RESEARCH ARTICLE

Automated software for counting and measuring *Hyalella* genus using artificial intelligence

Ludy Pineda-Alarcón[1] · Maycol Zuluaga[2] · Santiago Ruíz[2] · David Fernandez Mc Cann[2] · Fabio Vélez[3] · Nestor Aguirre[3] · Yarin Puerta[3] · Julio Cañón[1]

Abstract

Amphipods belonging to the *Hyalella* genus are macroinvertebrates that inhabit aquatic environments. They are of particular interest in areas such as limnology and ecotoxicology, where data on the number of *Hyalella* individuals and their allometric measurements are used to assess the environmental dynamics of aquatic ecosystems. In this study, we introduce HyACS, a software tool that uses a model developed with the YOLOv3's architecture to detect individuals, and digital image processing techniques to extract morphological metrics of the *Hyalella* genus. The software detects body metrics of length, arc length, maximum width, eccentricity, perimeter, and area of *Hyalella* individuals, using basic imaging capture equipment. The performance metrics indicate that the model developed can achieve high prediction levels, with an accuracy above 90% for the correct identification of individuals. It can perform up to four times faster than traditional visual counting methods and provide precise morphological measurements of *Hyalella* individuals, which may improve further studies of the species populations and enhance their use as bioindicators of water quality.

Introduction

Hyalella (Smith, 1874) is a crustacean of the Amphipod order that belongs to the ecological group of freshwater aquatic macroinvertebrates (Correa-araneda and Contreras 2010). Amphipods are an important link in the trophic chain that connects primary producers with higher trophic levels, thus constituting key pieces in the energetic transfer as prey for several depredators. (Arfianti and Costello 2020). This benthic organism lives in association with submerged aquatic vegetation (Bastos-Pereira and Bueno 2016; Reis et al. 2020) and reacts to the presence of pollutants in aquatic ecosystems (Amiard-Triquet and Berthet 2015). Therefore, *Hyalella* can serve as an

✉ Ludy Pineda-Alarcón
ludy.pineda@udea.edu.co

Maycol Zuluaga
maycol.zuluaga@udea.edu.co

Santiago Ruíz
santiago.ruiz3@udea.edu.co

David Fernandez Mc Cann
david.fernandez@udea.edu.co

Fabio Vélez
fabio.velez@udea.edu.co

Nestor Aguirre
nestor.aguirre@udea.edu.co

Yarin Puerta
yarin.puerta@udea.edu.co

Julio Cañón
julio.canon@udea.edu.co

[1] Environmental Management and Modeling Group (GAIA), Environmental School, Engineer Faculty, Universidad de Antioquia, Medellín, Colombia

[2] Power Electronics, Automation, and Robotics Group (GEPAR), Engineer Electronic, Engineer Faculty, Universidad de Antioquia, Medellín, Colombia

[3] Limnology and Environmental Modeling Group (GEOLIMNA), Environmental School, Engineer Faculty, Universidad de Antioquia, Medellín, Colombia

indicator species that responds to changes in the quality of water and sediments. The individual response of *Hyalella* can be used to assess potential damage caused by different environmental stressors (chemical or physical), while changes in their population dynamics are of interest for monitoring aquatic ecosystems (Correa-araneda and Contreras 2010). The *Hyalella* genus is used as a bioindicator of high organic matter contents in water (Posada et al. 2000). Likewise, within the group of aquatic macroinvertebrates, amphipods, and particularly *hyalellas*, due to their narrow range of tolerance, are sensitive to environmental changes. Therefore, studies that involving the periodic counting of *Hyalellas* can be used for toxicology monitoring to asses influence of pesticides on the population growth (Beristain-Castillo et al. 2023) and as a biological model in ecotoxicological studies, given their high sensitivity to heavy metals (Colla and César 2019). Likewise, *Hyalella's* high population density is related to fluctuations in electrical conductivity, temperature, and low concentrations of dissolved oxygen (Jacobsen and Marín 2008). These conditions could generate changes in dominance, diversity, reproductive strategies, and their relationship with physical and chemical environmental variables (Roldán-Pérez 2016; Stepanian et al. 2020), which determine the species permanence and survival strategies as a function of the quality of their environment (Nnoli et al. 2019; Zipkin et al. 2017). Other research identifies changes in the distribution of organisms that can represent a threat to the species of *Hyalella* genus, depending on the number of individuals and morphological size modifications (Zepon et al. 2021). The quantification of *Hyalella* organisms also helps to identify space–time trends at biological, ecological, and environmental levels (Bastos et al. 2021). These metrics are important in the description of new species, as described in the work of Rocha Penoni et al. (2021) which requires measuring the body length for species identification within *Hyalella* genus.

Currently, computer vision and digital image processing techniques are playing a relevant role in ecology and environmental engineering studies. These techniques facilitate various practical tasks in insect research, that involve observation, quantification, classification, measurement, and other data that can be inferred from images (Weinstein 2018).

Some examples of the applications of these techniques in insects and aquatic invertebrates include the application of digital imaging processing techniques to count fruit flies (Yati and Dey 2011) which is a fast, automatic, and simple process with a low error rate (1 to 4%); automatically classifying benthic macroinvertebrates into fine-grained categories using convolutional neural networks (CNN) such as MatConv, which has produced a new benchmark database with 64 categories of macroinvertebrates (Raitoharju et al. 2018), and automatically measuring the density of the

copepod *Acartia tonsa,* which enables monitoring of metrics such as hatching rate, mortality, development rate, and individual biomass (Alver et al. 2011).

Sorting, identification, morphological and biological metrics, and biomass estimation of terrestrial and aquatic organisms can be accomplished through image identification using machine and deep learning techniques. For example, machine learning and deep learning have been used for the automatic sorting and extraction of metrics such as body length and volume from 14 common terrestrial invertebrate specimens, with an average precision of 91.4% for all taxa (Wührl et al. 2022). Furthermore, studies about image-based identification of terrestrial invertebrates using machine and deep learning could lead to improved understanding of these organisms (Ärje et al. 2020).

Other examples of the use of images in aquatic organisms include the extraction of morphological descriptors from in situ images of the Arctic zooplankton community, which has allowed for a better understanding of ecological patterns in the Arctic melt zone (Vilgrain et al. 2021). In addition, the movement patterns of *Oryzias latipes* were identified and measured after treatment with diazinon, using two-dimensional fast Fourier transform to efficiently calculate the differences before and after treatment (Park et al. 2005); Finally, real-time image analysis using digital image processing by binary image sequence was used to determine the movements of *Daphnia magna* in response to copper effects (Untersteiner et al. 2003);

Some applications, for instance ImageJ, which have a user interface, provide an accurate and flexible method for automating the analysis of digital photographs of laboratory microcosms. These applications can detect, count, and measure organisms, such as collembola, ants, nematodes, and daphnias, moving on a fixed but heterogeneous substrate (Mallard et al. 2013) or detect changes in the color of *Ulva pertusa* macroalgae (Lee et al. 2020);

Image processing in ecology is being used at different scales and process. For example, it is used to extract the mass of migratory insects based on an ellipsoidal dispersion model using radar data (Kong et al. 2019). It is also used for the segmentation and counting of pollen grains in microscopic images (Johnsrud et al. 2013); as well as for the processing and classification of four groups of aquatic macroinvertebrates (*Thraulodes, Traverella, Anacroneuria, and Smicridea*) through digital images processing for recognition at the genus taxonomic level with a 97.1% success rate (Serna López et al. 2020). Additionally, image processing, computer vision, and machine learning techniques are used for the registration and classification of *Rhopalosiphum padi* aphids to assess and predict crop damage through the use of Aphid CV software (Lins et al. 2020).

