RESOURCE ARTICLE

DiversityScanner: Robotic handling of small invertebrates with machine learning methods

Lorenz Wührl[1] ⓘ　|　Christian Pylatiuk[1] ⓘ　|　Matthias Giersch[1]　|　Florian Lapp[1]　|
Thomas von Rintelen[2] ⓘ　|　Michael Balke[3] ⓘ　|　Stefan Schmidt[3] ⓘ　|　Pierfilippo Cerretti[4] ⓘ　|
Rudolf Meier[2] ⓘ

[1]Institute for Automation and Applied Informatics (IAI), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]Museum für Naturkunde, Center for Integrative Biodiversity Discovery, Leibniz-Institut für Evolutions- und Biodiversitätsforschung, Berlin, Germany

[3]SNSB – Zoologische Staatssammlung München, Munich, Germany

[4]Department of Biology and Biotechnology 'Charles Darwin', Sapienza University of Rome, Rome, Italy

Correspondence
Christian Pylatiuk, Institute for Automation and Applied Informatics (IAI), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.
Email: pylatiuk@kit.edu

Rudolf Meier, Museum für Naturkunde, Center for Integrative Biodiversity Discovery, Leibniz-Institut für Evolutions- und Biodiversitätsforschung, Berlin, Germany
Email: rudolf.meier@mfn.berlin

Funding information
Center for Integrative Biodiversity Discovery at the Museum für Naturkunde Berlin

Abstract

Invertebrate biodiversity remains poorly understood although it comprises much of the terrestrial animal biomass, most species and supplies many ecosystem services. The main obstacle is specimen-rich samples obtained with quantitative sampling techniques (e.g., Malaise trapping). Traditional sorting requires manual handling, while molecular techniques based on metabarcoding lose the association between individual specimens and sequences and thus struggle with obtaining precise abundance information. Here we present a sorting robot that prepares specimens from bulk samples for barcoding. It detects, images and measures individual specimens from a sample and then moves them into the wells of a 96-well microplate. We show that the images can be used to train convolutional neural networks (CNNs) that are capable of assigning the specimens to 14 insect taxa (usually families) that are particularly common in Malaise trap samples. The average assignment precision for all taxa is 91.4% (75%–100%). This ability of the robot to identify common taxa then allows for taxon-specific subsampling, because the robot can be instructed to only pick a prespecified number of specimens for abundant taxa. To obtain biomass information, the images are also used to measure specimen length and estimate body volume. We outline how the DiversityScanner can be a key component for tackling and monitoring invertebrate diversity by combining molecular and morphological tools: the images generated by the robot become training images for machine learning once they are labelled with taxonomic information from DNA barcodes. We suggest that a combination of automation, machine learning and DNA barcoding has the potential to tackle invertebrate diversity at an unprecedented scale.

KEYWORDS
automation, biodiversity, biomass, convolutional neural network, "dark taxa", DNA barcoding

1 | INTRODUCTION

Biodiversity science is currently at an inflection point. For decades, biodiversity loss had been mostly an academic concern, although many biologists had already predicted that the decline would eventually threaten whole ecosystems (May, 2011). Unfortunately, we are now at this stage, which explains why the World Economic Forum considers biodiversity loss as one of the top three global risks based on likelihood and impact for the next 10 years (World Economic Forum's Global Risk Initiative 2020). This new urgency is also leading to a reassessment of research priorities. Most biologists had traditionally focused on charismatic taxa (birds, mammals, butterflies, etc.) with a preference for endangered or even extinct species (Ceballos et al., 2017). However, with regard to quantitative arguments, this focus has always been poorly motivated. If one were to adopt a quantitative point of view of terrestrial animal diversity, invertebrates would receive most attention. They contribute more than 45 times the biomass of wild vertebrates (table S23 in Bar-On et al., 2018), contain >90% of the estimated species diversity (e.g., Groombridge, 1992), and comprise much of the functional and evolutionary diversity. In 2011 Robert May (2011) stated that "[w]e are astonishingly ignorant about how many species are alive on earth today, and even more ignorant about how many we can lose (and) yet still maintain ecosystem services that humanity ultimately depends upon." This situation can only be changed if new and efficient tools for assessing and monitoring invertebrate biodiversity are developed. Such tools need to be particularly suitable for clades that have come to be referred to as "dark taxa" and which Hartop et al. (2021) defined as those "for which the undescribed fauna is estimated to exceed the described fauna by at least one order of magnitude and the total diversity exceeds 1000 species."

Biomonitoring of these taxa with morphological tools or DNA barcoding is very time-consuming because it requires processing thousands of typically small specimens. This goes some way to explain why metabarcoding of bulk invertebrate samples has become increasingly applied. It allows for fast processing of bulk samples and yields information on species composition. However, using this method comes at the cost of not being able to obtain precise abundance data (Creedy et al., 2019), although monitoring population declines is important (Ceballos et al., 2017). Furthermore, the missing association between DNA sequences and individual specimens constrains follow-up research. For example, species new to science will remain undescribed because specimens belonging to the undescribed species cannot be readily located. Similarly, specimen- or species-specific studies addressing the role of species in the ecosystem cannot be carried out, although much could be deduced by, for example, sequencing the microbiome (e.g., Six, 2013) or gut content (e.g., Reeves et al., 2018). Overall, it is therefore desirable to develop not only bulk sequencing strategies for mass samples, but also to modernize specimen-based processing.

We here argue that three technical developments can help with achieving this goal. The first is cost-effective methods for obtaining and sequencing specimen-specific barcode amplicons with second- and third-generation sequencing technologies (Hebert et al., 2018; Srivathsan et al., 2019a, 2021; Wang et al., 2018). Indeed, today's consumable cost for barcoding a sample with 1000 specimens is <100 USD (Srivathsan et al., 2021) and portable sequencers produced by Oxford Nanopore Technologies are democratizing access to DNA sequences (Buchner et al., 2021; Pomerantz et al., 2018; Srivathsan et al., 2021; Watsa et al., 2020). Unfortunately, automation and data processing with neural networks, which present the other two developments, remain underutilized. Currently, automation mostly exists in the form of pipetting robots in molecular laboratories. However, the main challenge posed by bulk samples is the imaging and movement of individual specimens into microplates. With regard to the use of neural networks, they are currently widely used for identifying plant and charismatic vertebrate taxa (Fairbrass et al., 2019; Milošević et al., 2020; Stowell et al., 2019; Tabak et al., 2019), but invertebrates in bulk samples have benefited very little (but see Ärje et al., 2020b). Yet, thousands of samples are collected every day. They include plankton samples in marine biology, macroinvertebrate samples used for assessing freshwater quality, and mass insect samples (Borkent & Brown, 2015; Brown, 2005; Brown et al., 2018; Karlsson et al., 2020b). Here, well-trained convolutional neural networks (CNNs) would be important because they could use images to (a) identify specimens to species, (b) provide specimens for follow-up research (e.g., microbiome), (c) yield precise abundance information and (d) measure biomass. All this would enable semi-automated biomonitoring of invertebrates when samples obtained from the same place at different times are processed.

Computer-based identification systems for invertebrates are starting to yield promising results (Feng et al., 2016; Knyshov et al., 2021; Perre et al., 2016). For example, a recently developed system can size and identify stoneflies (Plecoptera) that are routinely used for freshwater quality assessment (Sarpola et al., 2008). Another system processes samples consisting of soil mesofauna (Chamblin et al., 2011). However, this system is comparatively expensive because it uses a robotic arm. Other robots have been designed for specific, commercial insect sorting purposes. This includes one that can separate intact mealworm larvae (*Tenebrio molitor*) from skins, faeces and dead worms (Kim, 2014) and one that sorts mosquitoes (Lepek et al., 2020) and is capable of distinguishing males from females. However, all these machines lack the ability to recognize a wide variety of insect specimens in bulk invertebrate samples. The machine closest to this capability is the BIODISCOVER by Ärje et al. (2020a), which can identify ethanol-preserved specimens, which have to be fed into the machine manually one by one. In addition, all specimens are returned into the same container after identification.

