolution matter? Taxonomic sufficiency in terrestrial arthropod biodiversity studies. *Insect Conservation and Diversity*, 6, 453–462. <https://doi.org/10.1111/icad.12004>
- Valan, M., Makonyi, K., Maki, A., Vondráček, D., & Ronquist, F. (2019). Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. *Systematic Biology*, 68(6), 876–895. <https://doi.org/10.1093/sysbio/syz014>
- Van Horn, G., Aodha, O. M., Song, Y., Cui, Y., Sun, C., Shepard, A., ... Belongie, S. (2018). The iNaturalist species classification and detection dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 8769–8778). IEEE.
- Wagner, D. L. (2019). Insect declines in the anthropocene. *Annual Review of Entomology*, 65, 457–480. <https://doi.org/10.1146/annurev-ento-011019-025151>
- Weeks, P. J. D., Gauld, I. D., Gaston, K. J., & O'Neill, M. A. (1997). Automating the identification of insects: A new solution to an old problem. *Bulletin of Entomological Research*, 87(2), 203–211. <https://doi.org/10.1017/S000748530002736X>
- Xia, D., Chen, P., Wang, B., Zhang, J., & Xie, C. (2018). Insect detection and classification based on an improved convolutional neural network. *Sensors (Basel)*, 18(12), 4169. <https://doi.org/10.3390/s18124169>
- Zhang, X., Gao, Y., & Caelli, T. (2010). Primitive-based 3D structure inference from a single 2D image for insect modeling: Towards an electronic field guide for insect identification. In *11th International Conference on Control Automation Robotics & Vision, IEEE* (pp. 866–871). IEEE.
- Zimmermann, J., Glockner, G., Jahn, R., Enke, N., & Gemeinholzer, B. (2015). Meta-barcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources*, 15, 526–542. <https://doi.org/10.1111/1755-0998.12336>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Ärje J, Melvad C, Jeppesen MR, et al. Automatic image-based identification and biomass estimation of invertebrates. *Methods Ecol Evol*. 2020;11:922–931. <https://doi.org/10.1111/2041-210X.13428>