While basic image processing technics can solve a wide variety of environmental problems, some cases require more

advanced tools related to pattern recognition. One such tool is YOLO (You Only Look Once) (Redmon et al. 2016) which is used to identify and locate objects in an image. YOLO consists of a convolutional neural network (CNN) based on deep learning that allows detecting and tracking down objects in an image, facilitating the quantification and obtaining of multiple metrics in images related to biological samples. According to previous work, when comparing YOLO with other object detection tools, it is much faster than other similar systems, since it applies a single CNN for the dual purpose of object classification and localization (Chen et al. 2022), which makes it ideal for real-time applications. Furthermore, YOLO has a higher accuracy and also is able to detect smaller objects than other systems (Wang et al. 2018). This is particularly useful in biological applications, such as detecting objects in microscopic images. Overall, YOLO's combination of speed, accuracy, and small object detection capabilities make it a top choice for many developers and researchers in the computer vision field. Additionally, YOLO outperforms other object detectors like Faster R-CNN in terms of accuracy, speed, efficiency, and it is simpler to construct and can be trained directly on full images (Joiya 2022).

These image analysis tools contribute to the improvement of techniques for both counting individuals and quantifying the body measurements of specimens in research related to ecology and water. For instance, Zhong et al. (2018) developed a fast and accurate method of capturing, detecting, and counting flying insects based on YOLO object detector and support vector machines (SVM). Similarly, Kvæstad et al. (2022) developed a method to make morphological measurements of larvae images of *Gadus morhua* fish using deep learning with mask R-CNN neural net architecture.

The study of *Hyalella* populations requires the collection, separation, counting, and measurement of the organisms (Bastos-Pereira and Bueno 2016). There are widely established sampling protocols for collecting *Hyalella* (Rice et al. 2012; Ntislidou et al. 2021). However, the traditional method of separation and counting is performed manually under a stereomicroscope, which is subject to human errors of perception and lack of expertise. Counting individuals is a time-consuming activity that increases the likelihood of errors, due to fatigue and visual illusions (Lins et al. 2020).

Given that extracting information from images is quick, objectively verifiable, and less prone to observational errors (Pech et al. 2004), it is worth investing efforts in development tools that complement and improve traditional methods of information extraction. Manoukis et al. (2019) suggests that using techniques based on digital image processing could improve the efficiency of counting and measuring body morphological parameters such as length, maximum thickness, eccentricity, arc length, perimeter, and area, of individuals from the *Hyalella*

genus. In this study, we are presenting the software HyACS (*Hyalella* Automatic Counting System), which is a form to call a tool developed to assist and enhance the work involving Hyalella and is aimed at detecting, counting, and extracting body metrics of the *Hyalella* genus individuals through deep learning and digital image processing techniques. This tool serves as an aid for researchers involved in aquatic ecology and water quality monitoring and offers advantages in efficiently and accurately obtaining data about metrics and counting *Hylella* organism. It allows researchers to automate the primary task related to this organism, which is quantifying the number of individuals with 92% of accuracy. Avoiding this task would rely on human observation, which can introduce errors. It allows for automatic analysis of a large set of individual image data, providing a novel approach for understanding both interspecific and intraspecific variability. Currently there is no other tool specifically designed for *Hyalella* organisms. Although there are similar tools, they are developed for other organisms, such as Aphid CV software for terrestrial aphids (Lins et al. 2020).

Materials and methods

The automated method for counting and extracting phenotypical characteristics of *Hyalella* individuals consists of two stages: the training of the model and the extraction of phenotypical features of the individuals (Fig. 1). The model training is based on the YOLOv3's architecture, and the extraction of characteristics from individuals was performed by using the OpenCV library (Open-Source Computer Vision) version 4. 1. 2 (Zelinsky 2009) in Python 3.7 (Rossum 2018).
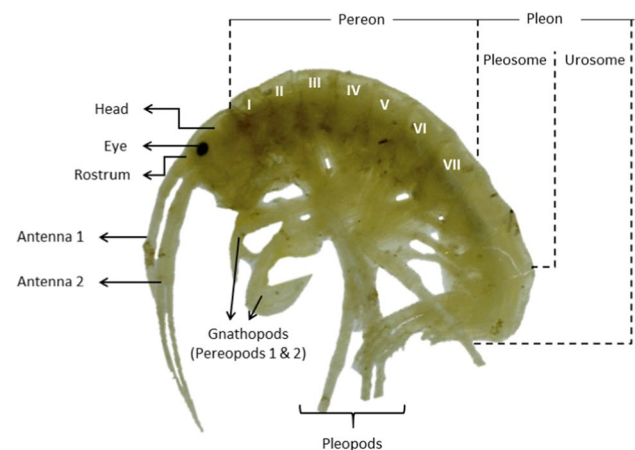

Fig. 1 Main phenotypical features of a *Hyalella*

Obtention of organisms and images

Hyalella individuals included in this study were collected from Lake Tota, located in Boyacá, Colombia (5°32′12. 8″ N 72°55′40. 7″ W). *Hyalella* is particularly abundant in this lake and has served as a bioindicator of local water quality (Pineda-Alarcón & Cañón 2023). The habitat of *Hyalella* in this lake is associated with the aquatic plant *Egeria densa* (Planch, 1849) along the lake's shoreline. For all macroinvertebrates samplers, ever is necessary the application of a standard protocol before counting or taxonomic determine macroinvertebrates. The sample preparation requires washing the collected material two or three times with clean water through the net, being careful to retain the sample inside the net. Afterwards, the net is inspected to remove large debris, detect organisms, and place the organisms found into the sample container. In the lab, the sample is rinsed thoroughly in a 500 μm mesh sieve to remove preservative and fine sediment, and large organic material (leaves, twigs, algal or macrophyte mats, etc.). Finally, the samples are preserved in alcohol (Rice et al. 2012; Ntislidou et al. 2021).

To obtain the images as a basic input in the process of quantification and characterization of *Hyalella*s, we designed an experimental assembly that allowed the visualization of a Petri dish with the individuals fixed in 70% alcohol. In this setup, we used a transparent plastic box ($25 \times 15 \times 20$ cm) with a lid on the top in which we practiced an orifice the size of the Petri dish. We put the sample of *Hyalella*s in the Petri dish, placing the set on the orifice, and covering it with a sheet of paper to prevent the influence of external light. On the inside of the plastic box, an oblique mirror reflects on one side the contents of the background of the Petri dish, allowing the projected scene to be captured with a camera Nikon D5600, which has a 24MP CMOS sensor with no optical low-pass filter (OLPF) and the company's latest EXPEED 4 processor. This combination offers an ISO range of 100–25,600 and 5 fps burst shooting, while the 39-point AF system can track subjects in "3D." Other features include a fully articulating 3.2″ touchscreen LCD with 1.04 M dots, 1080/60p video, and Wi-Fi with Bluetooth for a constant connection, plus NFC for quick pairing with Android devices (Fig. 2).

To avoid sources of noise at the imaging stage, samples from the Petri dish should contain evenly distributed individuals, removing any waste element as much as possible and avoiding the overlapping of individuals or the contact with the periphery of the Petri dish and air bubbles.

Method for counting and characterizing individuals

We based our detection method on YOLO (Redmon et al. 2016), taking an image $f(x, y)$ of a Petri dish containing individuals of the genus *Hyalella*, and returning an inventory of all regions $r_j\left(x_{1_j}, y_{1_j}, x_{2_j}, y_{2_j},\right) \in f(x, y)$ containing a *Hyalella*, indicating the organism's body, delimited by the rectangle with a top left vertex x_{1_j}, y_{1_j} and bottom right vertex x_{2_j}, y_{2_j} for each individual j identified. Figure 3 summarizes the prediction model for identifying and counting *Hyalella* individuals.