Most of these systems use deep CNNs with transfer learning (Ärje et al., 2020b) and thus require large sets of training images. Arguably, the lack of such sets is the main obstacle for applying CNN to invertebrates. Here, robotics could have a major impact if automatic imaging were to be combined with DNA barcoding. Robots could provide the images, which would then be assigned/labelled with taxonomic information obtained with DNA barcodes. Such barcodes can be used to sort specimens to putative species ("MOTUs"

[molecular operational taxonomic units]) with good overall congruence to morphospecies (e.g., Wang et al., 2018). Comparing the barcodes with public databases will then reveal for which specimens a preliminary MOTU ID can be replaced with a scientific name. Imaging combined with labelling at species-level resolution will thus be able to yield training sets for CNNs.

We here describe a new robot (DiversityScanner) that recognizes insect specimens based on an overview image of a sample and processes small specimens (<3 mm), which contribute >60% of all specimens in Malaise traps. This estimate is based on analyses of taxonomic compositions of samples from Sweden (Karlsson et al., 2020a) and Neotropical countries (Brown, 2005). The analyses revealed that 64%–84% of the specimens belonged to Diptera families that contain predominantly small species (e.g., Chironomidae, Sciaridae, Phoridae, Cecidomyiidae, Mycetophilidae). Furthermore, another >5% of all specimens in the Swedish samples were small parasitoid wasps (<3 mm: Diaprioidea, Chalcidoidea, Platygastroidea). After identifying specimens of appropriate size and distance to other specimens, the DiversityScanner images each suitable specimen and moves it into a microplate. The robot then uses these images to assign the specimens into 14 common "classes" of insects (usually family-level) using a CNN. Lastly, the images are used to estimate biomass based on insect length and an estimated volume.

2 | CONCEPT AND METHODS

We here present a compact insect sorting robot (Figure 1) that assigns objects (mostly insects) to different classes. Note that we here use the term in the context of machine learning. Indeed, most of the classes in our study are families in the Linnean system ($N = 10$), two contain two families and two are of higher rank (Calyptratae and the paraphyletic acalyptrate Diptera). To ensure accessibility, our robot relies mostly on standard, commercially available components that are connected via parts that can be printed on a commercial (Fused Deposition Modeling—FDM) 3D printer. The basic design uses a cube-shaped frame (50 × 50 × 50 cm) as well as three linear drives with accurately positioning stepper motors and is based on a zebrafish embryo handling robot that was developed earlier (Pfriem et al., 2012). The robot is equipped with two high-resolution cameras ("overview" and "specimen" camera) with customized lenses, LED lighting and image recognition software. Furthermore, a specimen transport system using a suction pump is integrated to transfer insects into the wells of a standard 96-well microplate. Thus, the robotic system can be divided into: (a) the transport system, (b) the image acquisition system and (c) the image processing system. All are operated by a graphical user interface (GUI) on a touchscreen.

2.1 | Transport system

The x- and y-axes of the robot are realized by LEZ1 linear drives (Isel AG) and connected to the outer frame of the robot at half height.
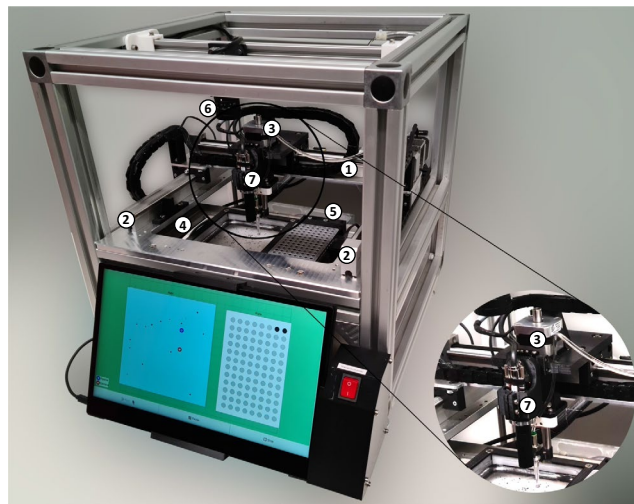

FIGURE 1 DiversityScanner with 1: x-axis; 2: y-axis; 3: z-axis; 4: Petri dish; 5: microwell plate; 6: overview camera (C1), 7: specimen camera (C2). The touch screen provides updates about the sorting process (e.g., insect detection, position of pipette tip and specimen carmera) and can be used to start and stop the robot. The Raspberry Pi, motor control unit and syringe pump are hidden behind the display panel

Both linear drives are driven by high-precision stepper motors to ensure good positioning accuracy. The y-axis is moved by the orthogonally connected shaft slide of the x-axis. The shaft slide of the y-axis transports the specimen camera and the z-axis with the suction hose. In order to move the suction hose in the z-direction (=up and down), the z-axis is driven by an AR42H50 spindle drive with stepper motor (Nanotec Electronic). All three axes are controlled by a single TMCM-3110 motor controller (Trinamic) that allows for precise, fast and smooth movements. The motor controller and the other electronics are protected from water and ethanol droplets by a box at the bottom of the robot. The transport system is controlled by a Raspberry Pi 4 (Model B, 4 GB) single-board computer that was programmed in Python. In order to pick up insects from a Petri dish and transfer them into a well of a 96-well microplate, a suction hose with a pipette tip is positioned above the target insect by the transportation system. The hose is connected to an LA100 syringe pump (Landgraf Laborsysteme) that is also controlled by the Raspberry Pi. The sorting process is illustrated in Figure 2.

2.2 | Image acquisition system

The sorting system includes two cameras with different lenses: the overview camera (C1) and the specimen camera (C2). The overview camera is a Ximea MQ042CG-CM camera using a CK12M1628S11 lens (Lensation) with a focal length of 16 mm and an aperture of 2.8. It is positioned directly above the Petri dish to take a detailed overview image of the sample. This image is used for detecting insects and their position within the Petri dish (Figure 3a).
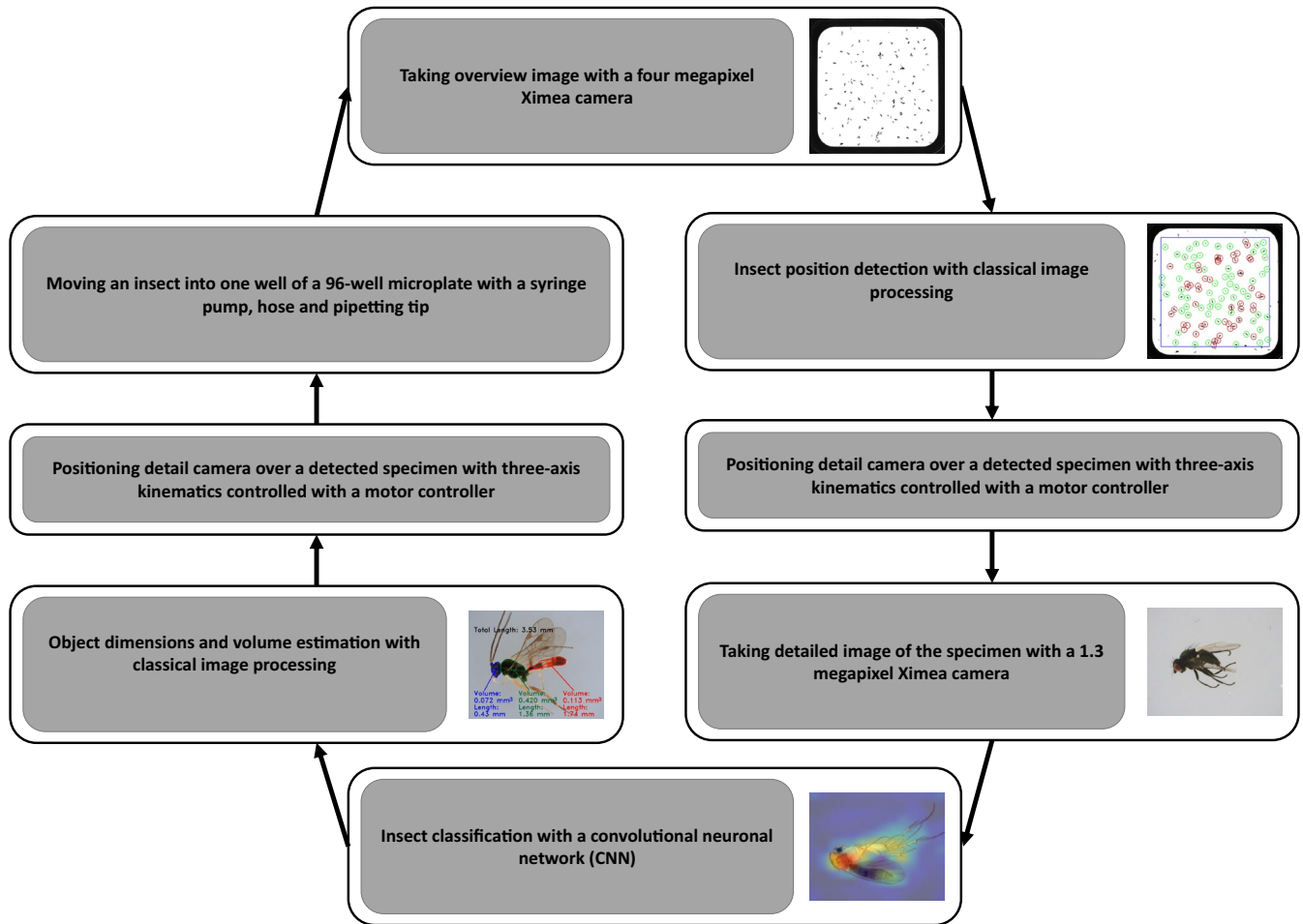
FIGURE 2 Eight-point process-chain for the sorting and classification process of the DiversityScanner. All relevant data are stored on a Raspberry Pi

