Computers and Electronics in Agriculture 170 (2020) 105274

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Deep Learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia

Arthur F.A. Fernandes[a], Eduardo M. Turra[b], Érika R. de Alvarenga[b], Tiago L. Passafaro[a], Fernando B. Lopes[a], Gabriel F.O. Alves[b], Vikas Singh[c], Guilherme J.M. Rosa[a,c,*]

[a] Department of Animal Sciences, University of Wisconsin-Madison, Animal Science Building, 1675 Observatory Dr., Madison, WI 53706, USA
[b] Departamento de Zootecnia, Escola de Veterinária da Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Caixa Postal 567, Campus da UFMG, CEP 30123-970 Belo Horizonte, MG, Brazil
[c] Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, 5795 Medical Sciences Center, 1300 University Avenue, Madison, WI 53706, USA

ARTICLE INFO

Keywords:
Artificial intelligence
Convolutional neural networks
Crowdsourcing
Aquaculture
Machine learning

ABSTRACT

Individual measurement of traits of interest is extremely important in aquaculture, both for production systems and for breeding programs. Most of the current methods are based on manual measurements, which are laborious and stressful to the animals. Therefore, the development of fast, precise and indirect measurement methods for traits such as body weight (BW) and carcass weight (CW) is of great interest. An appealing way to take noninvasive measurements is through computer vision. Hence, the objectives in the current work were to: (1) devise a computer vision system (CVS) for autonomous measurement of Nile tilapia body area (A), length, height, and eccentricity, and (2) develop linear models for prediction of fish BW, CW, and carcass yield (CY). Images from 1653 fish were taken at the same time as their BW and CW were measured. A set of 822 images had pixels labeled into three classes: background, fish fins, and A. This labeled dataset was then used for training of Deep Learning Networks for automatic segmentation of the images into those pixel classes. In a subsequent step, the segmentations obtained from the best network were used for extraction of A, length, height, and eccentricity. These variables were then used as covariates in linear models for prediction of BW, CW, and CY. A network with an input image of 0.2 times the original size and four encoder/decoder layers achieved the best results for intersection over union on the test set of 99, 90 and 64 percent for background, fish body and fin areas, respectively. The overall best predictive model included A and its square as predictor variables and achieved R^2 of 0.96 and 0.95 for fish BW and CW, respectively. Overall, the devised CVS was able to correctly differentiate fish body from background and fins, and the extracted area of the fish body could be successfully used for prediction of body and carcass weights.

1. Introduction

Fish consumption is growing vertiginously worldwide and, since capture is stagnated with most of the fisheries resources already overexploited, this expansion in consumption is directly related to an increase in aquaculture production. Specifically, aquaculture has grown from 47 million tons in 2006 to 80 million tons in 2016 (FAO, 2018), and thus, is an important protein source for the growing human population. Nevertheless, to continue with such a growth rate, the industry and the academic community need to constantly search for ways to increase productivity while controlling the production environmental impact.

An important tool to increase general aquaculture productivity and to improve traits of interest is genetic selection and breeding strategies. There are several known Nile tilapia breeding programs worldwide and the main focus of those programs is mostly on final body weight, although the ultimate interest is on meat production and quality (Gjedrem et al., 2012). However, direct measurement of meat traits, such as carcass yield and meat quality is impossible on live animals. Possible solutions are via selection based on the performance of

Abbreviations: A, body area; BW, body weight; CI, confidence interval; CNN, convolutional neural network; CV, convolutional layer; CPU, central processing unit; CVS, computer vision system; CW, carcass weight; CY, carcass yield; DL, Deep Learning; E, eccentricity; FCN, fully convolutional network; GPU, Graphics Processing Unit; H, height; IoU, intersection over union; L, length; RAM, random allocation memory

* Corresponding author at: Department of Animal Sciences, University of Wisconsin-Madison, Animal Science Building, 1675 Observatory Dr., Madison, WI 53706, USA.
E-mail address: grosa@wisc.edu (G.J.M. Rosa).

https://doi.org/10.1016/j.compag.2020.105274
Received 19 June 2019; Received in revised form 15 November 2019; Accepted 31 January 2020

relatives or via indirect selection using indicator traits. The selection based on data from relatives has the drawback of reduced selection intensity since some of the fish need to be harvested for data collection. For selection based on indicator traits, its success depends directly on the genetic correlation between the indicator traits and the traits of interest. Previous studies have shown that body and carcass weights are highly correlated with biometric traits, such as height, length, and width (Fernandes et al., 2015; Turra et al., 2018). However, direct measurement of these traits can be time-consuming and cumbersome. Moreover, it can be stressful for the fish, negatively affecting their growth performance and even resulting in increased mortality.

In this context, fast and non-invasive alternatives to measure traits of interest would be of great help to farmers and biologists. Computer vision systems are interesting in this regard since pictures or videos from each animal can be acquired and used to measure several traits remotely. In addition, recent advances in artificial intelligence are showing promising results in areas such as image processing, segmentation, and prediction of characteristics of interest (Goodfellow et al., 2016). However, in the available literature, most of the efforts of using computer vision systems (CVS) in aquaculture are related to fish classification, behavior, counting or measurement of body size (mostly length) (Hao et al., 2016; Saberioon et al., 2017), with very few works on measurement of body area (A) or prediction of body weight. Recent attempts employ traditional image analysis techniques such as the IMAFISH (Navarro et al., 2016) for measurement of body length and height in three marine species; however, it requires the fish to be frozen and with fins trimmed. There are also studies on live fish, such as Asian Seabass (*Lates calcarifer*) (Konovalov et al., 2018) and Halibut (*Hippoglossus stenolepis*) (Wang et al., 2019). Nonetheless, to the best of the authors' knowledge, there is no attempt to develop or apply CVS on Nile tilapia. Also, there is a need for the development of an efficient approach for direct segmentation of fish body from fins for appropriate measurement of fish body area, height and length for fish that have the ability of opening and closing their fins. Therefore, our objective was to devise an automated CVS based on encoder/decoder network for segmentation of fish body from fins and image background, measurement of biometric traits of interest, and prediction on body weight (BW), carcass weight (CW) and yield (CY) of live Nile Tilapia.