Training of the prediction model

To train a model using the YOLO's architecture, two types of inputs are needed: parameters configuration $\{p_1, p_2, \ldots, p_n\}$ and an image bank with delimited regions containing individuals to identify $z_j\left(x_{1_j}, y_{1_j}, x_{2_j}, y_{2_j}\right)$.

This study used YOLOv3 with the open-source framework darknet and the neural network structure

Fig. 2 Outline and assembly elements for image capture: lateral view (**A**), upper view (**B**), and front view (**C**). Elements: Petri dish (1), sheet of paper (2), side wall of the plastic box (3), mirror (4), lid (5), camera (6)
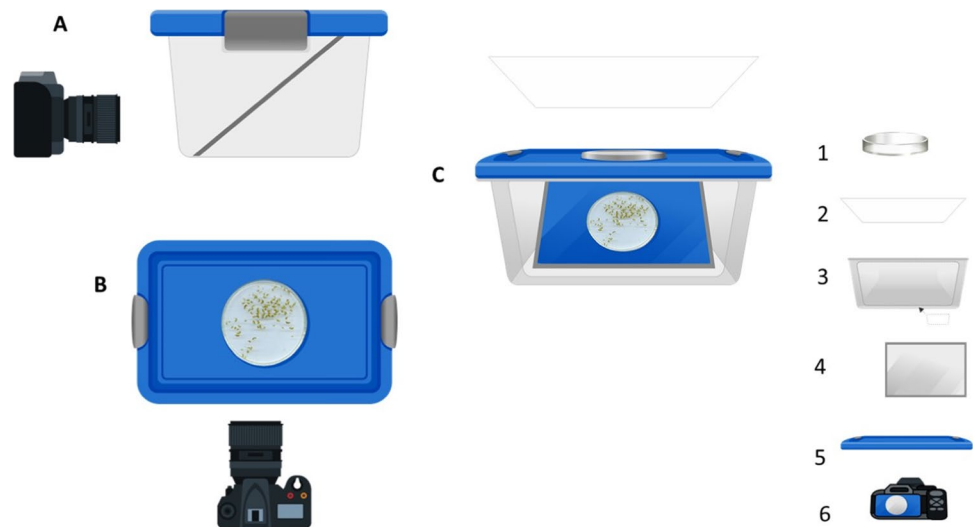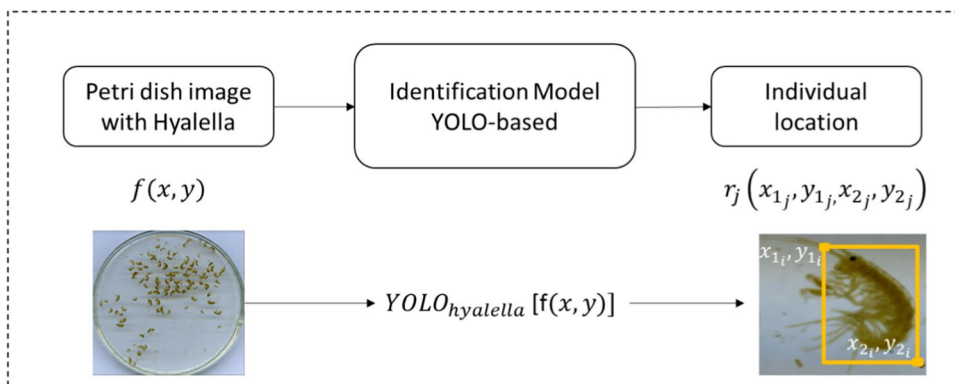
Fig. 3 Prediction model for the *Hyalella* count created with YOLO. From an original image $f(x, y)$, the model returns the total number of individuals and the region r_j occupied by each of the individuals

recommended by the community in the GitHub repository, https://github.com/AlexeyAB/darknet, that allows defining parameters such as network size (i.e., width of 416 pixels and height of 416 pixels), learning rate (i.e., 0.001), number of training iterations (i.e., 10000), among others (the file with the configuration parameters used in this work is provided as supplementary material). In addition, the labeling process is based on the tool recommended in the GitHub repository https://github.com/ManivannanMurugavel/Yolo-Annotation-Tool-New-, which provides a GUI to easily select the bounding boxes and automatically generate the files with the normalized values of the bounding boxes. We used a set of portions $g_k(x, y)$ of the original image to enter the data of the individuals for training. The regions $z_j \left(x_{1_j}, y_{1_j}, x_{2_j}, y_{2_j} \right)$, containing the individuals to be used for training, were manually determined from the set of portions (Fig. 4). We applied the data augmentation technique suggested by Mikołajczyk and Grochowski (2018) to the portion $g_k(x, y)$, thus, obtaining enough individuals to allow the development of a good model.

Dataset augmentation for training

YOLO is trained to predict the region that an object of interest occupies in a picture using a considerable number of sample images. These training images must contain multiple orientations of the object to be identified. Literature recommends no less than 1000 images for this procedure per type of individual. To achieve this training base, we used a *data-augmentation* technique (Shorten and Khoshgoftaar 2019) which consisted in dividing each image of a Petri dish $f(x, y)$ into a set of k sub-images $g_k(x, y)$, either of 4, 6, or 9 portions, which in turn were replicated four times by rotating them 0°, 90°, 180°, and 270° to obtain new images $g_{k_i}(x, y)$. With this procedure, we obtained 2781 images, from which we get the $z_j \left(x_{1_j}, y_{1_j}, x_{2_j}, y_{2_j} \right)$ regions containing the individuals to train the model (Fig. 5).

Extraction of morphological characteristics

After detecting the *Hyalella* organism in the image, we extracted the morphological characteristics listed in Table 1 for each organism framed in the regions $r_j \left(x_{1_j}, y_{1_j}, x_{2_j}, y_{2_j} \right)$

Fig. 4 Training YOLO algorithm for the identification of *Hyalella* individuals. The parameters set $\{ p_1, p_2, \ldots p_n \}$ and the interest regions $z_j \left(x_{1_j}, y_{1_j}, x_{2_j}, y_{2_j} \right)$ are the inputs to train the $YOLO_{\text{Hyalella}}[]$ model
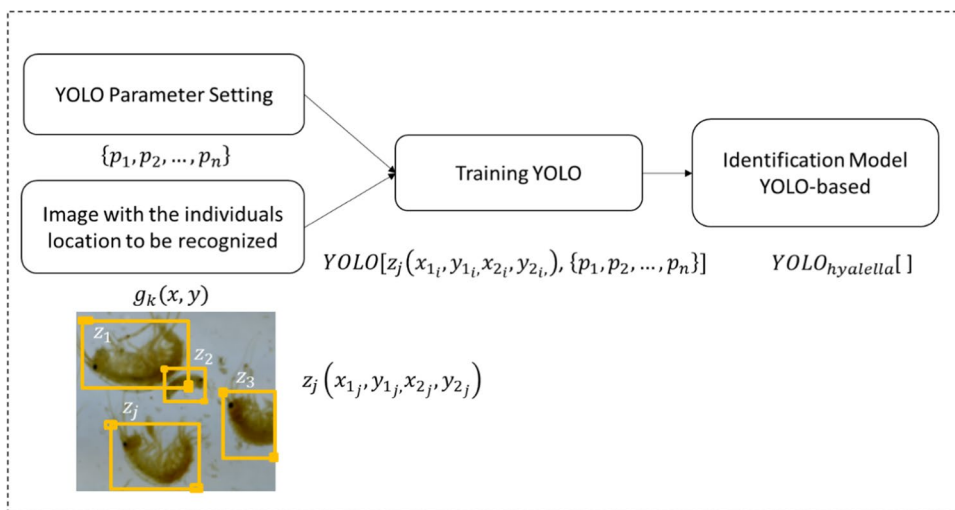
(Fig. 4). The length (*m2*) determines the linear measure of the individual. The arc length (*m3*) represents the total body measure along the curved shape of the individual. The maximum width (*m1*) refers to the width of the individual. The eccentricity (*e*) is a dimensionless parameter that indicates the degree of curvature of the individual. The perimeter (P) and the area (A) denote the contour and the space occupied by the organism

within the Petri dish respectively (Fig. 6). The characteristics of interest were obtained using basic digital image processing techniques. Initially, the image is converted to the HSV color space (hue, saturation, value), then binarized using the Otsu method (Gonzalez and Woods 2000), continuing with the application of morphological operations to finally identify the individual's contour. Based on this outline, the characteristics

