Automated Software for Counting and Measuring Hyalella Genus Using Artificial Intelligence

Ludy Pineda (ludy.pineda@udea.edu.co)  
Universidad de Antioquia

Maycol Zuluaga  
Universidad de Antioquia

Santiago Ruíz  
Universidad de Antioquia

David Fernandez McCann  
Universidad de Antioquia

Fabio Vélez  
Universidad de Antioquia

Nestor Aguirre  
Universidad de Antioquia

Yarin Puerta  
Universidad de Antioquia

Julio Cañón  
Universidad de Antioquia

Research Article

Keywords: measuring protocols, morphological measures, image capture, macroinvertebrates, YOLO

Posted Date: August 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1899518/v1

License: ☎️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Amphipods of the *Hyalella* genus are aquatic macroinvertebrates of interest in areas such as limnology and ecotoxicology that use data on the number of *Hyalella* individuals and their allometric measurements to assess the environmental dynamics of aquatic ecosystems. We introduce the software HyACS, aimed at counting individuals and extracting morphological metrics of the *Hyalella* genus by means of artificial vision based on YOLOv3 and digital image processing techniques. The software detects body metrics of length, arc length, maximum width, eccentricity, perimeter, and area of *Hyalella* individuals, using basic imaging capture equipment. The method requires images for training and prediction stages. Results show high prediction levels with HyACS, reaching metrics above 90%, in the correct identification of individuals, performing up to five times faster than traditional visual counting, and offering precise morphological measurements of *Hyalella* individuals that may improve further studies of the species populations and boost their use as bioindicators of water quality.

1. Introduction

The *Hyalella* (Smith, 1874) is a crustacean of the Amphipod order that belongs to the ecological group of freshwater aquatic macroinvertebrates (Correa-araneda y Contreras, 2010). *Hyalella* is a benthic organism that lives in association with submerged aquatic vegetation (Bastos-Pereira y Bueno, 2016; Reis et al., 2020) and reacts to the presence of pollutants in aquatic ecosystems (Amiard-Triquet y Berthet, 2015). Therefore, *Hyalella* are organisms that react to changes in the quality of water and sediments. The changes that individual *Hyalella* undergo serve to evaluate the responses to contact with different environmental stressors (chemical or physical), while the population dynamics of *Hyalella* is of interest for the monitoring of aquatic ecosystems. *Hyalella* is used as a bioindicator of high organic matter contents in water (Posada et al., 2000) and as a biological model in ecotoxicological studies due to its high sensitivity to heavy metals (Colla y César, 2019). Likewise, *Hyalella*'s high population density is related to fluctuations in electrical conductivity, temperature, and low concentrations of dissolved oxygen (Jacobsen y 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). The quantification of *Hyalella* organisms also helps to identify space-time trends at biological, ecological, and environmental levels (Bastos et al., 2021).

Currently, artificial 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, 2017). Some recent 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 y Dey, 2011); automatic fine-grained classification of benthic macroinvertebrates through the use of convolutional neural networks (CNN) (Raitoharju et al., 2018); automatic measurement
of the density of the copepod Acartia tonsa (Alver et al., 2011); sorting, identification and biomass estimation of terrestrial invertebrates by image-based identification machine (Ärje et al., 2020); morphological descriptors in Arctic zooplankton community measured from in situ images (Vilgrain et al., 2021); automatic sorting and extraction of metrics such as body length and volume from 14 common terrestrial invertebrate specimens (Würl et al., 2021); identification of movement patterns changes in Oryzias latipes (Park et al., 2005) and Daphnia magna (Untersteiner et al., 2003) using digital image processing; development of the software ImageJ as an accurate and flexible method to automate the analysis of digital photographs of the laboratory microcosm to detect, count and measure organisms such as collembola, ants, nematodes and daphnias moving on a fixed but heterogeneous substrate (Mallard et al., 2013); or changes in color of Ulva pertusa macroalgae (Lee 2020); a method to extract the mass of migratory insects based on an ellipsoidal dispersion model using a radar (Kong et al., 2019); segmentation and counting of pollen grains in microscopic images (Johnsrud 2013); the processing and classification of four groups of aquatic macroinvertebrates (Thraulodes, Traverella, Anacroneuria and Smicridea) through digital images processing (Serna López et al., 2020); and the registration and classification of Rhopalosiphum padi aphids to assess and predict crop damage by image processing, machine vision and machine learning through the Aphid CV software (Lins et al., 2020).