The specimen camera (C2) is a Ximea MQ013CG-E2 using a telecentric Lensation TCST-10-40 lens with a magnification of 1×. This camera is moved by the *x*- and *y*- axes of the robot to a position above the insect to take a detailed image for the purpose of classification and measuring size (Figures 4 and 7).

2.3 | Image processing system

Three different software algorithms are used. The first algorithm determines the position of each object within the Petri dish, the second estimates the biomass of each insect and the third is an artificial neural network to classify insects into different classes.

2.3.1 | Determining position of insects

Most objects in a Malaise trap sample are insects, but bulk samples also include insect parts and debris. After the overview image is taken, several image processing operations are used to detect only insects that are suitable for processing: (i) a median filter removes noise from the image, (ii) the RGB image is converted to greyscale,

(iii) an adaptive threshold filter segregates the objects and (iv) a contour finder identifies the boundaries of all objects. Three conditions must be met for an object to be considered for imaging and transfer: (i) the size must be within a specified interval, (ii) the object has to be >10 mm away from the Petri dish edge (blue line in Figure 3b) and (iii) its distance to other objects must exceed a minimum threshold value set by the user. For efficient operation, only small specimens (body length <3 mm) should be placed into the Petri dish. Size presorting of whole samples can be manual or employ the efficient sieving methods described by Buffington and Gates (2013). Furthermore, it is desirable to distribute insects more or less evenly in the Petri dish because clumping reduces the number of insects that are available for sorting.

After detection, the coordinates of the objects are stored in a list and used to determine where the pipetting tip and the specimen camera should be for processing a specimen. After each insect is moved, a new overview image is taken to determine the new coordinates of the remaining objects. The process continues until no additional suitable objects remain or all wells of the 96-well microplate are filled. Based on the size of the currently used Petri dish (120 × 120 mm), ~150 (±10) can be sorted. Since the work of the robot is automatically stopped after a 96-well microplate is filled,
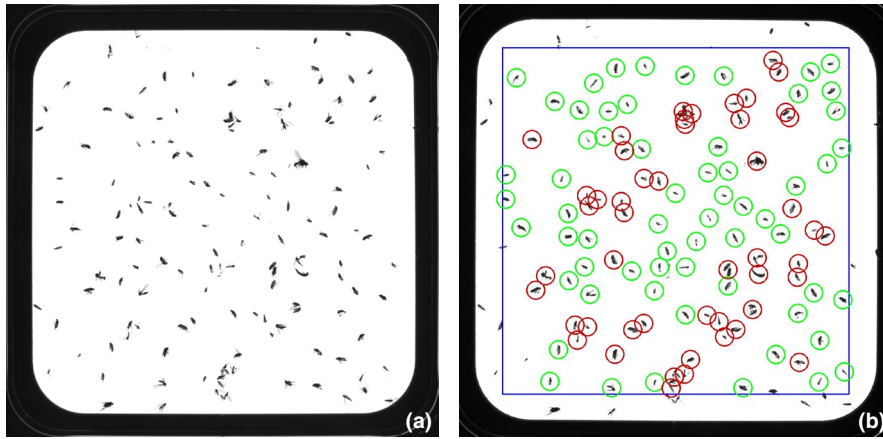
FIGURE 3 Overview of Petri dish with evenly distributed insects before (a) and after processing (b). (b) A region of interest has been defined (blue line 10 mm from the edge). Circles represent detected objects (green = meet size and distance conditions for imaging and movement; red = size too large and/or distance too small)
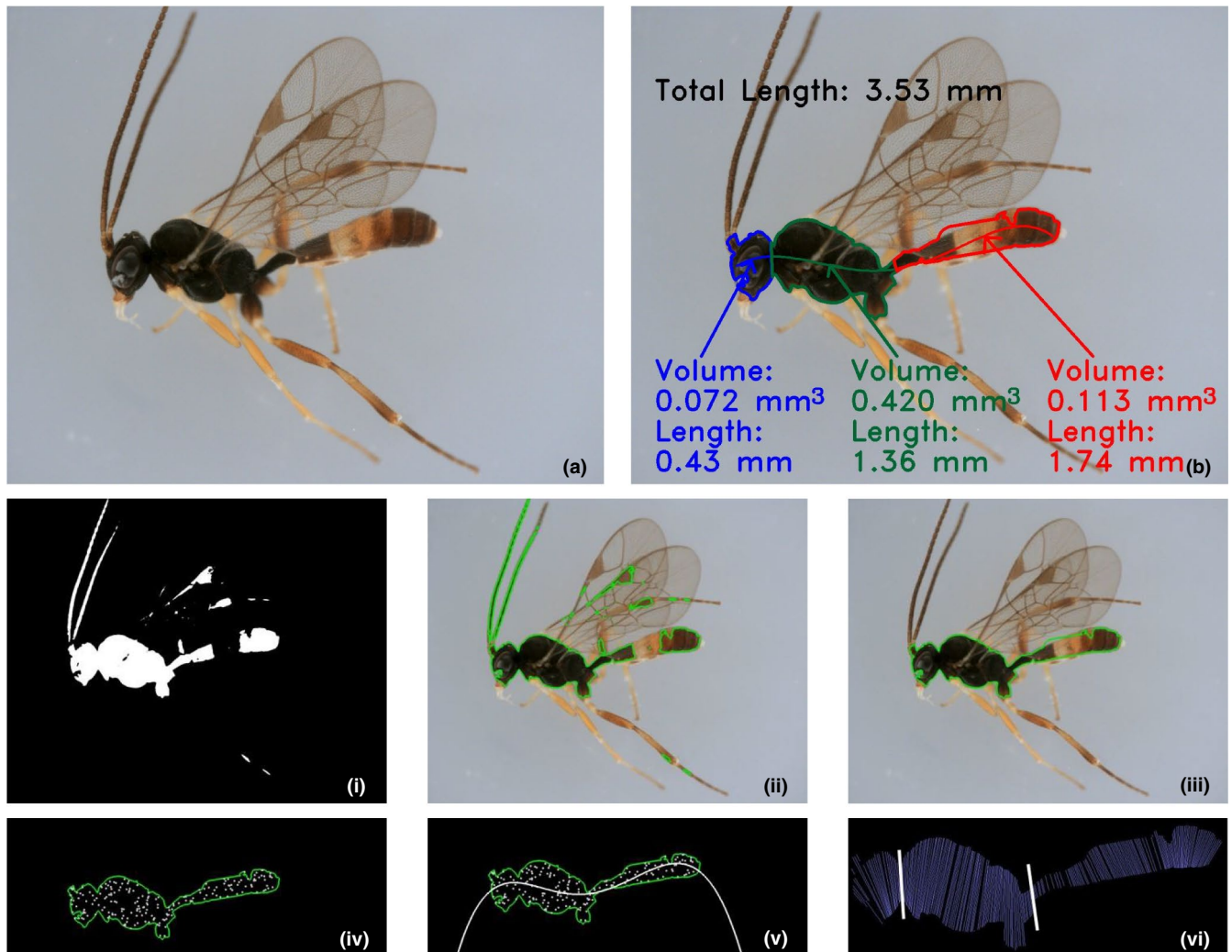