2. Material and methods

2.1. Fish breeding and grow-out

A total of 1653 Nile tilapia were used in this study. The fish were produced by mating 43 males and 86 females selected from the broodstock at the Aquaculture Laboratory (LAQUA) from the Veterinary School of the Federal University of Minas Gerais (UFMG), Brazil. The experiment was conducted following rigorous animal handling procedures that are in compliance with federal and institutional regulations regarding proper animal care practices (CONCEA, 2016). The breeding occurred for 58 days, from February to April 2014. During the breeding season, males were placed with 4 females in 1 m^3 tanks for reproduction. After one week of reproduction, females with eggs in their mouth were carefully restrained and had the eggs removed and sent to incubators. The males stayed in the reproduction tanks until they mated with 2 females, thus producing 2 full-sib families. Each full-family stayed in the incubators for approximately one week, after that the fish from the same family were reared in net tanks (*hapas*) until tagging. At approximately 60 days post-hatching, fish from the families closest in age were individually weighed, tagged and grouped for the grow-out phase. For tagging the animals, a passive integrated transponder (PIT) tag was inserted intramuscular close to the dorsal fin. The fish had an average initial BW of 19.0 (± 11.8) g. The grow-out phase had duration of approximately 220 days (min of 140 and max of 300 days). Thus, at harvest, the fish had an average age of 280 (± 36) days of life. Prior to harvest, the fish were again individually weighed, after having feed withdrawn to allow their digestive system to empty its content. In this experiment, the CW was considered as the weight of the fish body without scales and viscera, and CY was defined as the ratio between CW and BW. The fish had average final BW of 523.8 (± 224.4) g, CW of 294.3 (± 126) g, and CY of 56.25 (± 2.89)%.

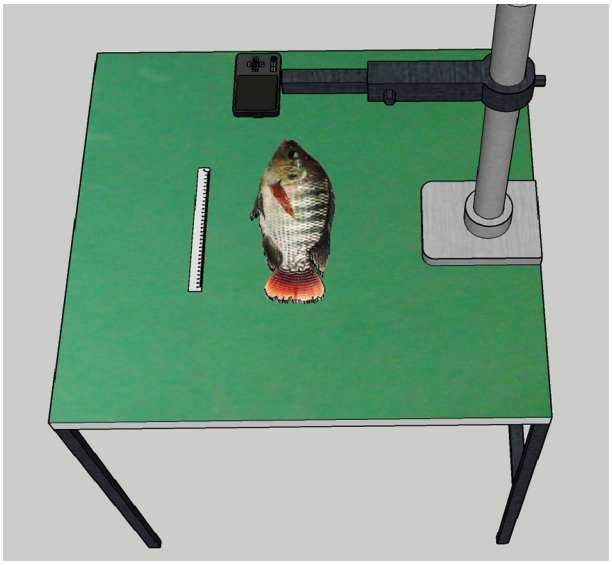

Fig. 1. Example of the station set up for image acquisition with a green background and a digital camera positioned on top of fish at an average distance of 0.5 m. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Image data acquisition and data sets

At the end of the grow-out phase along with with the measurement of BW and carcass traits, pictures were taken of each fish using a Sony DSCWX220 (Sony, Tokyo – JP) digital camera equipped with complementary metal-oxidesemiconductor (CMOS) sensor with a resolution of 18.2 megapixels. The camera was configured for maximum resolution, fast picture mode, and adaptive focus. For stability and improvement of the image acquisition, the camera was fixed at 0.5 m, via a table-top stand mount that was coupled to a table prepared to hold the fish. The table was covered with a green background for maximum contrast with the fish. The table also had rulers fixed on top of the green background for conversion of pixel measures to millimeters (Fig. 1). The Matlab image analysis toolbox (Matlab R2017b) was used for image calibration and acquisition of the pixel length in the metric system. Such calibration was performed via direct measurements made on the ruler length in pixels in the three first images from each time that the station was set up. Moreover, after the background segmentation step (Section 2.3), the length of the segmented base of the camera stand was acquired and the pixel length in centimeters was computed for a double check of the calculated value. At the moment the pictures were taken, the fish had their identification PIT tag read for tracking their identification, and their body weight was measured using an electronic scale.

The dataset (1653 fish with images) was split into two parts, dataset 1 had the 822 fish produced in the first month, whereas dataset 2 had the 831 fish produced in the second month of reproduction. The images from the fish from dataset 1 were used for training and evaluation of the image segmentation models for pixel classification. Thus, these images had a label matrix where each pixel had a final class that could be one of the previously defined classes, i.e. background, fish body or fins. For evaluation of the image segmentation models, the dataset 1 was split into training and testing sets. The training set contained 60% of the images (493 images) and the test set, 40% (329 images). Fish in dataset

2 did not have labeled images and were then used for the validation of the final prediction models for BW, CW, and CY. As a final assessment, the developed segmentation algorithm was tested on 40 images of other species of fish publicly available on the web. These images were either from more related fish species, such as the blue (*Oreochromis aureus* − 7 images), mossambicus (*Oreochromis mossambicus* − 4 images), red (*O. aureus* × *O. mossambicus* − 3 images), zanzibar (*Oreochromis urolepis* − 3 images) and zili (*Coptodon zillii* − 3 images) tilapias, or images from carp species, such as as common (*Cyprinus carpio* − 11 images), grass (*Ctenopharyngodon idella* − 7 images) and big head (*Hypophthalmichthys nobilis* − 4 images) carps.

2.3. Image segmentation and analysis

There were two different image segmentation algorithms used in the present work. The first corresponded to a color threshold algorithm for removal of the background, and identification and segmentation of the whole fish. The algorithm was written and implemented in Matlab R2017b comprised the following steps: (1) downsize of each image to 1944 × 2952 pixels; (2) conversion of images from RGB to CIE 1976 L * a * b space; (3) removal of green background via Global Image Threshold on the image channels based on their histogram distribution; and (4) removal of noise on the resultant black and white images by filling holes in the fish mask and opening the image with a disk kernel with a diameter of 50 pixels (Fig. 2).