While basic image processing technics enable the solution of a wide variety of environmental problems, there are many other cases that require the use of more robust pattern recognition tools such as You Only Look Once (YOLO) (Redmon et al., 2016). 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 obtention of multiple metrics in images related to biological samples. These image analysis tools contribute to the improvement of technics for both counting of individuals and quantification of specimen's body measurements in research related with 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), while Kvaestad, Hansen & Davies (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 y Bueno, 2016). There are widely established sampling protocols for collecting Hyalella (Brandi y Wilson-Wilde, 2013; EPA, 2003). The traditional method of separation and counting is performed manually under a stereomicroscope and is subject to human errors of perception and lack of expertise. The counting of individuals is a time-consuming activity that increases the likelihood of errors, due to fatigue and visual illusions (Lins et al., 2020).

Since the extraction of information from images is quick, objectively verifiable, and less uncertain in terms of observational errors (Pech et al. 2004), it is worth investing efforts in the development of tools that complement and improve the traditional methods of information extraction. Following Manoukis et al., (2019), the use of 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 introduce the software HyACS (*Hyalella* Automatic Counting System), aimed at counting individuals and extracting body metrics of *Hyalella* genus by means of artificial vision and digital image processing techniques as an aid for researchers involved in aquatic ecology and water quality monitoring.

2. Materials And Methods

To implement the automated method for counting and extracting phenotypical characteristics of *Hyalella* individuals, we developed a software called HyACS, which consists of two stages: the training of the model and the extraction of phenotypical features of the individuals. The model training uses the YOLO algorithm through the Deep Learning Darknet framework, and the extraction of characteristics from individuals is supported by the OpenCV library (Open-Source Computer Vision) version 4.1.2 (Zelinsky, 2009) in Python 3.7 (van Rossum y The Python Development Team, 2018).

2.1. 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, 2020). 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, it 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 500µm 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 (Brandi y Wilson-Wilde, 2013; EPA, 2003).

Figure 1 details the main components of the anatomy of a *Hyalella* organism.

To obtain the images as a basic input in the process of quantification and characterization of *Hyalellas*, 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 (25x15x20cm) with a lid on the top in which we practiced an orifice the size of the Petri dish. We put the sample of *Hyalellas* 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 100-25600 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 1.04M 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.

2.2. Method for counting and characterizing individuals

We based our counting 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 \( r_j \left( x_{1j}, y_{1j}, x_{2j}, y_{2j} \right) \) containing a Hyalella, indicating the organism's body, delimited by the rectangle with a top left vertex \( x_{1j}, y_{1j} \) and bottom right vertex \( x_{2j}, y_{2j} \) for each individual \( j \) identified. Figure 3 summarizes the prediction model for identifying and counting Hyalella individuals.

2.2.1. Training of the prediction model

The YOLO algorithm requires two types of inputs for its training: parameters configuration \( \{ p_1, p_2, \ldots, p_n \} \) and an image bank with delimited regions containing individuals to identify \( z_j \left( x_{1j}, y_{1j}, x_{2j}, y_{2j} \right) \).

This study used YOLOv3 with the open-source framework darknet and the neural network structure 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 \( g_k \left( x, y \right) \) of the original image to enter the data of the individuals for training. The regions \( z_j \left( x_{1j}, y_{1j}, x_{2j}, y_{2j} \right) \), 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 y Grochowski, 2018) to the portion \( g_k \left( x, y \right) \), thus, obtaining enough individuals to allow the development of a good model.

2.2.2. 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 1,000 images for this procedure per type of individual. To achieve this training base, we used a data-augmentation technique (Shorten y 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 $0^\circ$, $90^\circ$, $180^\circ$, and $270^\circ$ to obtain new images $g_k(x, y)$. With this procedure, we obtained 2,781 images, from which we get the $z_j(x_1, y_1, x_2, y_2)$ regions containing the individuals to train the model (Fig. 5).