FIGURE 4 Specimen images obtained with the specimen camera before (a) and after processing (b). (i–iv) Image processing steps used to distinguish head, thorax/mesosoma and abdomen/metasoma. (i, ii) Contour determination; (iii) connecting surfaces; (iv) placing random points; (v) regression; (vi) defining dividing lines

new insects can also be added to the Petri dish during this work step. Given that only the Petri dish and the suction tube touch specimens, only these parts need cleaning between samples. The dish and suction tube can be cleaned with bleach or replaced.

2.3.2 | Biomass estimation

Several image processing operations are needed to measure the length and volume of each insect. First, the contour is determined using

morphological operators. Only those surfaces that have a minimum value are selected to avoid mistaking very thin structures (wings or debris) as belonging to the body. If more than one surface is found (e.g., two body parts of the same specimen separated by a light area), they are connected so that there is only one contour. Within this contour, points are placed randomly and then used to create a regression; therefore, a larger number of points increases the accuracy of the regression and the precision of the estimate. To find the dividing lines between body parts (i.e. head, thorax/mesosoma, abdomen/

metasoma), straight lines are placed at right angles to and along the regression line. Only those points of a line that lie within the contour are used. Subsequently, the dividing line between the head and thorax or between the thorax and abdomen is determined by examining the changes in length. Note that because not all species have a clear dividing line between the body tagmata, some dividing lines are set incorrectly and need to be adjusted manually before the total volume can be determined (see Results for details). To estimate the volume of a specimen, a straight line is drawn through each body part. Afterwards,

TABLE 1 Classes and number of images available for training, validation and testing to train the CNN to distinguish 14 insect classes and one "other" class

Class (Taxon)	Training	Validation	Testing	Total
Acalyptrate Diptera	377	69	148	594
Diptera Calyptratae	57	10	12	79
Diptera Cecidomyiidae	280	70	117	467
Diptera Chironomidae	140	20	32	192
Diptera Dolichopodidae	112	14	14	140
Diptera Empididae & Hybotidae	254	80	112	446
Diptera Mycetophilidae & Keroplatidae	251	79	110	440
Diptera Phoridae	461	167	209	837
Diptera Psychodidae	91	19	19	129
Diptera Sciaridae	219	54	90	363
Hemiptera Cicadellidae	102	14	21	137
Hymenoptera Braconidae	74	17	22	113
Hymenoptera Diapriidae	166	38	51	255
Hymenoptera Ichneumonidae	100	13	20	133
Other	498	113	147	758
Total	3182	777	1124	5083

FIGURE 5 Class activation maps for specimens belonging to four different insect classes. The warmer the colour, the more important is the region for classifying the insects (red = very important, blue: less important). (a) Hymenoptera Diapriidae: important areas are antennae, head, mesosoma and wing venation; (b) Diptera Calyptratae: important areas are head and eye; (c) Diptera Keroplatidae and Mycetophilidae: important areas are thorax and legs; (d) Diptera Psychodidae: important area is the wings
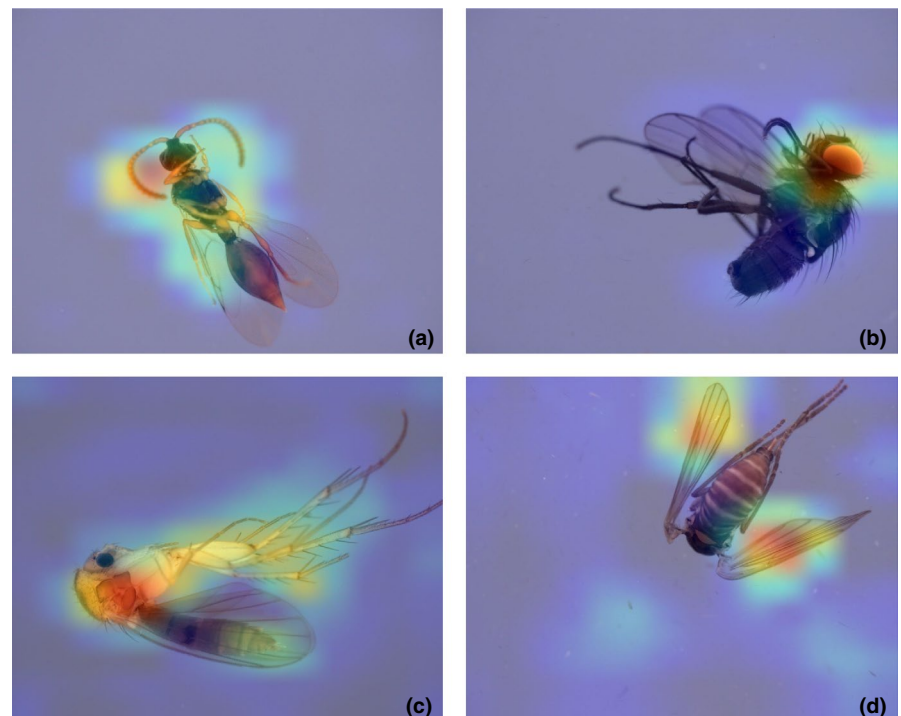
additional perpendicular straight lines are drawn which must be within the body contour. The distance and length of the straight lines is then used to determine the volume, one slice at a time. The lengths and volumes of the individual and its body parts are displayed on the screen of the sorting robot and the measurements are stored. Figure 4 shows an example of a detailed picture (a) before and (b) after the volume estimation, as well as the necessary steps (i–vi). All operations use the free OPENCV program library (version 4.5.1) and Python scripts (version 3.8.6). Currently, volume estimates perform best for body parts that are rotationally symmetrical; that is, the method works better for insects with rotationally symmetrical morphology (e.g., many Hymenoptera).

2.3.3 | Classification with artificial neural network

We apply machine-learning algorithms based on CNNs to assign insects to different classes. In our study, we used 5083 colour images split into 3182 for training (~62.5%), 777 for validation (~15%) and 1124 for testing (22.5%; Table 1). The images were obtained with the specimen camera for insects from five Malaise trap samples: three from Germany (near Rastatt, Kitzing and Framersbach) and two from Italy (Province of L'Aquila: Valle di Teve and Foresta Demaniale Chiarano-Sparvera). We used the abundances of the different taxa (usually family-level) to decide on how many classes (in the machine learning sense) could be covered because they had a sufficiently large number of training, validation and test images. We trained the CNN for 14 taxa and created a 15th class for the residual specimens ($N = 758$), which also includes images of body parts (mainly legs and wings). Data augmentation was performed to increase the number of images and the invariance within a class. The following processing operations were applied randomly: rotation, width shift, height shift, shear, zoom, horizontal flip and fill mode nearest.