The algorithm described above was used to generate cropped images containing only the fish. Then, each cropped image was used in the following steps for labeling of fish body area (A) and fins. The fish A in each image was manually labeled using a web application developed in house. The manual labels of fish A were then grouped with the resultant segmented images from the color threshold method for the construction of a labeled dataset where each pixel in the original image

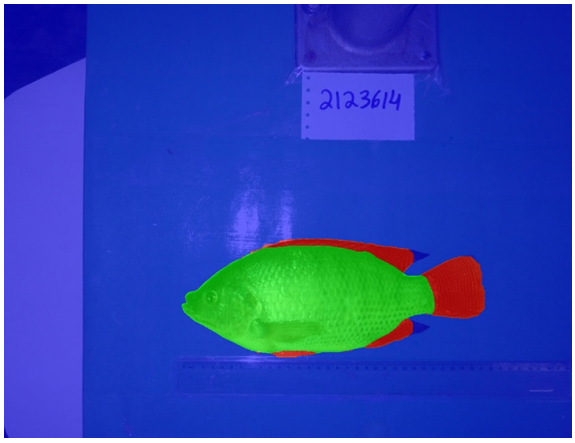

Fig. 3. Example of the labeled image in the crowd-sourcing step, with pixel labels classes specified as background (blue), fish body (green) and fish fins (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was labeled as one of the following classes: background, fish body, or fish fins (Fig. 3). To generate this dataset in a fast and robust manner, the manual image labeling was crowdsourced in Amazon Mechanical Turk (MTurk) (Amazon, Seattle-WA). In the crowdsourcing step, each image was labeled by an anonymous worker, and all labels were later curated and approved by a trained evaluator. If a label was not accepted, the same image was sent back to the crowdsourcing service where it was assigned to a different worker.

The Deep Learning (DL) methods used for image segmentation were also developed in Matlab. These methods are known as segmentation networks, a class of DL models for which the main goal is to classify

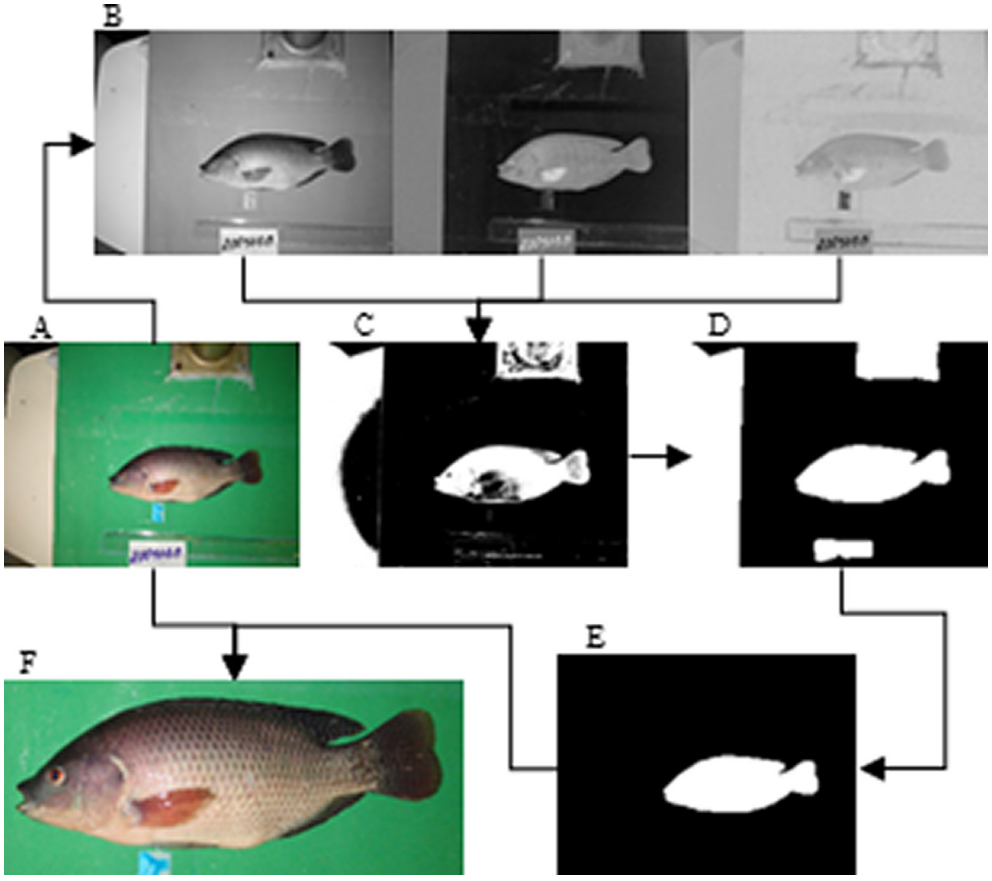

Fig. 2. Representation of background removal and fish identification. (A) Original image. (B) The channels of the image converted to the CIE 1976 L * a * b color space. (C) Global thresholding on the LAB channels. (D) Morphological image opening and filling. (E) The binary mask of the identified fish. (F) Fish cropped from the original image using the binary mask. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

every input node properly. More specifically, the image segmentation models developed in the current study apply an encoder-decoder architecture that is based on the SegNet approach (Badrinarayanan et al., 2015). The segmentation networks evaluated had encoder layers that were composed of a convolutional layer (CV1), batch normalization (BN1), rectified linear unit (ReLu1), convolutional layer (CV2), batch normalization (BN2), rectified linear unit (ReLu2), and a max-pooling layer (MP). Therefore, each encoder had a total of 7 layers. For each encoder in the model, there was a respective decoder that had a max un-pooling layer (MU), a convolutional layer (CV2), batch normalization (BN2), rectified linear unit (ReLu2), convolutional layer (CV1), batch normalization (BN1), and rectified linear unit (ReLu1). Finally, a fully connected layer and a final pixel-wise classification layer were added to the networks, coupled to the last decoder layer. Each network was trained for 100 epochs, with mini-batch size of 4 samples and data shuffled every epoch. The optimization algorithm used was a stochastic gradient descending with momentum of 0.9 and initial learning rate of 10^{-3}. Also, L2 regularization with a shrinkage factor of 5^{-4} was used. To improve training performance, the training set was augmented using image rotation, reflection, and translation. The augmentation was performed during the training phase to increase the diversity of the data and provide more examples to the network.

The networks evaluated diverged in the size of the input image, which could be 0.1, 0.2, 0.3 and 0.4 times of the original size (1944 × 2592 pixels), and the number of complex encoder-decoder layers, which ranged from 1 to 5. For evaluation of the predictive quality of pixel classification, the overall prediction accuracy and the intersection over union (IoU) ratio were calculated. The IoU is defined by, $IoU_{A,B} = (A \cap B)/(A \cup B)$, where A and B represent the predicted and expected areas. Also, the networks were compared on usage of computational resources, such as disk space for model allocation (MB), computational time (s) and random allocation memory (RAM) (MB) required for evaluation of new frames.