### 2.2.3. 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 $r_j(x_1, y_1, x_2, y_2)$ (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 y Woods, 2000), continuing with the application of morphological operations to finally identify the individual’s contour. Based on this outline, the characteristics are extracted with algorithms implemented in several functions of the OpenCV library.

**Table 1**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Detail for the individual</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>Maximum length in pixels of the largest side of the ellipse containing the specimen.</td>
<td>$m1$</td>
</tr>
<tr>
<td>Maximum width</td>
<td>Maximum length in pixels between the ends of the specimen, perpendicular to the direction of its longest length.</td>
<td>$m2$</td>
</tr>
<tr>
<td>Eccentricity</td>
<td>The ratio of the distance between the focal points of the ellipse and the length of its major axis.</td>
<td>$e = \sqrt{1 - m2^2}$</td>
</tr>
<tr>
<td>Arc length</td>
<td>Measurement of the distance along a curve in an interval $[a, b]$.</td>
<td>$m3$</td>
</tr>
<tr>
<td>Perimeter</td>
<td>Distance between each pair of contiguous pixels around the edge of the region.</td>
<td>$P$</td>
</tr>
<tr>
<td>Area</td>
<td>Actual number of pixels in the region.</td>
<td>$A = \sum_x \sum_y I(x, y)$</td>
</tr>
</tbody>
</table>

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).

2.3. Model validation

2.3.1. Model performance

To evaluate the performance of the software 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 *Hyalella*, 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 $r_j \left( x_{1j}, y_{1j}, x_{2j}, y_{2j} \right)$ with the regions provided by the expert $z_j \left( x_{1j}, y_{1j}, x_{2j}, y_{2j} \right)$ 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.

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

**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.

$$\text{Precision} = \frac{\text{Number of correct detections}}{\text{Total number of detections}}$$

**Recall** percentage of *Hyalellas* that the system can correctly identify based on the quantity supplied by the expert.

$$\text{Recall} = \frac{\text{Number of correct detections}}{\text{Total number of Hyalellas}}$$

**F1-score** measurement whose purpose is to weigh the importance of accuracy and recall.

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

2.3.2. Comparison between traditional counting and HyACS

To compare the traditional and automated quantification methods, the *Hyalellas* of 36 images were counted by both visual inspection of a trained person in the microscope and using the automated method of our software HyACS. We recorded the time spent in seconds for each method and applied a non-
parametric Mann Whitney test to identify differences (Rencher, 2002). The results obtained were compared with the Spearman’s correlation method.

3. Results

Figure 8 shows the HyACS software interface and the step-by-step approach to the Hyalellas quantification and characterization results. The software uses as input an image of a Petri dish containing an unknown amount of Hyalellas (Fig. 7a). To identify and quantify the individuals in the image, HyACS relies on the model generated by YOLO, which indicates the regions where the Hyalellas are located as the result of the processing (Fig. 7c); the user can track the status of the process through the progress bar displayed on the software interface (Fig. 7b). Subsequently, the software extracts each of the detected Hyalellas as an individual image to determine their defined morphological characteristics (Fig. 7d). Finally, HyACS provides the number of individuals and the average characteristics for each image provided (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).

3.1. Software validation metrics

The HyACS software performance in identifying Hyalellas is evaluated through the metrics in Table 2, which correspond to the values of accuracy, precision, recall and F1-score, highlighting both the recall, which indicates that the software identifies more than 90% of Hyalellas correctly and the accuracy that shows that the model prediction is 90% correct.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.92</td>
<td>0.80</td>
<td>0.92</td>
<td>0.86</td>
</tr>
</tbody>
</table>