We here used the VGG19 architecture as the base model for classification (Simonyan & Zisserman, 2014). The model is initialized with pretrained ImageNet weights and the last layer is removed. For the new classification layer, a global average pooling, a dense layer with 1024 units and a reLU-activation, and a linear layer with a dropout rate during training of 0.4 are added. For the final classification, a softmax and an
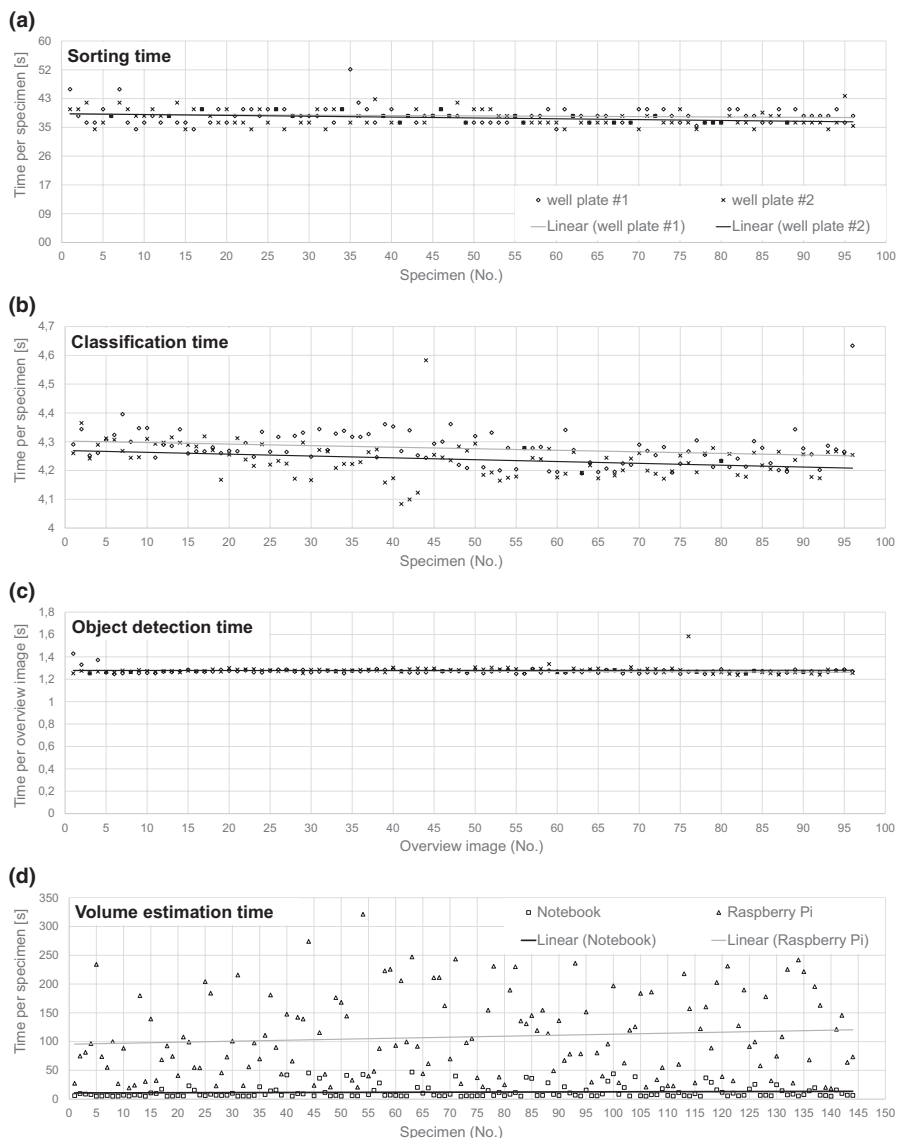

FIGURE 6 Specimen processing times. (a) Time per specimen for sorting based on two microplates; average times were 37 s for plate 1 and 38 s for plate 2. (b) Time per specimen for classification. Plate 1: 4.3 s; plate 2: 4.2 s. (c) Time for object detection on each overview image. A new overview image is needed for each sorted object. Average processing time for both plates was 1.3 s. (d) Time per specimen for volume estimation. Determined for 144 specimens of Hymenoptera Diapriidae on Raspberry Pi 4 (Model B; 4 GB) and notebook with Intel Core i7-4510 U with 2.0 GHz. Raspberry Pi runtime: 108 s; Notebook: 12 s. Note that symbol legend for (a) also applies to (b) and (c)

L2-regularization with a value of 0.02 are applied. The input size of each image is 224 × 224 pixels and the model has about 20.5 million parameters in total. The number of nodes in the last layer corresponds to the number of classes in the experiment. For training, the parameters of the original model are frozen and only the classification layer is trained. Afterwards, the whole model is optimized, whereby training is applied to all layers. Lastly, we use class activation maps to illustrate those features used by the neural network for specimen assignment. They were obtained by creating a global average pooling layer (Figure 5).

The model was implemented in KERAS (version 2.4.3) based on TENSORFLOW (version 2.2.1) and all experiments were conducted in the

"Hymenoptera Diapriidae" and "Hemiptera Cicadellidae," where all insects were correctly classified (100%), whereas insects of the class "Hymenoptera Ichneumonidae" had the lowest correct classification rate (75%). The performance details for the different classes are summarized in a "confusion matrix" (Table 3) that compares results of the "predicted" (CNN) identification with the "true" labels (taxonomists). Note, that the good performance of the CNN allowed for the implementation of taxon-specific processing. The robot then only sorts insects belonging to a predefined class.

Biomass estimation is the slowest process and performing

TABLE 2 Classification accuracy (predicted label = true label) for each of the 15 classes that can be distinguished by the DiversityScanner

Class (Taxon)	Result	Class (Taxon)	Result
Acalyptrate Diptera	91%	Diptera Psychodidae	89%
Diptera Calyptratae	83%	Diptera Sciaridae	92%
Diptera Cecidomyiidae	91%	Hemiptera Cicadellidae	100%
Diptera Chironomidae	97%	Hymenoptera Braconidae	82%
Diptera Dolichopodidae	86%	Hymenoptera Diapriidae	100%
Diptera Empididae & Hybotidae	87%	Hymenoptera Ichneumonidae	75%
Diptera Mycetophilidae & Keroplatidae	99%	Other	81%
Diptera Phoridae	97%	Overall result[a]	91.4%

[a]The number of images used is included in the calculation of the overall result, which is why it differs from the arithmetic mean of all individual results.

Python programming language (version 3.8.6). The networks were trained using the online tool COLABATORY using a single board computer (Nvidia). The working principles of the robot are illustrated in the following video clip: https://www.youtube.com/watch?v=ElJ5VSHa4OI.

3 | RESULTS

To test how fast the DiversityScanner sorts, we used 192 specimens (=two microplates). The average time per specimen was 37 s for the first and 38 s for the second plate, with some specimens taking much longer (e.g., #1, #8, #35: Figure 6a). The reported time consists of the time needed for activating the GUI, the write operations on the SD card, the movement time of the axes, the runtimes of the algorithms for object detection and classification, and the times for using the syringe pump. Faster sorting is feasible, but reduces quality because the specimens need to settle before high-quality images can be taken. In addition, the specimens have to "sink" within the pipette tip before they can be safely expulsed into a well of the microplate. In contrast, object recognition and classification are fast (Figure 6b,c). The average time for object detection is <1.3 s and classification ~4 s. Currently, the robot classifies the detected insects into 14 different classes. All other insects and noninsect objects are combined in a class labelled "other" (Table 2). The best classification results are obtained for

it during sorting adds significantly to processing times because the Raspberry Pi requires almost 2 min per specimen (Figure 6d: 108.06 s). We therefore recommend that the images be exported to another computer before the algorithm is applied. On a notebook with Intel Core i7-4510 U with 2.0 GHz, the average processing time per specimen is ~12 s. Note that total volume is estimated as the sum of the volumes for head, thorax and abdomen/metasoma and that our tests used only Hymenoptera Diapriidae because they have clearly separated body parts.