From the segmented images, the value of each pixel in the images was linearly transformed to centimeters by the measurement of the correspondent distance in pixels from the ends of the ruler affixed in the measurement table. Fish body area (A), length (L), and height (H) in centimeters were then calculated using this linear transformation. The eccentricity (E) of the fish body was calculated as the ratio between the foci and the major axis length of the ellipsis with the same second moments (major and minor axis) as the fish body area (Fig. 4).

2.4. Statistical analysis

Estimates of *Pearson's* correlation between body measurements and body weight (BW), carcass weight (CW) and carcass yield (CY) were computed. Multiple linear regression models were evaluated for prediction of BW, CW, and CY. These models differed in terms of the inclusion of specific combinations of the body measurements, their squared values, and possible interactions. The search for the best possible model on the training set was performed using the Bayesian Information Criteria (BIC) to rank the models and a genetic algorithm search (Calcagno and de Mazancourt, 2010). In this analysis, the fish body predicted from the best segmentation model (model with depth 4 and input image size of 20%) developed in the previous section was used as input for automated acquisition of fish body measurements. For this purpose, the segmentation model was retrained using the full dataset 1, and the predicted segmented fish body from both datasets were used. The predictive models were fitted using only the fish that had information on BW, CW and CY from dataset 1 (that were 764 fish labeled images) and evaluated on dataset 2 (that were 765 fish). These statistical analyses were implemented in R, using the packages MASS, glmulti and ggplot2.

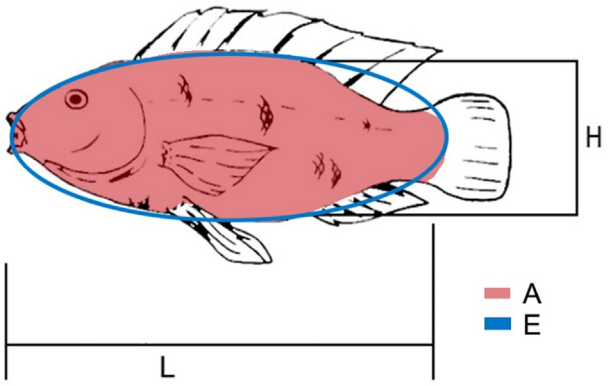

Fig. 4. Extracted body measurements from image: body area (A), height (H), length (L) and ellipsis (E) that has the same second moments of the body area.

3. Results

3.1. Semantic segmentation

Fig. 5 presents the mean accuracy and intersection over union (IoU) and their respective 95% confidence interval (CI) for the segmentation networks evaluated. The results show an increase in both overall accuracy and IoU as the depth of the network increased, reaching a maximum at depth 4. Similarly, an increase in IoU was also observed for each of the three different pixel classes when evaluated separately, as the depth of the model increased (Fig. 6). However, the background and body had higher IoU with the best models achieving 0.99 and 0.90, respectively, while for fins the best models achieved at maximum 0.65.

The differences in classification performance between fish fin versus body or background were expected. This is mostly because the fin class had the lowest number of pixels and it was also very difficult to differentiate the boundaries between fish fins and body or in some cases even with the background, even for human evaluators. Fig. 7 presents examples of the labeled images with their correspondent pixel classification image. By visual inspection of the resulting pixel classifications versus the input label image it is possible to observe that background pixels near to the fin or on the border of the fish were more prone to be erroneously classified as fin (Fig. 7.2 and 7.3). Also, patches of the background that had more texture due to water splashes or higher light reflection could be incorrectly classified as fish (Fig. 7.2, 7.3 and 7.4).

Regarding the images of other fish species retrieved from the web, the results were, in general, worse than what was observed for the Nile tilapia dataset, with an average accuracy of 0.57 and IoU of 0.52, 0.34 and 0.1 for background, body, and fins respectively (Table 1). This result is reflected in the visual inspection of the images (Fig. 8). It can be observed that images with fish that had either body shape or color different from the fish used to generate the models and/or with a more complex background, such as grass or a fishnet, presented the worse predictions.

In the evaluation of computational resources used, there was an overall increase in disk and RAM usage along with computational time as the model complexity increased (Table 2). The main factor in increasing the demand for computational resources was the size of the input image. Also, the evaluation using graphics processing unit (GPU) was overall 20 to 40 times faster but used from 1.5 to 10 times more RAM than with central processing unit (CPU). It is also worth noting that the usage of RAM increased at a steeper rate on CPU evaluations compared to GPU. Thus, RAM usage on a CPU-based version of the model probably would surpass the RAM usage of the GPU version for more complex models.

3.2. Image derived body measurements and body weight prediction

It was observed that the pixel length ranged from 0.021 to 0.026 cm, due to the difference in the height that the camera was