Fig. 5 Method of database augmentation for training. Grids of 4, 6, or 9 elements divide the $f(x, y)$ image into $g_k(x, y)$ sub-images. Rotations of the sub-images create four new $g_{k_i}(x, y)$ sets. These sets provide the $z_j\left(x_{1_j}, y_{1_j}, x_{2_j}, y_{2_j}\right)$ regions with the individuals for training
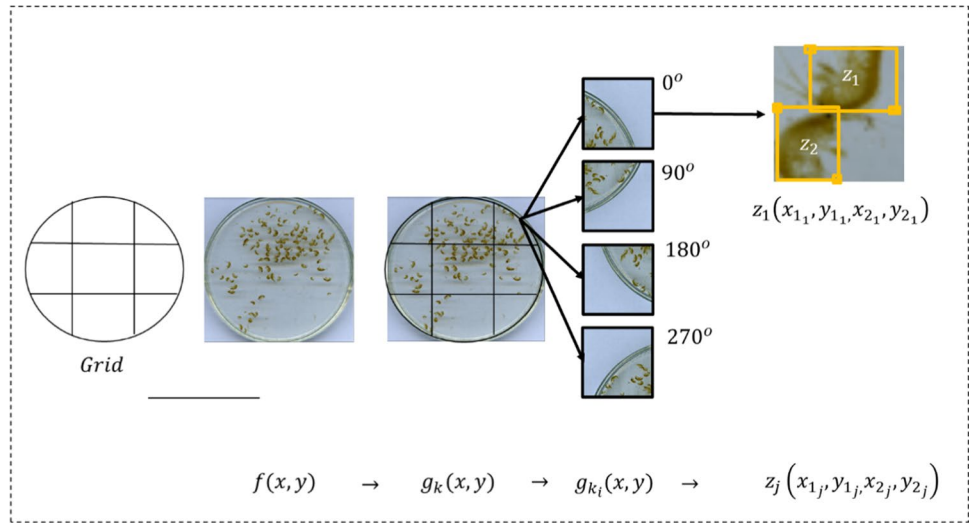
Table 1 Interest characteristics extracted from each *Hyalella* and its mathematical representation

Characteristics	Detail for the individual	Representation
Length	Maximum length in pixels of the largest side of the ellipse containing the specimen	$m1$
Maximum width	Maximum length in pixels between the ends of the specimen, perpendicular to the direction of its longest length	$m2$
Eccentricity	The ratio of the distance between the focal points of the ellipse and the length of its major axis	$e = \sqrt{1 - \frac{m2}{m1}}$
Arc length	Measurement of the distance along a curve in an interval [a, b]	$m3$
Perimeter	Distance between each pair of contiguous pixels around the edge of the region	P
Area	Actual number of pixels in the region	$A = \sum_x \sum_y I(x, y)$

Fig. 6 Representation of the extracted characteristics from each *Hyalella*. **a** *Hyalella*'s length (m2), maximum thickness (m1), arc length (m3), and perimeter (P); **b** *Hyalella*'s area (A)
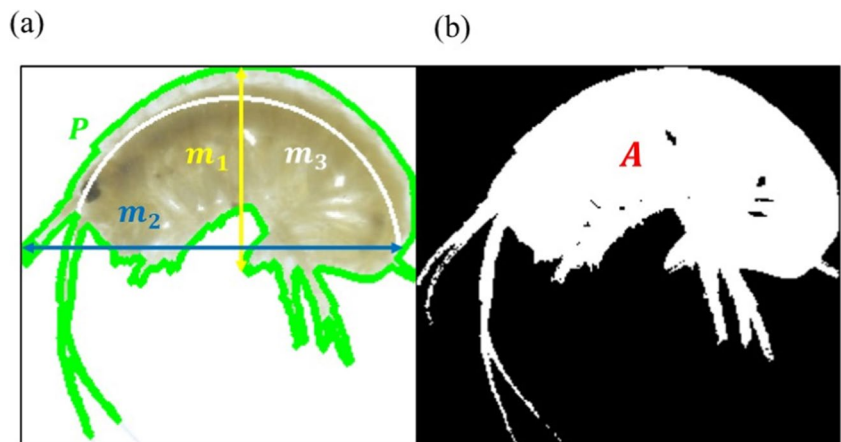
are extracted with algorithms implemented in several functions of the OpenCV library.

The values associated with the extracted characteristics were transformed from pixels to millimeters, using a scale factor based on the standardized diameter of a Petri dish, corresponding to 10 cm (Corkidi et al. 1998).

Model validation

Model performance

To evaluate the performance of the model in detecting *Hyalellas*, we took as a basis the statistical values of true positives (TP-correct identification of *Hyalellas*), false positives (FP-incorrect identification of *Hyalellas*), and false negatives (FN-no detection of *Hyalellas*, when present). From these values, we constructed the confusion matrix to obtain different validation metrics such as accuracy, precision, recall, and F1-score. The training process automatically provides the results of the validation test (confusion matrix), comparing the regions generated by the model $r_j\left(x_{1_j}, y_{1_j}, x_{2_j}, y_{2_j,}\right)$ with the regions provided by the expert $z_j\left(x_{1_j}, y_{1_j}, x_{2_j}, y_{2_j}\right)$ that were obtained from a set of test images. The following are the formulas used for the calculation of each of these metrics.

Accuracy: percentage of *Hyalellas* correctly identified, considering the total number of *Hyalellas* delivered by the expert and the number of objects other than *Hyalellas* that were identified as such (Eq. 1).

$$\text{Accuracy} = \frac{\text{Number of correct detections}}{\text{Total number of Hyalellas} + FP} \quad (1)$$

Precision: number of *Hyalellas* correctly identified, considering the number of identified objects, i.e., the cases that were correctly detected plus the cases other than *Hyalellas* that were identified as such (Eq. 2).

$$\text{Precision} = \frac{\text{Number of correct detections}}{\text{Total number of dectections}} \quad (2)$$

Recall: percentage of *Hyalellas* that the system can correctly identify based on the quantity supplied by the expert (Eq. 3).

$$\text{Recall} = \frac{\text{Number of correct detections}}{\text{Total number of Hyallelas}} \quad (3)$$

F1-score: measurement whose purpose is to weigh the importance of accuracy and recall (Eq. 4).

$$\text{F1score} = \frac{2 \times \text{Precision} \times \text{recall}}{\text{Presicion} + \text{recall}} \quad (4)$$

Comparison between traditional counting and the model developed

To compare traditional and automated quantification methods, we counted the *Hyalellas* in 36 images using both visual inspections by a trained person in the microscope and the model developed. We recorded the time spent in seconds for each method and applied a non-parametric Mann–Whitney test to identify differences (Rencher and Schimek 1997). The results obtained were also compared with the Spearman's correlation method.