Currently, the sorting robot handles only insects up to 3 mm in length (Figure 7a–o), because larger insects do not fit through the pipetting tip. However, solutions for larger insects are in development. Lower-bound size limits for specimens do not exist, but very small specimens may not be detectable on the overview image.

4 | DISCUSSION

The use of CNNs for the identification of charismatic species is becoming routine (Fairbrass et al., 2019; Milošević et al., 2020; Stowell et al., 2019; Tabak et al., 2019). However, these methods have been largely unavailable for small invertebrates, even though they comprise much of the multicellular animal species diversity (Groombridge, 1992; Stork et al., 2015) and contribute many ecosystem services (Wagner, 2020). The main problem is the lack of

TABLE 3 Confusion matrix for 15 classes that can be distinguished by the DiversityScanner, including the "other" class. The true label is shown on the y-axis, the predicted label on the x-axis

True label	acalyptrate Dipt.	Dipt. Calyptratae	Dipt. Cecidomyiidae	Dipt. Chironomidae	Dipt. Dolichopodidae	Dipt. Empididae & Hybotidae	Dipt. Mycetophilidae & Keroplatidae	Dipt. Phoridae	Dipt. Psychodidae	Dipt. Sciaridae	Hemiptera Cicadellidae	Hyenoptera Braconidae	Hymenoptera Diapriidae	Hymenoptera Ichneumonidae	Other
acalyptrate Dipt.	0.912	0.020	0.000	0.000	0.014	0.014	0.000	0.007	0.000	0.000	0.007	0.000	0.000	0.000	0.027
Dipt. Calyptratae	0.083	0.833	0.000	0.000	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Dipt. Cecidomyiidae	0.000	0.000	0.915	0.017	0.000	0.000	0.000	0.000	0.009	0.051	0.000	0.000	0.000	0.000	0.009
Dipt. Chironomidae	0.000	0.000	0.031	0.939	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Dipt. Dolichopodidae	0.143	0.000	0.000	0.000	0.857	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Dipt. Empididae & Hybotidae	0.027	0.000	0.000	0.000	0.000	0.866	0.009	0.009	0.000	0.018	0.009	0.018	0.000	0.000	0.045
Dipt. Mycetophilidae & Keroplatidae	0.000	0.000	0.000	0.000	0.000	0.000	0.991	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Dipt. Phoridae	0.014	0.000	0.000	0.005	0.000	0.000	0.005	0.967	0.005	0.005	0.000	0.000	0.000	0.000	0.000
Dipt. Psychodidae	0.053	0.000	0.000	0.053	0.000	0.000	0.000	0.000	0.895	0.000	0.000	0.000	0.000	0.000	0.000
Dipt. Sciaridae	0.000	0.000	0.056	0.022	0.000	0.000	0.000	0.000	0.000	0.922	0.000	0.000	0.000	0.000	0.000
Hemiptera Cicadellidae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
Hymenoptera Braconidae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.818	0.000	0.091	0.091
Hymenoptera Diapriidae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
Hymenoptera Ichneumonidae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.150	0.000	0.750	0.100
Other	0.034	0.000	0.000	0.027	0.000	0.034	0.000	0.007	0.020	0.014	0.020	0.020	0.014	0.000	0.810

trained CNNs, which cannot be obtained without first producing sets of training images for thousands of species. We believe that the best strategy for obtaining these sets is combining automated specimen imaging with DNA barcoding. Each DiversityScanner can image 1000 specimens per day so that a laboratory equipped with a few DiversityScanners will be able to process several full Malaise trap samples per day. Each contains thousands of specimens that can be imaged with minimal manual labour. After imaging, the specimens are automatically transferred to microplates for DNA barcoding. Once barcoded, the images can be relabelled using the taxonomic information obtained from DNA barcodes. This can produce image training sets that have approximately species-level resolution given that specimen sorting with DNA barcodes yields MOTUs that are mostly congruent with morphospecies even when rigorously assessed based on barcoding thousands of specimens (90%: Wang et al., 2018; Yeo et al., 2018). Some of the MOTUs will represent described species so that the training image sets can even be labelled with scientific names. This requires that the MOTUs are matched to a scientific name via DNA barcodes from a barcode database or through identifying specimens. Common species, genera and families can rapidly acquire sufficiently large numbers of training images. Indeed, for the most common 14 classes of insects in Malaise traps, we already had a sufficiently large number of images (5083) for creating such networks after partially imaging only five Malaise trap samples.

An additional useful feature of the DiversityScanner is that it can be instructed to only transfer a limited number of specimens for particularly abundant taxa. For example, the robot can be programmed to move only one or two microplates' worth of nonbiting midges (Chironomidae) if this taxon is too abundant for complete treatment. This ability to only find and move some taxa helps with implementing clade-specific molecular recipes (e.g., different DNA extraction or PCR recipes for taxa that are difficult to barcode: e.g., Hymenoptera) or restricting barcoding to either males or females given that often only one sex has species-specific morphological differences (Eberhard, 2010). Overall, we would thus predict that the DiversityScanner will prove useful for many studies using the toolkit of molecular ecology. The robot can rapidly generate barcodes for an unknown fauna, which helps with improving the quality of barcode databases and the interpretation of metabarcoding data. The robot can also prepare large numbers of specimens for molecular work on microbiomes or species interactions that have been sorted semi-automatically to the species level. By facilitating the barcoding of all specimens, the DiversityScanner furthermore highlights which common species belonging to dark taxa should be prioritized for taxonomic treatment.

One of the unresolved issues is whether CNNs will be sufficiently powerful to yield species-level identifications for closely related species (but see Ärje et al., 2020b; Knyshov et al., 2021). It is likely that the main limitation will be the number, quality and orientation of training images. Figure 7 illustrates the latter problem. Insects are imaged from many different angles and each will require enough training images before the CNNs will have a realistic chance for

achieving accuracy at high taxonomic resolution. One solution for this problem is imaging specimens in many orientations. Fortunately, this is now feasible because modern, high-quality cameras can acquire large numbers of images at different magnifications and orientations. This is particularly straightforward once specimens have been presorted to putative species based on DNA barcodes. As illustrated by the BIODISCOVER robot, inserting these specimens into a cuvette allows imaging from many sides. This is why we predict that once large numbers of species have been extensively imaged and included in CNNs, robots such as the DiversityScanner should be able to identify many specimens to species based on images only. Note also that not all would be lost if CNNs were eventually found to be incapable of distinguishing closely related species. Specimens identified to genus- or species-group level would still be suitable for many biomonitoring purposes.

Eventually, DNA barcoding might become restricted to those specimens that are not identifiable based on visual information; that is, the DiversityScanner would learn how to sort specimen to species, but also learn how to identify those specimens that still require barcoding. This will make the robot a powerful tool for discovering rare or new species in large samples. This ability would be particularly important in the 21st century, given that new species continue to arrive at well-characterized sampling sites (Parmesan, 2006). These new arrivals are due to both distribution shifts in response to climate change (Fartmann et al., 2021; Wilson et al., 2007) and anthropogenic introductions (Bertelsmeier, 2021). It would be desirable for both to have an early-alert system based on automated workflows.

With regard to the classification accuracy rates of our current CNN, we observe only a very weak correlation between the number of training images, morphological heterogeneity and classification accuracy (Figure S1). There are classes with large numbers of training images that perform better than classes with lower numbers (e.g., "Diptera Calyptratae," 57 training images: 83% vs. "Diptera Phoridae," 64 training images: 97%), but the better performance of "Diptera Phoridae" could also be due to higher morphological uniformity. However, this is not in line with the observation of a comparatively high classification accuracy that was obtained for the class "other" that has the highest morphological heterogeneity. Indeed, this class performed better (81%) than "Hymenoptera Ichneumonidae" (75%, Table 3). Clearly, more data are needed to understand the training needs of CNNs for insects.