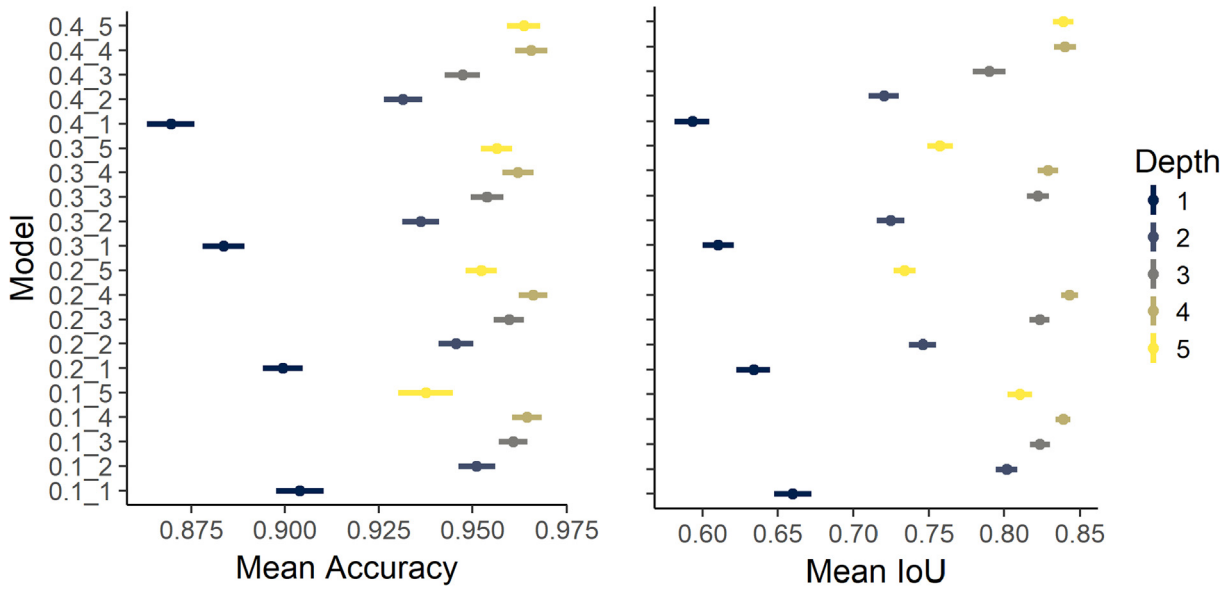

Fig. 5. Average model accuracy and intersection over union (IoU) and respective 95% CI for an input image of 0.1–0.4 times of the original size and encoder depth varying from 1 to 5 stacks of layers on the test dataset.

attached in each day of measurement. However, there was no difference in the pixel length at the plane of the table for fish measured on the same day. Thus, the final body measurements extracted from each image were rescaled to the metric system using the corresponding pixel length in cm.

Table 3 shows the estimated *Pearson's* correlation between body measurements, BW, CW and CY for both datasets. Correlations between body measurements and BW or CW were positive and high, ranging from 0.90 for L and CW in dataset 1 to 0.98 for A and BW in dataset 2. However, E presented moderate to low negative correlations with the other variables in dataset 1, while in dataset 2 the correlations were from low to moderate. Estimates of correlations between CY and the other variables were mostly low and positive (apart from E) in both datasets.

Regarding the predictions of BW and CW, it was observed that the simpler model that included only A as predictor variable had similar performance on the testing dataset than more complex models selected from the genetic algorithm (Table 4 and Fig. 9). The overall best model had A and A^2 as predictors and achieved predicted square correlations (R^2) of 0.96 and 0.95 and mean absolute percentage errors (MAPE) of 11.35 and12.35, for BW and CW, respectively. On the other hand, the MAPE for the model including only A as predictor variable was 4.09 for prediction of CY. However, the predicted R^2 was very low for every predictive model evaluated for CY, showing that there is not a significant difference from a model that included only the intercept.

4. Discussion

In the current paper, a CVS for biometric measurement of fish bodies and prediction of fish BW, CW, and CY was presented. To the best of the author's knowledge, this is the first work to present such results for Nile tilapia. However, there are previous attempts in the literature to develop automated CVS for other fish species, such as seabass (*Dicentrarchus labrax, L.*) (Costa et al., 2013), olive flounder (*Paralichthys olivaceus*) (Jeong et al., 2013), gilthead seabream (*Sparus aurata L.*), meagre (*Argyrosomus regius*) and red porgy (*Pagrus pagrus*) (Navarro et al., 2016). These studies focused on the development of an automated CVS for indirect measurement and prediction of traits of interest. However, none of the previously presented methods was developed for live animals, even if removed from the water, as in the current study. Moreover, the images were taken in an environment with controlled illumination conditions, and for some species (e.g., seabass, meagre, and porgy) the fins were manually cut to improve predictions. Thus, the CVS presented in this manuscript is an improvement to these

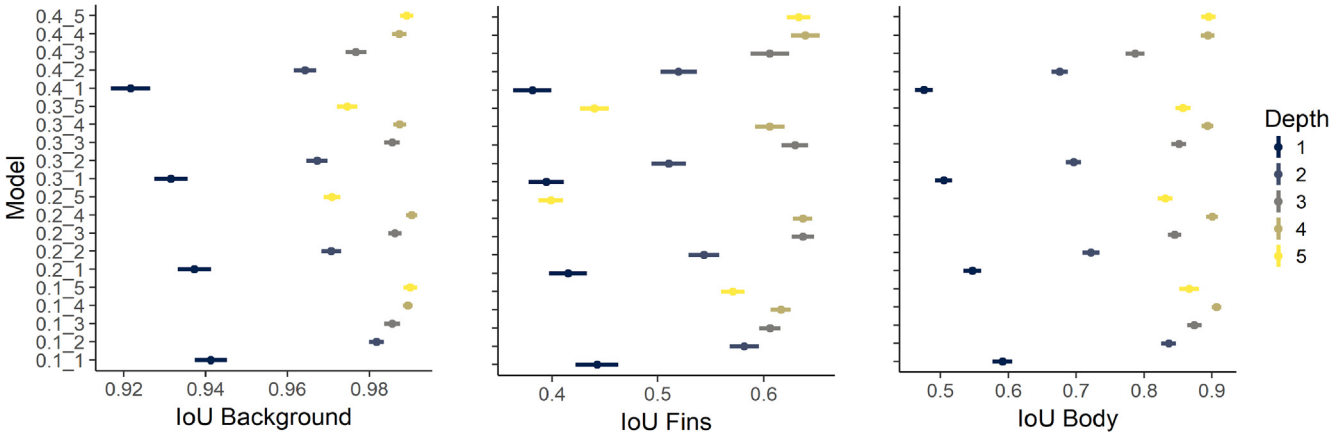

Fig. 6. Average intersection over union (IoU) for the 3 classes (Background, Fins, and Body), and respective 95% CI for an input image of 0.1–0.4 times of the original size and encoder depth from 1 to 5 stacks of layers on the test dataset.