Results

Figure 7 shows the HyACS software interface and the step-by-step process of the *Hyalellas* identification, counting, and body metrics extraction. The software uses as input an image of a Petri dish containing an unknown amount of *Hyalellas* (Fig. 7a), and the process starts where the user can track the status through the progress bar displayed on the software interface (Fig. 7b). To identify and quantify the individuals in the image, HyACS relies on the model generated using YOLO, which indicates the regions where the *Hyalellas* are located as the result of the processing (Fig. 7c). Subsequently, the software extracts each of the detected *Hyalellas* as an individual image to determine their defined morphological characteristics (Fig. 7d). Finally, it is provided the number of individuals and the metric's extracted average (Fig. 7e). The user can access the data estimated by the software through two Excel spreadsheets. The first sheet corresponds to the data per individual and the second sheet contains the average data of the processed image (Fig. 8).

Model validation metrics

The model performance is evaluated through the metrics shown in Table 2, which correspond to the values of *accuracy, precision, recall, and F1-score*, highlighting both the *recall*, which indicates that the model identifies more than 90% of *Hyalellas* correctly and the *accuracy* that denotes that the model success in 90% of its predictions.

Visual counting vs. model developed

Figure 9 shows the Spearman correlation between the number of *Hyalellas* counted by visual inspection (average manual counts from 3 different people) and the number of *Hyalellas* detected by the model. The correlation coefficient $r_2 = 0.94$ indicates a high correlation degree between the counts carried out by each method.
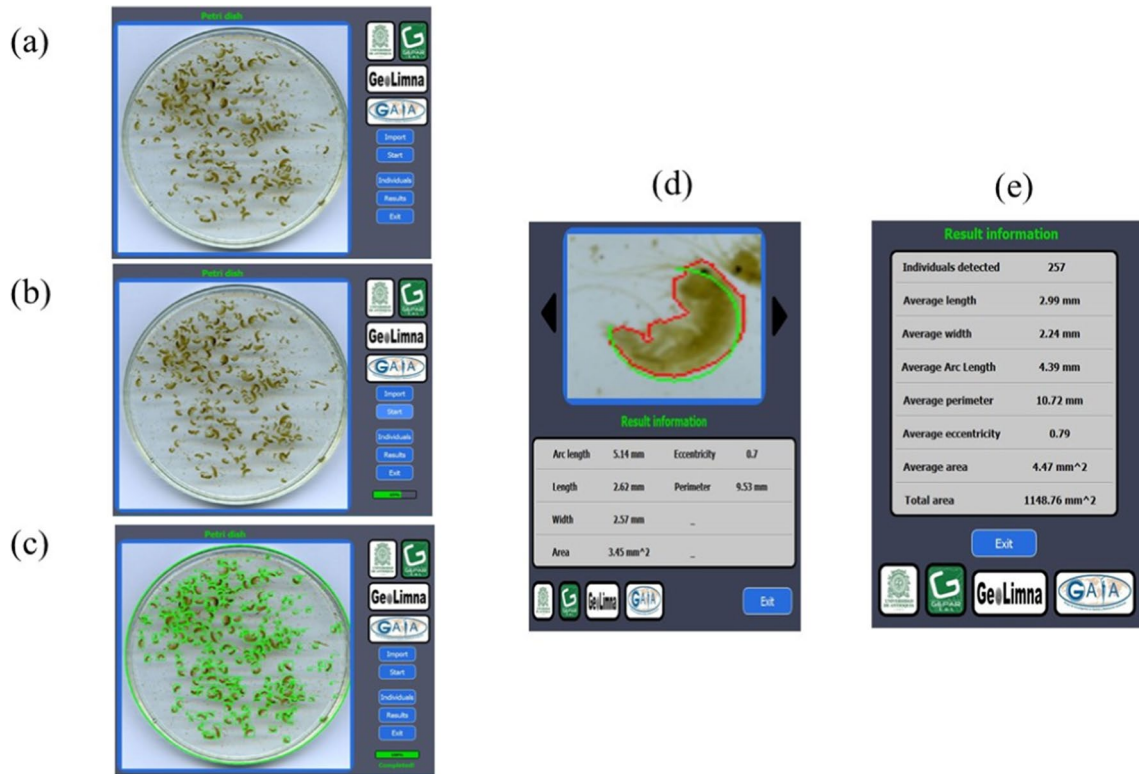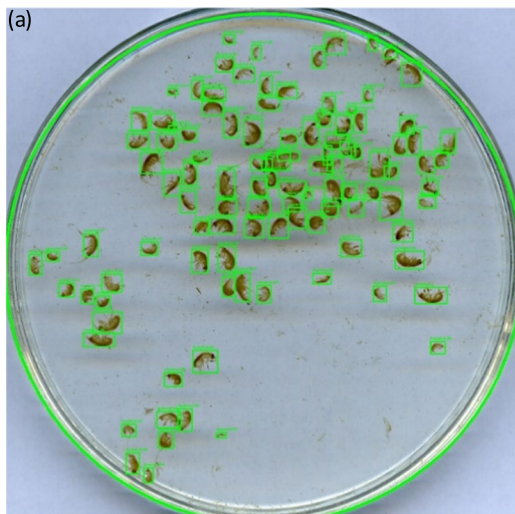
Fig. 7 *Hyalellas* identification, counting, and body metrics extraction in its various stages: **a** importing the input image; **b** detecting and extracting characteristics; **c** visualizing detected individuals; **d** visualizing individual morphological characteristics; **e** averaging characteristics per processed image

(b)

Id	Area (mm^2)	Perimeter (mm)	Eccentricity	Arc length (mm)	Length (mm)	Width (mm)
1	5.43	14.95	0.83	5.82	3.74	3.37
2	3.8	10.85	0.81	4.78	3.06	2.34
3	2.84	11.2	0.85	3.22	3.48	2.28
4	4.64	11.54	0.94	4.62	4.21	2.28
5	4.66	11.98	0.89	4.62	3.53	2.65
6	5.46	11.75	0.87	4.57	3.84	2.23
7	4.53	10.34	0.91	4.83	3.63	2.28
8	6.71	15.71	0.78	6.39	4.67	2.96
9	4.57	10.45	0.7	5.4	2.86	2.49
10	4.4	10.01	0.86	3.84	3.06	2.6
11	5.43	13.66	0.85	5.4	3.79	2.7
12	4	10.26	0.62	4.67	2.86	2.28
13	3.99	9.04	0.76	4.52	2.8	2.02

(c)

Individuals detected	Total area (mm^2)	Average area (mm^2)	Average Arc length (mm)	Average length (mm)	Average width (mm)	Average perimeter (mm)	Average eccentricity
108	573.68	5.31	4.8	3.56	2.6	12.64	0.79

Fig. 8 Individuals detected and spreadsheet report generated by HyACS after processing: **a** Labels of the Hyalella individuals detected. **b** Metrics per individual identified. **c** Average metrics of the individuals identified per image

Table 2 Results of the metrics to evaluate the model's performance

Accuracy	Precision	Recall	F1-score
0.92	0.80	**0.92**	0.86

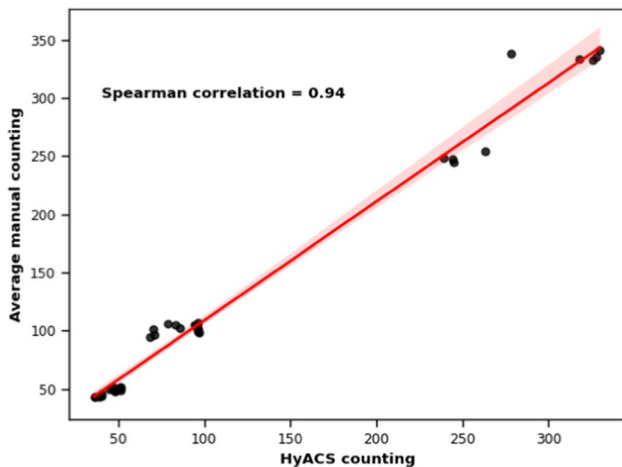Data in bold emphasis indicate the recall and accuracy because these metrics are above 90%

Fig. 9 Correlation of visual counting by an expert vs. automatic counts by the model. The red line represents the best linear fit

Upon comparing the counting time of visual inspection and the model, we observed a significant difference in their distribution and dispersion. The time taken for visual counting ranged widely from 46 to 409 s, with an average of 150 s. Conversely, automatic counting demonstrated minimal variation, ranging from 22 to 58 s, with an average of 35 s. The statistical analysis indicated that there is a statistically significant difference between the two counts, with a *p*-value less than 0.05.