This first version of the DiversityScanner still struggles with several aspects of complex Malaise trap samples. Two specimens lying on top of each other may be erroneously recognized as one insect based on the overview image. This may lead to the transfer of several specimens into one well. Future versions of the scanner will have improved object detection based on the overview image. In addition, object detection can also be applied to detailed specimen images in order to detect cases where the image contains more than one specimen. The latter would also avoid instances where an insect was detected but not picked up by the pipette so that a well of the microplate remains empty. Particularly high on the list of development needs is also the handling of specimens larger than 3 mm.

FIGURE 7 Sample images for the 15 classes that can be distinguished by the CNN. (a) Acalyptrate Diptera; (b) Diptera Calyptratae; (c) Diptera Cecidomyiidae; (d) Diptera Chironomidae; (e) Diptera Dolichopodidae; (f) Diptera Empididae & Hybotidae; (g) Diptera Keroplatidae & Mycetophilidae; (h) Diptera Phoridae; (i) Diptera Psychodidae; (j) Diptera Sciaridae; (k) Hemiptera Cicadellidae; (l) Hymenoptera Braconidae; (m) Hymenoptera Diapriidae; (n) Hymenoptera Ichneumonidae; (o) other insects (e.g., Hemiptera Aphididae); (p) other objects

They can be accommodated by increasing the suction tip diameter or introducing a gripper with a sensor-based feedback system. Note that these changes could be implemented readily within the existing object detection algorithms because only the object detection parameters would have to be changed.

High-quality imaging of samples containing a mixture of small and large specimens requires several lenses with different focal lengths. The only alternative is image cropping. This approach is favoured by the BIODISCOVER robot (Ärje et al., 2020a) or the live moths photographing system of Bjerge et al. (2021). Large images are taken from which the areas with insects are cropped. This causes the resolution of the individual images to be comparatively low (BIODISCOVER: 496 × 496 pixels; Moth light trap: average size: 368 × 353 pixels). Instead, the DiversityScanner moves the specimen camera over the insects, which can then be photographed at high resolution (1280 × 1024 pixels). For the classification using neural networks, low resolution would be sufficient, but for determining

the volume as well as for taxonomic work on the specimens high resolution is needed.

Overall, we believe that robots like the DiversityScanner can solve many of the problems that Robert May mentioned when he bemoaned our lack of biodiversity knowledge. Automation can expedite biodiversity discovery and monitoring of neglected taxa. However, it will be important to keep the design of the DiversityScanner simple, low-cost (currently<5000 €) and open access. This will render it feasible to have many robots running in parallel and all around the world. Imagine 100 robots processing 1000 specimens a day for 100 days/year. They would generate 10 million imaged and sorted specimens every year. Many new species would be discovered and imaged. The data and images would be an important scientific and educational resource. Of course, the newly discovered species would still need description and we would still know very little about the ecological roles that these species play, but molecular approaches to species interaction research, diet analysis and life history stage matching can

help fill these gaps (e.g., Reeves et al., 2018; Six, 2013; Srivathsan et al., 2019b; Yeo et al., 2018).

AUTHOR CONTRIBUTIONS

Conceptualization: R.M., T.v.R., L.W. and C.P.; writing original draft preparation: L.W., R.M. and M.G.; writing review and editing: C.P., R.M., S.S., P.C., M.B. and T.v.R.; visualization: L.W., M.G. and S.S.; supervision: C.P., R.M. and T.v.R.; funding acquisition: C.P., T.v.R. and R.M. All authors have read and agreed to the published version of the manuscript.

DATA AVAILABILITY STATEMENT

All image data used for training and testing are accessible at the media repository of the Museum für Naturkunde Berlin: https://doi.org/10.7479/4tbx-qm72. All files for printing the robot parts and the software code are available from the repository of the Open Science Framework: https://osf.io/en594/. Benefits from this research accrue from the sharing of our data, software and robot assembly information on the public databases as described above.

ORCID

Lorenz Wührl https://orcid.org/0000-0002-0734-6093
Christian Pylatiuk https://orcid.org/0000-0002-3507-7134
Thomas von Rintelen https://orcid.org/0000-0002-6253-3078
Michael Balke https://orcid.org/0000-0002-3773-6586
Stefan Schmidt https://orcid.org/0000-0001-5751-8706
Pierfilippo Cerretti https://orcid.org/0000-0002-9204-3352
Rudolf Meier https://orcid.org/0000-0002-4452-2885

REFERENCES

Lepek, H., Nave, T., Fleischmann, Y., Eisenberg, R., Karlin, B. E., & Tirosh, I. (2020). Method for sex sorting of mosquitoes and apparatus therefore: US Patent (16/479,648).

Ärje, J., Melvad, C., Jeppesen, M. R., Madsen, S. A., Raitoharju, J., Rasmussen, M. S., Iosifidis, A., Tirronen, V., Gabbouj, M., Meissner, K., & Høye, T. T. (2020a). Automatic image-based identification and biomass estimation of invertebrates. *Methods in Ecology and Evolution*, 11, 922–931. https://doi.org/10.1111/2041-210X.13428

Ärje, J., Raitoharju, J., Iosifidis, A., Tirronen, V., Meissner, K., Gabbouj, M., Kiranyaz, S., & Kärkkäinen, S. (2020). Human experts vs. machines in taxa recognition. *Signal Processing: Image Communication*, 87, 115917.

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 6506–6511. https://doi.org/10.1073/pnas.1711842115

Bertelsmeier, C. (2021). Globalization and the anthropogenic spread of invasive social insects. *Current Opinion in Insect Science*, 46, 16–23. https://doi.org/10.1016/j.cois.2021.01.006

Bjerge, K., Nielsen, J. B., Sepstrup, M. V., Helsing-Nielsen, F., & Høye, T. T. (2021). An automated light trap to monitor moths (Lepidoptera) using computer vision-based tracking and deep learning. *Sensors (Basel, Switzerland)*, 21, 343. https://doi.org/10.3390/s21020343

Borkent, A., & Brown, B. V. (2015). How to inventory tropical flies (Diptera)–One of the megadiverse orders of insects. *Zootaxa*, 3949, 301–322. https://doi.org/10.11646/zootaxa.3949.3.1

Brown, B. V. (2005). Malaise Trap Catches and the Crisis in Neotropical Dipterology. *American Entomologist*, 51, 180–183. https://doi.org/10.1093/ae/51.3.180

Brown, B. V., Borkent, A., Adler, P. H., Amorim, D. D. S., Barber, K., Bickel, D., Boucher, S., Brooks, S. E., Burger, J., Burington, Z. L., Capellari, R. S., Costa, D. N. R., Cumming, J. M., Curler, G., Dick, C. W., Epler, J. H., Fisher, E., Gaimari, S. D., Gelhaus, J., … Zumbado, M. A. (2018). Comprehensive inventory of true flies (Diptera) at a tropical site. *Communications Biology*, 1, 21. https://doi.org/10.1038/s42003-018-0022-x

Buchner, D., Macher, T.-H., Beermann, A. J., Werner, M.-T., & Leese, F. (2021). Standardized high-throughput biomonitoring using DNA metabarcoding: Strategies for the adoption of automated liquid handlers. *Environmental Science and Ecotechnology*, 8, 100122. https://doi.org/10.1016/j.ese.2021.100122

Buffington, M., & Gates, M. (2013). The Fractionator: A simple tool for mining 'Black Gold'. Skaphion.

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E6089–E6096. https://doi.org/10.1073/pnas.1704949114

Chamblin, M. A., Paasch, R. K., Lytle, D. A., Moldenke, A. R., Shapiro, L. G., & Dietterich, T. G. (2011). Design of an automated system for imaging and sorting soil mesofauna. *Biological Engineering Transactions*, 4, 17–41. https://doi.org/10.13031/2013.37174

Creedy, T. J., Ng, W. S., & Vogler, A. P. (2019). Toward accurate species-level metabarcoding of arthropod communities from the tropical forest canopy. *Ecology and Evolution*, 9, 3105–3116. https://doi.org/10.1002/ece3.4839

Eberhard, W. G. (2010). Rapid divergent evolution of genitalia. In: *The evolution of primary sexual characters in animals* (pp. 40–78). Oxford University Press.