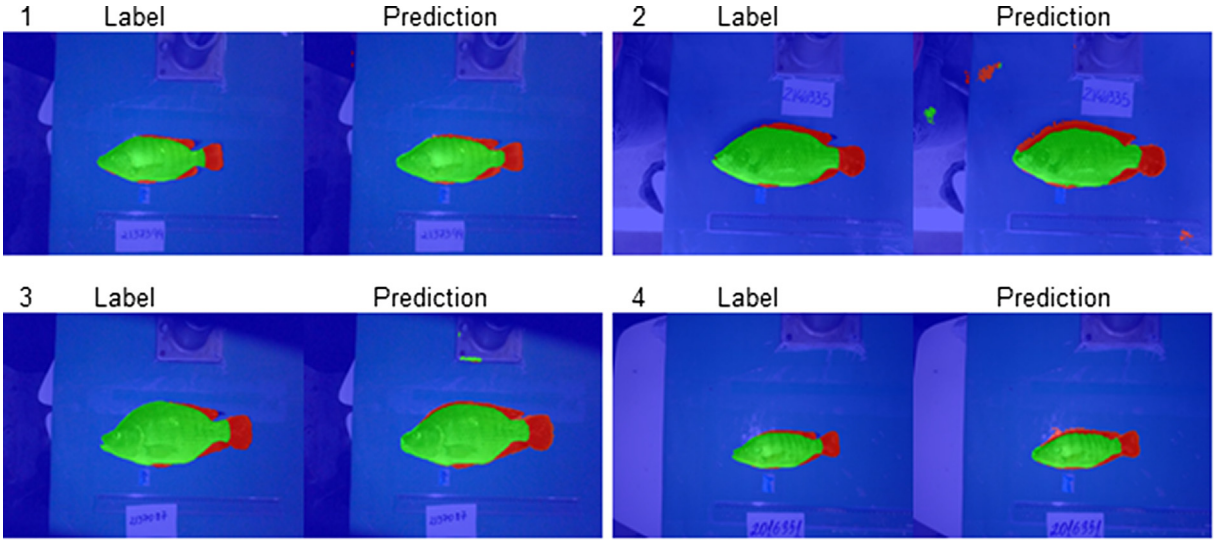

Fig. 7. Randomly selected examples of fish images from the test dataset showing the expected result (Label) and model output predicted pixel classification (Prediction) with the background, fish body and fins coded as blue, green and red respectively.

Table 1
Average and 95% confidence interval (CI) for accuracy (Acc) and intersection over union (IoU) for the best segmentation network on 40 fish images retrieved from the web.

	Acc	Mean IoU	IoU Background	IoU Fins	IoU Body
Average	0.573	0.318	0.526	0.100	0.340
95% CI	± 0.068	± 0.049	± 0.083	± 0.031	± 0.063

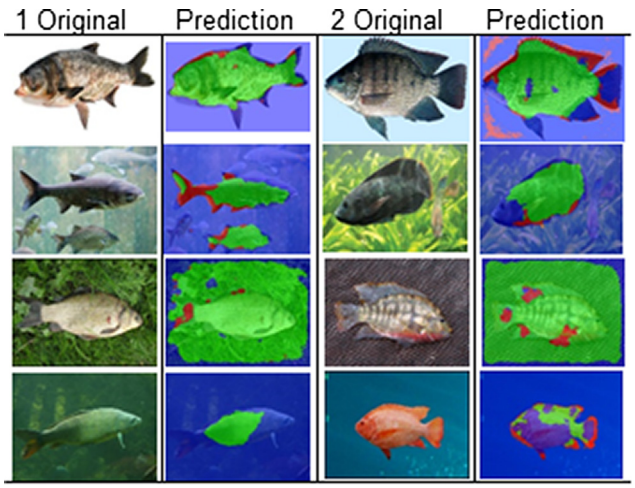

Fig. 8. Randomly selected examples of carp (1) and tilapia (2) images retrieved from the web showing the original image and model pixel classification (Prediction) with the background, fish body and fins coded as blue, green and red respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

previous methods, because the images of live fish were taken on multiple sites without the aid of controlled illumination. The only other work found in the literature that attempted to automatically remove fins from fish body and to predict BW was with Asian Seabass (Konovalov et al., 2018). In their study, a DL for extraction of fish body area was trained on 200 fish and the extracted area from images of 1072 fish (BW ranging from 200 to 2700 g) was evaluated for prediction of BW, achieving an R^2 of 0.98.

A drawback of the proposed CVS is that it still requires the removal of fish from the water to set them on a table covered with a green background for the acquisition of the images. Moreover, as expected, the models developed are applicable only for Nile tilapia on a flat surface, as it became evident after the evaluation of the semantic segmentation on images from other fish species and in different background conditions. For the 40 images of other fish species, the average accuracy dropped to 0.57 and the IoU to 0.32 (Table 1). By visual inspection of such images (Fig. 8) it is possible to evaluate that the best results were achieved if the fish were on a more homogeneous background compared to texturized backgrounds, such as grass or inside a net. It is also worth noticing that the model performs better for fish with a body color that is on the dark brown palette, similar to the Nile tilapia, since the best results were for the top images of Fig. 8 representing a blue tilapia while the red tilapia in the bottom image was not identified.

4.1. Semantic segmentation

From the several existing image classification and segmentation methods, Deep Learning methods have gained notoriety in recent years since they are showing better performance in tackling many problems of interest in the Machine Learning community (Goodfellow et al., 2016). The most classic and first successful application of such models was in optic character recognition (Lecun et al., 1998). In this previous study, a special case of an artificial neural network model known as a convolutional neural network (CNN) or fully convolutional network (FCN) was developed. The basic structure of such a model is the stack of convolutional, and pooling layers followed by a series of fully connected layers and a final classification layer. The motivation for convolution and pooling operations is to include sparsity in the network, parameter sharing, and to account for small translational variances (Goodfellow et al., 2016). However, FCN has the drawback of, generally, demanding extremely large datasets for the training process and to reduce the size of the image after each set of convolution and pooling operations. To address these drawbacks, adaptations and extensions for semantic segmentation have been developed. The models known as encoder-decoder networks have presented an improvement to the overall quality of the prediction by coupling the FCN to another set of layers known as deconvolution network; examples of such models are U-Net and Segnet (Badrinarayanan et al., 2015; Ronneberger et al., 2015). The function of the deconvolution layers is to rescale the network output to the original image size without losing the information gained during the previous convolutional layers. However, SegNet has an advantage over other similar methods, in that the decoder layers

Table 2
Comparison of the different Deep Learning architectures evaluated on image size (pixels), number (N) of encoders, layers and weights, and computational resources as disk space for model allocation (MB), computational time (s) and random allocation memory (RAM) (MB) required on evaluation step to predict each pixel in a single image on GPU or CPU.