Discussion

Through the correlation analysis between manual and automatic counting, we identify a positive correlation coefficient, suggesting that the model performs correctly. Likewise, we compared the time required for counting and extracting individual characteristics between the two methods and we found that the use of HyACS led to a time reduction of approximately four times. This is one of the highlights of the software, because, in similar studies, Lins et al. (2020), also achieved a fourfold reduction in time using the software (Aphid CV). Therefore, HyACS contributes efficiently to the counting and measurement of individuals of the *Hyalell*a genus.

In the validation of the model's performance, we obtained high prediction rates with metrics above 90% for the correct identification of individuals (accuracy and recall). These

results are close to those obtained in the work of Lins et al. (2020) who use a similar number of training and validation images. Meanwhile, Ding and Taylor (2016) report 93% accuracy for automatic detection of moths with a database five times larger than ours, which indicates that the greater the database the higher the accuracy.

Using HyACS, it is possible to obtain accurate body measurements and other morphological characteristics of individual organisms, which can be of biological and environmental significance (Duckworth et al. 2019). Even in studies focused on describing new species of the genus *Hyalella*, body length is often used as a distinctive measure for identifying different stages of the population from various sizes (Marrón-Becerra et al. 2020). This metric contribute in the classified as males (M), immature females (F), sexually mature females (MF), ovigerous females (OF), and juveniles (J), according to specific criteria (de Paula et al. 2021). The genus *Hyalella* exhibits a wide range of body length, varying between 1.5 and 7.2 mm in the literature (González and Watling 2003; Kühr et al. 2018). However, it is important to note that obtaining of this metric of body length is through linear length measurement that may not fully account for the curved shape of the genus or differences in curvature that may arise when organisms are fixed for analysis (Duckworth et al. 2019).

Studies, such as de Paula et al. (2021), calculate a more accurate body length, by sliding *Hyalella*'s body, measuring body segments individually and adding these measurements to obtain the total length (in millimeters). This process requires a greater investment of time and work. However, HyACS provide a more accurate and efficient body length measurement by detecting automatically the arc length along the curve of the individual. In this regard, HyACS body lengths measurements exceed the maximum length reported in the literature, indicating that traditional techniques may underestimate the body length of organisms (Duckworth et al. 2019).

Ecological studies often involve analyzing the size spectra and organization of communities, where body measurements play a crucial role (Pomeranz and Wesner 2022). Some studies use body length and maximum thickness as dispersion limiting factors among others to understand changes in *Hyalella* populations (Li et al. 2021). For instance, changes in thickness whether the organisms are thicker or thinner could indicate changes in its environment conditions (De Marchi et al. 2019). Various metrics are used in biotic communities to determine the biomass of individuals. Some studies determine the biomass through weight and length (Gualdoni et al. 2013; Stoffels et al. 2003), while others determine the weight from biomass and length to identify the communal energy contribution within an aquatic ecosystem (Rivera-Usme et al. 2015). Additionally, the relationship between body size and dry biomass can be estimated based on the

measured body length to determine toxicity of fipronil and 2,4-D in *Hyalella meinernti* (da Silva Pinto et al. 2021).

Eccentricity is a measure that allows the evaluation of the curvature of the *Hyalella*. This characteristic indicates circular shapes for values close to zero and linear shapes for values close to one. This measure may be a response to various environmental conditions or a species differentiating characteristic (Marrón-Becerra et al. 2020). These characteristics can be obtained through HyACS for the *Hyalella* genus. However, depending on the research question and purpose, the software can be adjusted using tools such as OpenCV library to estimate multiple morphological measurements. In addition, measures such as area and length can be inputs to estimate biologically important metrics such as biomass or weight (Gualdoni et al. 2013; Stoffels et al. 2003).

For better results, both in counting and in measuring characteristics, it is essential to locate the organisms in a way that avoids the edges of the Petri dish. Additionally, it is important to avoid the formation of air bubbles that prevent the bottom of the box from being delimited and the overlapping of organisms.

Using a high-resolution camera that produces quality images can also facilitate and improve the counting and measuring of *Hyalella*'s characteristics. Other authors with similar models have provided this type of recommendation (Lins et al 2020), reaffirming the importance of sample preparation for the acquisition of the image to get a better performance.

The use of software such as HyACS provides a new approach to counting and measuring *Hyalella* that challenges traditional standardized methods. This method also offers the possibility of modifying the sampling protocol since images can be acquired directly in the field. As a result, organisms can be returned to their environment without the need to fix and transport the samples. This approach not only eliminates the need to stress the organisms with the traditional process but it also avoids the alteration of their shape and limbs, which can occur due to the stress of the process and result in their death (Rice et al. 2012).

Conclusions

HyACS is a software that accurately identifies, counts, and extracts shape characteristics from *Hyalella* organisms. It provides fast and reliable results for researchers interested in analyzing data related to the ecology and management of an aquatic ecosystem.

Compared to traditional methods, HyACS offers significant advantages in terms of time spent on counting and the reliability of the results, providing a more accurate population dynamic description. This opens up new possibilities for improving sample processing of the *Hyalella* genus in

laboratory analysis and sampling protocols of individuals in the field. HyACS provides more reliable results on population dynamics and realistic view of the ecosystem.

HyACS can be applied in research related with the demographic, ecological, and toxicological implications of *Hyalellas*, where precise measurements of size and abundance are required. The software would be advantageous to obtain body measurements that consider the curvature of the body as a determining factor in population morphology (Duckworth et al. 2019). These results about metrics brought by HyACS are fast, automatic, and reliable.

The integration of machine learning techniques in fields like biology and ecology is the result of joint efforts from various disciplines. It is therefore important to make progress on methodologies that facilitate the collection of data in populations that are difficult to count or measure and that have been standardized for years using methods that do not consider species-inherent variations, such as the *Hyalella* curvature. HyACS is a pioneering software that uses convolution neural network YOLO to identify objects and employs basic image processing techniques to determine relevant morphological characteristics in the genus *Hyalella*. Its introduction marks a significant advancement in the counting and quantification of characteristics in this genus. We hope that future studies in this area will use HyACS to enhance the quality and reliability of their results.

Data availability The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval Not applicable.

Consent to participate All authors consent to participate in this research.

Consent for publication All authors consent to publish this research in ESPR if accepted.

Competing interests The authors declare no competing interests.