Fairbrass, A. J., Firman, M., Williams, C., Brostow, G. J., Titheridge, H., & Jones, K. E. (2019). CityNet—Deep learning tools for urban eco-acoustic assessment. *Methods in Ecology and Evolution*, 10, 186–197. https://doi.org/10.1111/2041-210X.13114

Fartmann, T., Poniatowski, D., & Holtmann, L. (2021). Habitat availability and climate warming drive changes in the distribution of grassland grasshoppers. *Agriculture, Ecosystems & Environment*, 320, 107565. https://doi.org/10.1016/j.agee.2021.107565

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

Groombridge, B. (1992). *Global biodiversity status of the Earth's living resources*, No. 333.95 G562gl. World Conservation Monitoring Centre.

Hartop, E., Srivathsan, A., Ronquist, F., & Meier, R. (2021) Large-scale Integrative Taxonomy (LIT): Resolving the data conundrum for dark taxa.

Hebert, P. D. N., Braukmann, T. W. A., Prosser, S. W. J., Ratnasingham, S., deWaard, J. R., Ivanova, N. V., Janzen, D. H., Hallwachs, W., Naik, S., Sones, J. E., & Zakharov, E. V. (2018). A Sequel to Sanger: Amplicon sequencing that scales. *BMC Genomics*, 19, 219. https://doi.org/10.1186/s12864-018-4611-3

Karlsson, D., Forshage, M., Holston, K., & Ronquist, F. (2020a). The data of the Swedish Malaise Trap Project, a countrywide inventory of Sweden's insect fauna. *Biodiversity Data Journal*, 8, e56286. https://doi.org/10.3897/BDJ.8.e56286

Karlsson, D., Hartop, E., Forshage, M., Jaschhof, M., & Ronquist, F. (2020b). The Swedish Malaise Trap Project: A 15 Year Retrospective on a Countrywide Insect Inventory. *Biodiversity Data Journal*, 8, e47255.

Kim, M. (2014). Mealworm sorting unit and sorting apparatus. Korean Patent Number, KR101464734B1.

Knyshov, A., Hoang, S., & Weirauch, C. (2021). Pretrained convolutional neural networks perform well in a challenging test case: Identification of plant bugs (Hemiptera: Miridae) using a small number of training images. *Insect Systematics and Diversity*, 5, 3. https://doi.org/10.1093/isd/ixab004

Sarpola, M. J., Paasch, R. K., Mortensen, E. N., Dietterich, T. G., Lytle, D. A., Moldenke, A. R., & Shapiro, L. G. (2008). An aquatic insect imaging system to automate insect classification. *Transactions of the ASABE*, 51, 2217–2225. https://doi.org/10.13031/2013.25375

May, R. M. (2011). Why worry about how many species and their loss? *PLoS Biology*, 9, e1001130. https://doi.org/10.1371/journal.pbio.1001130

Milošević, D., Milosavljević, A., Predić, B., Medeiros, A. S., Savić-Zdravković, D., Stojković Piperac, M., Kostić, T., Spasić, F., & Leese, F. (2020). Application of deep learning in aquatic bioassessment: Towards automated identification of non-biting midges. *The Science of the Total Environment*, 711, 135160. https://doi.org/10.1016/j.scitotenv.2019.135160

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, 37, 637–669. https://doi.org/10.1146/annurev.ecolsys.37.091305.110100

Perre, P., Faria, F. A., Jorge, L. R., Rocha, A., Torres, R. S., Souza-Filho, M. F., Lewinsohn, T. M., & Zucchi, R. A. (2016). Toward an automated identification of Anastrepha fruit flies in the fraterculus group (Diptera, Tephritidae). *Neotropical Entomology*, 45, 554–558. https://doi.org/10.1007/s13744-016-0403-0

Pfriem, A., Pylatiuk, C., Alshut, R., Ziegener, B., Schulz, S., & Bretthauer, G. (2012). A modular, low-cost robot for zebrafish handling. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. *Annual International Conference*, 2012, 980–983.

Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L. A., Barrio-Amorós, C. L., Salazar-Valenzuela, D., & Prost, S. (2018). Real-time DNA barcoding in a rainforest using nanopore sequencing: Opportunities for rapid biodiversity assessments and local capacity building. *GigaScience*, 7, giy033. https://doi.org/10.1093/gigascience/giy033

Reeves, L. E., Gillett-Kaufman, J. L., Kawahara, A. Y., & Kaufman, P. E. (2018). Barcoding blood meals: New vertebrate specific primer sets for assigning taxonomic identities to host DNA from mosquito blood meals. *PLoS Neglected Tropical Diseases*, 12, e0006767. https://doi.org/10.1371/journal.pntd.0006767

Simonyan, K., & Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556v6..

Six, D. L. (2013). The Bark Beetle Holobiont: Why microbes matter. *Journal of Chemical Ecology*, 39, 989–1002. https://doi.org/10.1007/s10886-013-0318-8

Srivathsan, A., Hartop, E., Puniamoorthy, J., Lee, W. T., Kutty, S. N., Kurina, O., & Meier, R. (2019a). Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biology*, 17, 96. https://doi.org/10.1186/s12915-019-0706-9

Srivathsan, A., Lee, L., Katoh, K., Hartop, E., Kutty, S. N., Wong, J., Yeo, D., & Meier, R. (2021). ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. *BMC Biology*, 19, 217. https://doi.org/10.1186/s12915-021-01141-x

Srivathsan, A., Nagarajan, N., & Meier, R. (2019b). Boosting natural history research via metagenomic clean-up of crowdsourced feces. *PLoS Biology*, 17, e3000517. https://doi.org/10.1371/journal.pbio.3000517

Stork, N. E., McBroom, J., Gely, C., & Hamilton, A. J. (2015). New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proceedings of the National Academy of Sciences*, 112, 7519–7523. https://doi.org/10.1073/pnas.1502408112

Stowell, D., Wood, M. D., Pamuła, H., Stylianou, Y., & Glotin, H. (2019). Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge. *Methods in Ecology and Evolution*, 10, 368–380. https://doi.org/10.1111/2041-210X.13103

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., Halseth, J. M., Di Salvo, P. A., Lewis, J. S., White, M. D., Teton, B., Beasley, J. C., Schlichting, P. E., Boughton, R. K., Wight, B., Newkirk, E. S., Ivan, J. S., Odell, E. A., Brook, R. K., … Miller, R. S. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, 10, 585–590. https://doi.org/10.1111/2041-210X.13120

Wagner, D. L. (2020). Insect declines in the anthropocene. *Annual Review of Entomology*, 65, 457–480. https://doi.org/10.1146/annurev-ento-011019-025151

Wang, W. Y., Srivathsan, A., Foo, M., Yamane, S. K., & Meier, R. (2018). Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: Validating a reverse workflow for specimen processing. *Molecular Ecology Resources*, 18, 490–501. https://doi.org/10.1111/1755-0998.12751

Watsa, M., Erkenswick, G. A., Pomerantz, A., & Prost, S. (2020). Portable sequencing as a teaching tool in conservation and biodiversity research. *PLoS Biology*, 18, e3000667. https://doi.org/10.1371/journal.pbio.3000667

Wilson, R. J., Gutiérrez, D., Gutiérrez, J., & Monserrat, V. J. (2007). An elevational shift in butterfly species richness and composition accompanying recent climate change. *Global Change Biology*, 13, 1873–1887. https://doi.org/10.1111/j.1365-2486.2007.01418.x

World Economic Forum's Global Risk Initiative. (2020) The global risks report 2020. Available from: http://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf. Accessed September 27, 2021.

Yeo, D., Puniamoorthy, J., Ngiam, R. W. J., & Meier, R. (2018). Towards holomorphology in entomology: Rapid and cost-effective adult-larva matching using NGS barcodes. *Systematic Entomology*, 43, 678–691. https://doi.org/10.1111/syen.12296

SUPPORTING INFORMATION

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