3.1.1. Visual counting vs. HyACS

Figure 9 shows the Spearman correlation between the number of Hyalellas counted by visual inspection and the number of Hyalellas detected by the software. The correlation coefficient $r^2 = 0.97$ indicates the high association degree between the counts carried out by each method. In addition, statistically significant differences were found with a value of $p < 0.05$ between the time used in the manual counting of Hyalellas and the automated counting using HyACS. Figure 10 shows a wide distribution and dispersion of the time of visual counting, varying in a range of 45 to 450 seconds. In contrast, the automated count shows minimal variation from the time spent on Hyalellas identification and counting varying in a range of 5 to 55 seconds.
4. Discussion

Through the correlation analysis between manual and automatic counting carried out by HyACS software, we identify a positive association, without neglecting the human error in the count. Likewise, the time used for this count was compared between the two methods, finding that the time for counting and extracting characteristics of the individuals was reduced by about 5 times. This is one of the highlights of the software, because, in similar studies (Lins et al., 2020), achieved a 4-fold reduction in time using the software (Aphid CV). Therefore, the implementation of the HyACS software contributes efficiently to the counting and measurement of individuals of the *Hyalella* genus.

For the validation of the HyACS software, we obtained high prediction rates with metrics above 90% regarding the correct identification of individuals and the count of the expert (Accuracy and Recall). These results are close to those obtained in the work of Lins et al., (2020) who use a similar amount of training and validation images. Meanwhile, Ding and Taylor, (2016) report 93% accuracy for automatic detection of moths with a database 5 times larger than ours, which indicates that the greater the database the higher the accuracy.

Identification of whole bodies through HyACS allows obtaining the body measurement of individuals, as well as other morphological characteristics of biological and environmental interest (Duckworth et al., 2019). Even in studies describing new species of the genus *Hyalella*, body length is estimated as a distinctive measure of the species, serving to define the stages of the population from various sizes (Marrón-Becerra et al., 2020). According to the literature, the genus *Hyalella* varies in body length between 1.5 and 7.2 mm (González y Watling, 2003; Kühr et al., 2018). However, this is a linear length that does not consider the curved shape of the genus and the differences in curvature that may occur when fixing organisms (Duckworth et al., 2019).

Therefore, HyACS detects the arc length that allows determining the body length along the curve of the individual, which is a more accurate body measurement for individuals of the same species compared to the linear length. In this regard, HyACS measured body lengths that exceed the maximum length reported in the literature (Duckworth et al., 2019), indicating that traditional techniques may underestimate the body length of organisms.

Another characteristic is the maximum thickness that could indicate changes in *Hyalella* populations, considering whether the organisms are thicker or thinner, which may be a biological response to changes in the conditions of their environment (Marchi et al. 2019). Regarding the use of various metrics in biotic communities, some studies determine the biomass of individuals 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, J. J., Pinilla-Agudelo, G. A.; Camacho-Pinzón, D. L.; Castro-Rebolledo, M. I.; & Rangel-Churio, 2014). Other studies analyze size spectra (Pomeranz y Wesner, 2020) and the organization of communities through the analysis of body measurement as a dispersion limiting factor (Li et al. 2021), among others. Therefore, measurements of body length and maximum thickness obtained through HyACS would facilitate ecological studies of this
type. Another characteristic is the maximum thickness that could indicate changes in *Hyalella* populations, considering whether the organisms are thicker or thinner, which may be a biological response to changes in the conditions of their environment (Marchi et al. 2019). Regarding the use of various metrics in biotic communities, some studies determine the biomass of individuals 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, J. J., Pinilla-Agudelo, G. A.; Camacho-Pinzón, D. L.; Castro-Rebolledo, M. I.; & Rangel-Churio, 2014). Other studies analyze size spectra (Pomeranz y Wesner, 2020) and the organization of communities through the analysis of body measurement as a dispersion limiting factor (Li et al. 2021), among others. Therefore, measurements of body length and maximum thickness obtained through HyACS would facilitate ecological studies of this type.