References

Alver MO, Storøy W, Bardal T, Overrein I, Onsøyen MK, Tennøy T, Øie G (2011) Automatic measurement of Acartia tonsa nauplii density, and estimation of stage distribution. Aquaculture 313(1–4):100–106. https://doi.org/10.1016/j.aquaculture.2011.01.033

Amiard-Triquet C, Berthet B (2015) Endobenthic invertebrates as reference species. In Aquatic Ecotoxicology (pp. 229–252). Academic Press. https://doi.org/10.1016/B978-0-12-800949-9.00010-3

Arfianti T, Costello M (2020) Global biogeography of marine amphipod crustaceans: latitude, regionalization, and beta diversity. Mar Ecol Prog Ser 638:83–94. https://doi.org/10.3354/meps13272

Ärje J, Melvad C, Jeppesen MR, Madsen SA, Raitoharju J, Rasmussen MS, Iosifidis A, Tirronen V, Gabbouj M, Meissner K, Høye TT (2020) Automatic image-based identification and biomass estimation of invertebrates. Methods Ecol Evol 11(8):922–931. https://doi.org/10.1111/2041-210X.13428

Bastos RC, Brasil LS, Oliveira-Junior JMB, Carvalho FG, Lennox GD, Barlow J, Juen L (2021) Morphological and phylogenetic factors structure the distribution of damselfly and dragonfly species (Odonata) along an environmental gradient in Amazonian streams. Ecol Ind 122:107257. https://doi.org/10.1016/j.ecolind.2020.107257

Bastos-Pereira R, Bueno AADP (2016) Dynamics of a natural population of a hyallelid amphipod from Brazil. J Crustac Biol 36(2):154–162. https://doi.org/10.1163/1937240X-00002416

Beristain-Castillo E, Salazar-Bautista JL, Moyaho A (2023) Guarding success of the rarest sex in the amphipod Hyalella azteca Saussure, 1858 (Amphipoda: Hyalellidae). J Crust Biol 43(3):ruad051. https://doi.org/10.1093/jcbiol/ruad051

Chen S, Tang Y, Zou X, Huo H, Hu K, Hu B, Pan Y (2022) Identification and detection of biological information on tiny biological targets based on subtle differences. Machines 10(11):1–13. https://doi.org/10.3390/machines10110996

Colla MF, César II (2019) Ecological aspects of natural populations of Hyalella pampeana (Crustacea, amphipoda, hyalellidae) from the natural reserve island of Martín García (Río de La Plata, Argentina). Anais da Academia Brasileira de Ciências 91. https://doi.org/10.1590/0001-3765201920170928

Corkidi G, Diaz-Uribe R, Folch-Mallol JL, Nieto-Sotelo J (1998) COVASIAM: an image analysis method that allows detection of confluent microbial colonies and colonies of various sizes for automated counting. Appl Environ Microbiol 64(4):1400–1404. https://doi.org/10.1128/AEM.64.4.1400-1404.1998

Correa-Araneda F, Contreras A, De Los Ríos P (2010) Amphipoda and Decapoda as potential bioindicators of water quality in an urban stream (38°S, Temuco, Chile). Crustaceana 83(8):897–902. https://doi.org/10.1163/001121610X502948

da Silva Pinto TJ, Moreira RA, da Silva LCM, Yoshii MPC, Goulart BV, Fraga PD, da Silva Rolim VL, Montagner CC, Daam MA, Espindola ELG (2021) Toxicity of fipronil and 2,4-D formulations (alone and in a mixture) to the tropical amphipod Hyalella meinerti. Environ Sci Pollut Res 28(28):38308–38321. https://doi.org/10.1007/s11356-021-13296-9

de Paula DR, Cunha KVS, Garcia ICP, de Paiva Barros-Alves S, Jacobucci GB (2021) Population biology of sympatric species of Caprella (Amphipoda: Caprellidae) in a tropical algal bed. Sci Mar 85(1):39–47. https://doi.org/10.3989/scimar.05065.004

De Marchi L, Neto V, Pretti C, Chiellini F, Morelli A, Soares AM, ..., Freitas R (2019) The influence of Climate Change on the fate and behavior of different carbon nanotubes materials and implication to estuarine invertebrates. Comp Biochem Physiol Part - C: Toxicol Pharmacol, 219:103–115. https://doi.org/10.1016/j.cbpc.2019.02.008

Ding W, Taylor G (2016) Automatic moth detection from trap images for pest management. Comput Electron Agric 123:17–28. https://doi.org/10.1016/j.compag.2016.02.003

Duckworth J, Jager T, Ashauer R (2019) Automated, high-throughput measurement of size and growth curves of small organisms in well plates. Sci Rep. https://doi.org/10.1038/s41598-018-36877-0

Ntislidou C, Bobori D, Lazaridou M (2021) Suggested sampling methodology for lake benthic macroinvertebrates under the requirements of the European Water Framework Directive. Water 13(10):1353. https://doi.org/10.3390/w13101353

Gonzalez RC, Woods RE (2000) Processamento De Imagens Digitais. Editora Blucher, São Paulo

González ER, Watling L (2003) A new species of Hyalella from Colombia, and the redescription of H. meinerti Stebbing, 1899 from Venezuela (Crustacea: Amphipoda). J Nat Hist 37(17):2095–2111. https://doi.org/10.1080/00222930210133255

Gualdoni C, French PW, Oberto AM (2013) Relaciones longitud-biomasa en macroinvertebrados bentónicos de un arroyo serrano del sur de Córdoba, Argentina. Ecol Austral. https://doi.org/10.25260/ea.13.23.3.0.1155

Jacobsen D, Marín R (2008) Bolivian Altiplano streams with low richness of macroinvertebrates and large diel fluctuations in temperature and dissolved oxygen. Aquat Ecol. https://doi.org/10.1007/s10452-007-9127-x

Johnsrud S, Yang H, Nayak A, WadugePunyasena S (2013) Semi-automated segmentation of pollen grains in microscopic images: a tool for three imaging modes. Grana 52(3):181–191

Joiya F (2022) Object detection: Yolo Vs Faster R-Cnn. Int Res J Modern Eng Technol Sci 09:1911–1915. https://doi.org/10.56726/irjmets30226

Kong S, Wang R, Lang T, Liu C (2019) Measurement of insect mass based on ellipsoid scattering model. J Eng 2019(21):7455–7458. https://doi.org/10.1049/joe.2019.0499

Kühr S, Schneider S, Meisterjahn B, Schlich K, Hund-Rinke K, Schlechtriem C (2018) Silver nanoparticles in sewage treatment plant effluents: chronic effects and accumulation of silver in the freshwater amphipod Hyalella azteca. Environ Sci Eur 30(1):1–11. https://doi.org/10.1186/s12302-018-0137-1

Kvæstad B, Hansen BH, Davies E (2022) Automated morphometrics on microscopy images of Atlantic cod larvae using Mask R-CNN and classical machine vision techniques. MethodsX 9(November 2021):101598. https://doi.org/10.1016/j.mex.2021.101598

Lee H, Park J, Shin K, Depuydt S, Choi S, De Saeger J, Han T (2020) Application of a programmed semi-automated Ulva pertusa bioassay for testing single toxicants and stream water quality. Aquat Toxicol 221:105426

Li Z, Heino J, Chen X, Liu Z, Meng X, Jiang X, Ge Y, Chen J, Xie Z (2021) Understanding macroinvertebrate metacommunity organization using a nested study design across a mountainous river network. Ecol Ind 121:107188. https://doi.org/10.1016/j.ecolind.2020.107188

Lins EA, Pedro J, Rodriguez M, Scoloski SI, Pivato J, Balotin M, Maurício J, Fernandes C, Roberto P, Pereira S, Lau D, Rieder R, Fundo P, Fundo P, Grande R (2020) A method for counting and classifying aphids using computer vision. Comput Electron Agric 169(January):105200. https://doi.org/10.1016/j.compag.2019.105200

Mallard F, Le Bourlot V, Tully T (2013) An automated image analysis system to measure and count organisms in laboratory microcosms. PLoS ONE 8(5):1–10. https://doi.org/10.1371/journal.pone.0064387

Manoukis NC, Collier TC (2019) Computer Vision to Enhance Behavioral Research on Insects. Ann Entomol Soc Am 112(3):227–235. https://doi.org/10.1093/aesa/say062