Image size	N Encoders	N Layers	N Weights	N Nodes	Model Size (MB)	GPU[a] RAM (MB)	GPU Time (s)	CPU[b] RAM (MB)	CPU Time (s)
200 × 237	1	17	77,967	398,364	1.8	1202.46	0.06	91	1.1
	2	31	226,447	546,844	2.84	1511.664	0.07	99	1.66
	3	45	374,927	695,324	3.89	1322.706	0.09	105	1.81
	4	59	523,407	843,804	4.93	1288.35	0.1	117	1.82
	5	73	671,887	992,284	5.97	1202.46	0.11	125	1.82
390 × 520	1	17	77,967	1,294,764	4.4	1717.8	0.15	270	4.35
	2	31	226,447	834,844	5.08	1889.58	0.19	357	6.3
	3	45	374,927	983,324	6.13	1803.69	0.2	405	6.8
	4	59	523,407	1,131,804	7.17	1803.69	0.21	465	6.9
	5	73	671,887	1,280,284	8.25	1631.91	0.23	461	6.9
585 × 780	1	17	77,967	1,446,864	7.67	1975.47	0.3	750	10.7
	2	31	226,447	1,595,344	8.71	1975.47	0.38	1100	15.75
	3	45	374,927	1,743,824	9.75	2061.36	0.39	1240	16.5
	4	59	523,407	1,892,304	10.8	2147.25	0.4	1230	16.85
	5	73	671,887	2,040,784	11.8	2147.25	0.42	1300	17
780 × 1040	1	17	77,967	2,511,564	12.5	2576.7	0.57	1640	24.5
	2	31	226,447	2,660,044	10.9	3006.15	0.79	2207	31
	3	45	374,927	2,808,524	12	3006.15	0.8	2268	33.5
	4	59	523,407	2,957,004	13	3092.04	0.82	2245	33.7
	5	73	671,887	3,105,484	14.1	3092.04	0.84	2272	33.7

[a] The graphic processing unit (GPU) used in this study was an NVIDIA Quadro M4000.
[b] The central processing unit (CPU) used in the current study was an Intel Core i7-7700K.

share the same feature map indexes with the encoder layers. This design generates networks that have a high reduction in the number of parameters (Badrinarayanan et al., 2015).

Despite the fact that the DL models used are state of the art, they are still considerably complex with several layers in each encoder or decoder stack. Moreover, the difference in the size of the input image increased the network complexity, because it directly affects the size of the input and output layers. Table 2 presents these differences in computational resources required between the networks evaluated for prediction of pixel classes in a single image. As the networks increase in complexity, it also will increase the demand for computational resources, more accentuated in RAM and also in processing time if using only a CPU. In the current application, the main factor driving the increased demand for computational resources was the size of the image used as input.

This analysis is important for the deployment of the CVS since it directly affects the hardware requirement. For instance, the model that has an input image 20% of the original size (390 × 520 pixels) and 4 encoder stacks had similar to better accuracy and IoU than more complex models (Figs. 5 and 6). On the other hand, if the models were compared based on their computational performance, the selected model (input of 390 × 520 and 4 encoder stacks) uses on average 2 times less RAM and it is 4 times faster on graphics processing unit (GPU) and uses 5 times less RAM and it is 5 times faster on the CPU than the more complex models. Therefore, this compact smaller model is more useful for deployment in embedded or mobile systems, which can have a positive impact on the fish industry by potentially reaching more farmers. This concept had been already applied for other DL models (Lane and Georgiev, 2015). Thus, a more complex and robust model can possibly be deployed in embedded systems if it proved more accurate. Such a model may be important if there is a need to improve classification or to include different fish species in the model as well as different background scenarios. However, for the current application, the real interest is to correctly separate the Nile tilapia body for the estimation of body measurements and prediction of BW, CW, and CY, thus the lower IoU for fish fins is not of major concern. It is also worth noting that maybe the dataset used in this study is not large and diverse enough for training more complex models.

In the current study, data augmentation and batch normalization were applied during the training of the segmentation network as a way to increase regularization, and to improve learning performance and stability of the networks (Goodfellow et al., 2016). An approach to further improve the training would be via other data augmentation techniques, such as synthetic data generated via rendering of computer graphics models (Abu Alhaija et al., 2018), thus creating a more diverse, simulated dataset for training. It is also important to notice that a current trend in computer sciences when working with small datasets and/or limited computing resources is to use transfer learning strategies. The basic idea is to reuse a proven network model developed and trained in a larger and more complex data set for initialization of the

Table 3
Pearson's correlation between fish body area (A), length (L), height (H), eccentricity (E), body weight (BW), carcass weight (CW) and carcass yield (CY) on dataset 1 (below diagonal) and on dataset 2 (above diagonal).

	A	L	H	E	BW	CW	CY
A	–	$0.97^{(0.00)}$	$0.97^{(0.00)}$	$-0.40^{(0.00)}$	$0.98^{(0.00)}$	$0.97^{(0.00)}$	$0.08^{(0.02)}$
L	$0.97^{(0.00)}$	–	$0.92^{(0.00)}$	$-0.20^{(0.00)}$	$0.93^{(0.00)}$	$0.93^{(0.00)}$	$0.06^{(0.11)}$
H	$0.98^{(0.00)}$	$0.92^{(0.00)}$	–	$-0.55^{(0.00)}$	$0.95^{(0.00)}$	$0.94^{(0.00)}$	$0.07^{(0.04)}$
E	$-0.56^{(0.00)}$	$-0.37^{(0.00)}$	$-0.69^{(0.00)}$	–	$-0.40^{(0.00)}$	$-0.40^{(0.00)}$	$-0.06^{(0.11)}$
BW	$0.97^{(0.00)}$	$0.92^{(0.00)}$	$0.96^{(0.00)}$	$-0.63^{(0.00)}$	–	$0.99^{(0.00)}$	$0.07^{(0.06)}$
CW	$0.96^{(0.00)}$	$0.90^{(0.00)}$	$0.95^{(0.00)}$	$-0.63^{(0.00)}$	$0.99^{(0.00)}$	–	$0.18^{(0.00)}$
CY	$0.21^{(0.00)}$	$0.18^{(0.00)}$	$0.22^{(0.00)}$	$-0.21^{(0.00)}$	$0.22^{(0.00)}$	$0.34^{(0.00)}$	–

Superscripts are the respective p-values.

Table 4
Comparison of selected models for predictions of body weight (BW), carcass weight (CW) and carcass yield (CY) accounting for area (A), length (L), height (H) and eccentricity (E).