Eccentricity 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. Therefore, this measure may be a response to various environmental conditions or a species differentiating characteristic (Marrón-Becerra et al., 2020). The characteristics mentioned above are some that can be obtained through HyACS for the *Hyalella* genus. However, depending on the question and the purpose of the research, the software can be adjusted using tools such as OpenCV library to estimate multiple morphological measurements. In addition, measures such as area and length may 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. The formation of air bubbles that prevent the bottom of the box from being delimited and the overlapping of organisms must also be avoided. The quality of the images from a high-resolution camera also facilitates and improves the counting and the measuring of *Hyalella* characteristics. This type of recommendation has been provided by other authors with similar models (Lins et al, 2020), re-affirming the importance of sample preparation for the acquisition of the image to get a better performance.

The use of software such as HyACS opens a path for changing traditional standardized methods for counting and measuring *Hyalella*. Furthermore, the method opens a venue for modifying the sampling protocol, since it will be feasible to acquire the images directly in the field, returning the organisms to their environment without the need to fix and transport the samples. This possibility would also avoid the death of the organisms and the alteration of their shape and limbs due to the stress of the traditional process (Baird et al., 2017).

5. Conclusions

HyACS is a software for identifying, counting, and extracting shape characteristics from *Hyalella* organisms that works with high precision and accuracy, therefore providing fast and reliable results for
researchers interested in further data analysis associated with the ecology and management of an aquatic ecosystem.

Compared to traditional methods, HyACS shows great advantages regarding the time spent on counting and the reliability of the automated results, opening new possibilities for improving the processing of samples of the *Hyalella* genus in laboratory analysis and sampling protocols of individuals in the field, obtaining results that are more reliable to population dynamics and realistic of the ecosystem.

HyACS can be applied in research of the demographic, ecological, and toxicological implications of *Hyalellas*, where exact 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 morphologies (Duckworth et al., 2019), thus, obtaining metrics that are fast, automatic, and reliable.

The joint work of different disciplines contributes to the integration of tools based on machine learning techniques in areas such as biology and ecology. 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. Therefore, HyACS is a pioneering software in the counting and quantification of characteristics in the genus *Hyalella*, which identifies objects using convolution neural network YOLO and determines relevant morphological characteristics with basic image processing techniques. We hope that future studies with this genus will use HyACS to support the quality of their results.

**Declarations**

**Author contributions**


**Funding**

Partial funding for this research was provided by the call 733 for the formation of high-level human capital for the Department of Boyacá 2015, from the Colombian Ministry of Science and Technology and Innovation.

**Availability of data and material**
The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of Competing Interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the Graduate Office at the School of Engineering at Universidad de Antioquia for facilitating the human, material, and computational resources to develop this research. We are also grateful to the research groups in environmental management and modeling (GAIA), the group on power electronics, automation, and robotics (GEPAR), and the research group in geography, limnology, and environmental modeling (GEOLIMNA) of Universidad de Antioquia.

References


quality. Aquatic Toxicology, 221, 105426.


**Supplementary Materials**

The Supplementary Materials are not available with this version.

**Figures**
Figure 1

Main phenotypical features of a *Hyalella* individual.
Figure 2

Outline and assembly elements for image capture: lateral view (A), upper view (B), front view (C). Elements: Petri dish (1), sheet of paper (2), side wall of the plastic box (3), mirror (4), lid (5), camera (6).

Figure 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.

Figure 3

See image above for figure legend.
Figure 4. Training YOLO algorithm for the identification of Hyalella individuals. The parameters set \( \{p_1, p_2, ..., p_n\} \) and the interest regions \( z_j(x_{1j}, y_{1j}, x_{2j}, y_{2j}) \) are the inputs to train the YOLO_{Hyalella}[]} model.

Figure 4

See image above for figure legend.
Figure 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 \( g_{k_t}(x, y) \) sets. These sets provide the \( z_j(x_1, y_1, x_2, y_2) \) regions with the individuals for training.

Figure 5

See image above for figure legend.
Figure 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).
Figure 7

Performance images of HyACS 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.
Figure 8

Spreadsheets of the report generated by HyACS in Excel: a) results per image; b) results per individual identified.
Figure 9

Correlation of visual counting by an expert vs. automatic counts by HyACS. The red line represents the best linear fit.
Figure 10

Distribution of time (in seconds) spent in counting the *Hyalellas* with HyACS and by visual inspection.