Marrón-Becerra A, Hermoso-Salazar M, Rivas G (2020) A new species of the genus hyalella (Crustacea, amphipoda) from Northern Mexico. ZooKeys 2020(942):1–19. https://doi.org/10.3897/zookeys.942.50399

Mikołajczyk A, Grochowski M (2018) Data augmentation for improving deep learning in image classification problem. In 2018 international interdisciplinary PhD workshop (IIPhDW) (pp. 117–122). IEEE. https://doi.org/10.1109/IIPHDW.2018.8388338

Nnoli H, Kyerematen R, Adu-Acheampong S, Hynes J (2019) Change in aquatic insect abundance: evidence of climate and land-use change within the Pawmpawm River in Southern Ghana. Cogent Environ Sci 5(1):1594511. https://doi.org/10.1080/23311843.2019.1594511

Park Y-S, Chung N-I, Choi K-H, Cha EY, Lee S-K, Chon T-S (2005) Computational characterization of behavioral response of medaka (Oryzias latipes) treated with diazinon. Aquat Toxicol 71(3):215–228

Pech D, Condal AR, Bourget E, Ardisson PL (2004) Abundance estimation of rocky shore invertebrates at small spatial scale by high-resolution digital photography and digital image analysis. J Exp Mar Biol Ecol. https://doi.org/10.1016/j.jembe.2003.08.017

Pineda-Alarcón LY, Cañón JE (2023) Modelación de la Relación Predador-Presa para la Comunidad de Macroinvertebrados en el Litoral del Lago de Tota. Acta Biológica Colombiana 28(2). https://doi.org/10.15446/abc.v28n2.97983

Pomeranz JP, Junker JR, Wesner JS (2022) Individual size distributions across North American streams vary with local temperature. Glob Chang Biol 28(3):848–858. https://doi.org/10.1111/gcb.15862

Posada JA, Roldán G, Ramírez JJ (2000) Caracterización fisicoquímica y biológica de la calidad de aguas de la cuenca de la quebrada Piedras Blancas, Antioquia, Colombia. Revista de Biología Tropical 48(1):59–70. http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S0034-77442000000100008&lng=en&tlng=es

Raitoharju J, Riabchenko E, Ahmad I, Iosifidis A, Gabbouj M, Kiranyaz S, Tirronen V, Ärje J, Kärkkäinen S, Meissner K (2018) Benchmark database for fine-grained image classification of benthic macroinvertebrates. Image Vis Comput 78:73–83. https://doi.org/10.1016/j.imavis.2018.06.005

Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. Proc IEEE Comput Soc Conf Comput Vis Pattern Recog. https://doi.org/10.1109/CVPR.2016.91

Reis GO, Penoni LR, Bueno AAP (2020) First record of the genus Hyalella (Amphipoda: Hyalellidae) from Santa Catarina State, Brazil, with description of two new species. Biota Neotropica 20(2). https://doi.org/10.1590/1676-0611-BN-2019-0879

Rencher AC, Schimek MG (1997) Methods of multivariate analysis. Comput Stat 12(4):422

Rice EW, Bridgewater L, American Public Health Association (2012) Standard methods for the examination of water and wastewater (Vol. 10). Washington, DC: American public health association. Standard-Methods-23rd-Perv.pdf (yabesh.ir)

Rivera-Usme J, Pinilla G, Rangel-Churio J, Castro M, Camacho-Pinzón D (2015) Biomass of macroinvertebrates and physicochemical characteristics of water in an Andean urban wetland of Colombia. Braz J Biol 75(1):180–190. https://doi.org/10.1590/1519-6984.10613

Rocha Penoni L, Alvarenga Lares MM, de Pádua Bueno AA (2021) Description of two new species of the freshwater amphipod Hyalella Smith, 1874 (Amphipoda: Hyalellidae) from southeastern Brazil, with remarks on their population biology and reproduction. J Crustac Biol 41(3):ruab050. https://doi.org/10.1093/jcbiol/ruab050

Roldán-Pérez G (2016) Los macroinvertebrados como bioindicadores de la calidad del agua: cuatro décadas de desarrollo en Colombia y Latinoamerica. Rev Acad Colomb Cienc Exactas Fís Nat 40(155):254. https://doi.org/10.18257/raccefyn.335

Rossum GV (2018) The Python Library Reference, release 3.6.4. 12th Media Services

Serna López JP, Fernández Mc Cann DS, de J Vélez Macías F, Aguirre Ramírez NJ, (2020) An image processing method for recognition of four aquatic macroinvertebrates genera in freshwater environments in the Andean region of Colombia. Environ Monit Assess 192(10):617. https://doi.org/10.1007/s10661-020-08545-2

Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data. https://doi.org/10.1186/s40537-019-0197-0

Stepanian PM, Entrekin SA, Wainwright CE, Mirkovic D, Tank JL, Kelly JF (2020) Declines in an abundant aquatic insect, the burrowing mayfly, across major North American waterways. Proc Natl Acad Sci 117(6):2987–2992. https://doi.org/10.1073/pnas.1913598117

Stoffels RJ, Karbe S, Paterson RA (2003) Length-mass models for some common New Zealand littoral-benthic macroinvertebrates, with a note on within-taxon variability in parameter values among published models. NZ J Mar Freshwat Res. https://doi.org/10.1080/00288330.2003.9517179

Untersteiner H, Kahapka J, Kaiser H (2003) Behavioural response of the cladoceran Daphnia magna Straus to sublethal Copper stress—validation by image analysis. Aquat Toxicol 65(4):435–442

Vilgrain L, Maps F, Picheral M, Babin M, Aubry C, Irisson J-O, Ayata S-D (2021) Trait-based approach using in situ copepod images reveals contrasting ecological patterns across an Arctic ice melt zone. Limnol Oceanogr 66(4):1155–1167

Wang X, Xu T, Zhang J, Chen S, Zhang Y (2018) SO-YOLO based WBC detection with Fourier ptychographic microscopy. IEEE Access 6:51566–51576. https://doi.org/10.1109/ACCESS.2018.2865541

Weinstein BG (2018) A computer vision for animal ecology. J Anim Ecol 87(3):533–545. https://doi.org/10.1111/1365-2656.12780

Wührl L, Pylatiuk C, Giersch M, Lapp F, Rintelen T, Balke M, Schmidt S, Cerretti P, Meier R (2022) DiversityScanner: robotic handling of small invertebrates with machine learning methods. Mol Ecol Resour 22(4):1626–1638. https://doi.org/10.1111/1755-0998.13567

Yati A, Dey S (2011) FlyCounter: a simple software for counting large populations of small clumped objects in the laboratory. Biotechniques 51(5):347–348. https://doi.org/10.2144/000113753

Zelinsky A (2009) Learning OpenCV---Computer vision with the OpenCV library (Bradski, GR et al.; 2008)[On the Shelf]. IEEE Robot \Autom Mag 16(3):100

Zepon T, Resende LPA, Bueno AAP, Bichuette ME (2021) New records of the troglobitic Hyalella veredae Cardoso & Bueno, 2014 (Crustacea, Amphipoda, Hyalellidae) from Minas Gerais caves, southeast Brazil, with notes on its natural history. Check List 17(1):115–123. https://doi.org/10.15560/17.1.115

Zhong Y, Gao J, Lei Q, Zhou Y (2018) A vision-based counting and recognition system for flying insects in intelligent agriculture. Sensors (Switzerland) 18(5):1–23. https://doi.org/10.3390/s18051489

Zipkin EF, Rossman S, Yackulic CB, Wiens JD, Thorson JT, Davis RJ, Grant EHC (2017) Integrating count and detection–nondetection data to model population dynamics. Ecology 98(6):1640–1650. https://doi.org/10.1002/ecy.183