		Training Dataset					Test Dataset				
Target	Predictors	R^2 [a]	MAPE[b]	MAE[c]	RMSE[d]	SE[e]	R^2	MAPE	MAE	RMSE	SE
BW	A	0.95	6.72	35.47	48.26	1.75	0.95	14.80	62.34	77.45	2.80
	$A^{1.5}$	0.95	5.87	32.84	45.17	1.63	0.96	11.37	58.53	78.83	2.85
	A^2	0.95	6.62	34.57	46.60	1.69	0.95	13.49	61.40	83.20	3.01
	$A + A^2$	**0.95**	**5.84**	**32.56**	**44.81**	**1.62**	**0.96**	**11.35**	**57.77**	**77.52**	**2.80**
	A + LxE	0.96	5.69	30.88	42.81	1.55	0.95	12.97	60.33	79.78	2.88
	A + H + HxL + AxE	0.96	5.20	28.69	39.53	1.43	0.95	11.87	59.62	81.98	2.96
	$H + A^2xH + A^2xL^2 + H^2xL + H^2xE + H^2xA^2$	0.96	5.18	28.47	39.31	1.42	0.95	11.91	58.30	79.37	2.87
CW	A	0.92	8.38	25.36	34.53	1.25	0.94	16.87	39.44	49.43	1.79
	$A^{1.5}$	0.93	7.26	23.26	32.35	1.17	0.95	12.60	35.88	48.14	1.74
	A^2	0.93	7.60	23.49	32.46	1.17	0.95	13.92	37.53	52.22	1.89
	$A + A^2$	**0.93**	**7.22**	**23.04**	**32.05**	**1.16**	**0.95**	**12.35**	**36.07**	**49.98**	**1.81**
	A + LxE	0.94	7.21	22.33	30.67	1.11	0.94	14.42	38.17	51.39	1.86
	A + H + HxL + AxE	0.95	6.50	20.57	28.65	1.04	0.94	13.26	37.96	53.63	1.94
	$H + A^2xH + A^2xL^2 + H^2xL + H^2xE + H^2xA^2$	0.95	6.50	20.48	28.52	1.03	0.94	13.19	36.78	51.58	1.86
CY	A	0.05	3.96	2.21	2.78	0.10	0.01	4.09	2.27	2.84	0.10
	$A^{1.5}$	0.05	3.96	2.21	2.78	0.10	0.01	4.09	2.27	2.84	0.10
	A^2	0.05	3.96	2.21	2.77	0.10	0.01	4.10	2.28	2.84	0.10
	$A + A^2$	0.05	3.96	2.21	2.77	0.10	0.01	4.11	2.28	2.84	0.10
	A + LxE	0.06	3.94	2.20	2.76	0.10	0.01	4.10	2.28	2.84	0.10
	A + H + HxL + AxE	0.06	3.94	2.20	2.75	0.10	0.01	4.13	2.29	2.87	0.10
	$H + A^2xH + A^2xL^2 + H^2xL + H^2xE + H^2xA^2$	0.06	3.94	2.20	2.75	0.10	0.00	4.11	2.29	2.87	0.10

[a] Squared Correlation.
[b] Mean absolute percentage error.
[c] Mean Absolute Error.
[d] Residual Mean Square Error.
[e] Residual Standard Error.

weights, or as part of a new model for a specific case. Examples in computer vision are FCN as ImageNet and VGG (Krizhevsky et al., 2012; Simonyan and Zisserman, 2015) that were developed using large and general purpose datasets. These datasets are composed of hundreds of thousands of labeled images from diverse scenes and applications. The models developed for such general datasets have been fully adapted for other problems (Konovalov et al., 2018; Wang et al., 2019; Zhou et al., 2017) or used as part of the network in novel architectures. Therefore, the use of transfer learning and other data augmentation approaches may be useful for the case of including other fish species and different backgrounds.

4.2. Image derived body measurements and body weight prediction

It is well known that BW has a strong correlation with morphometric traits for many animal species. Previous studies with Nile tilapia shows that phenotypic and genetic correlations between BW or CW with morphometric traits tend to be positive and high, while for CY with other traits tend to be positive but low to moderate (He et al., 2017; Turra et al., 2018). The results for phenotypic correlations presented in the current work (Table 3) agree with previous results for Nile tilapia, mainly with A being the trait that achieved the best results. These correlation results are reflected in the predictive linear models evaluated because the simple model that included only A as a predictor variable achieved high predictions for both BW and CW on the test dataset (Table 4). In fact, the model that included also squared A achieved a slightly better predictive accuracy than including only A, and presented similar performance compared to the more complex model retrieved by the genetic algorithm. In fact, even though the genetic algorithm selected models including more morphometric traits, and interaction terms the results show that the improvement in accuracy achieved on the training set is not observed for the test dataset, which is indicative of overfitting for both BW and CW. It is also important to notice that the model including only $A^{1.5}$ had been evaluated before in a study for Asian seabass (Konovalov et al., 2018), resulting in MAPE of about 5% for fish with BW ranging from 200 g to 2.7 kg, which is close to the 6% reported in the present study. Therefore, the results showed the potential for the application of CVS in the Nile tilapia industry. The possibility of prediction of BW and CW on live animals without the need of weighing the animals on a scale could speed-up the measurement process since several traits of interest can be measured at the same time with the proposed CVS. Regarding the predictions for CY, even though the MAPE, MAE and the root mean square error (RMSE) were low, the R^2 was also low. Some hypotheses for such results are: (1) The small variation observed for CY could be due to random noise or simply variation in processing (cleaning of the carcass); (2) It is possible that CY does not have indeed any relationship to the variables evaluated, and some other variables would need to be tested for prediction of CY, such as body volume and widths, which unfortunately could not be evaluated with this dataset.

It is important to note that the developed methodology is applicable only to Nile tilapia, and the images need to be taken using a homogeneous green background. As an ultimate goal, it would be interesting to have a CVS where there is no need for handling the animals. An alternative in this regard would be by the implementation of a CVS in association to a chute for fish management (Wang et al., 2019), reducing labor effort. Another strategy would be with images acquired underwater, without the need for interfering with the fish. Such an idea is not new and there are some attempts that have been made for other fish species, with the more successful ones for Tuna (Muñoz-Benavent et al., 2018; Shafait et al., 2017). However, all those applications use restrictive body shape models and thus can only be applied for the species it was developed. Meanwhile, we predict that the image segmentation methodology described in this study can be extrapolated to underwater images and also for other species. The main constraint for such is the lack of a database rich enough for training such a model. Such database would need to include images from different fish species, in different poses and under different water/system condition. Future directions on the development of such underwater CVS should also evaluate other imaging technologies and sensors, such as stereo imaging, time of flight,

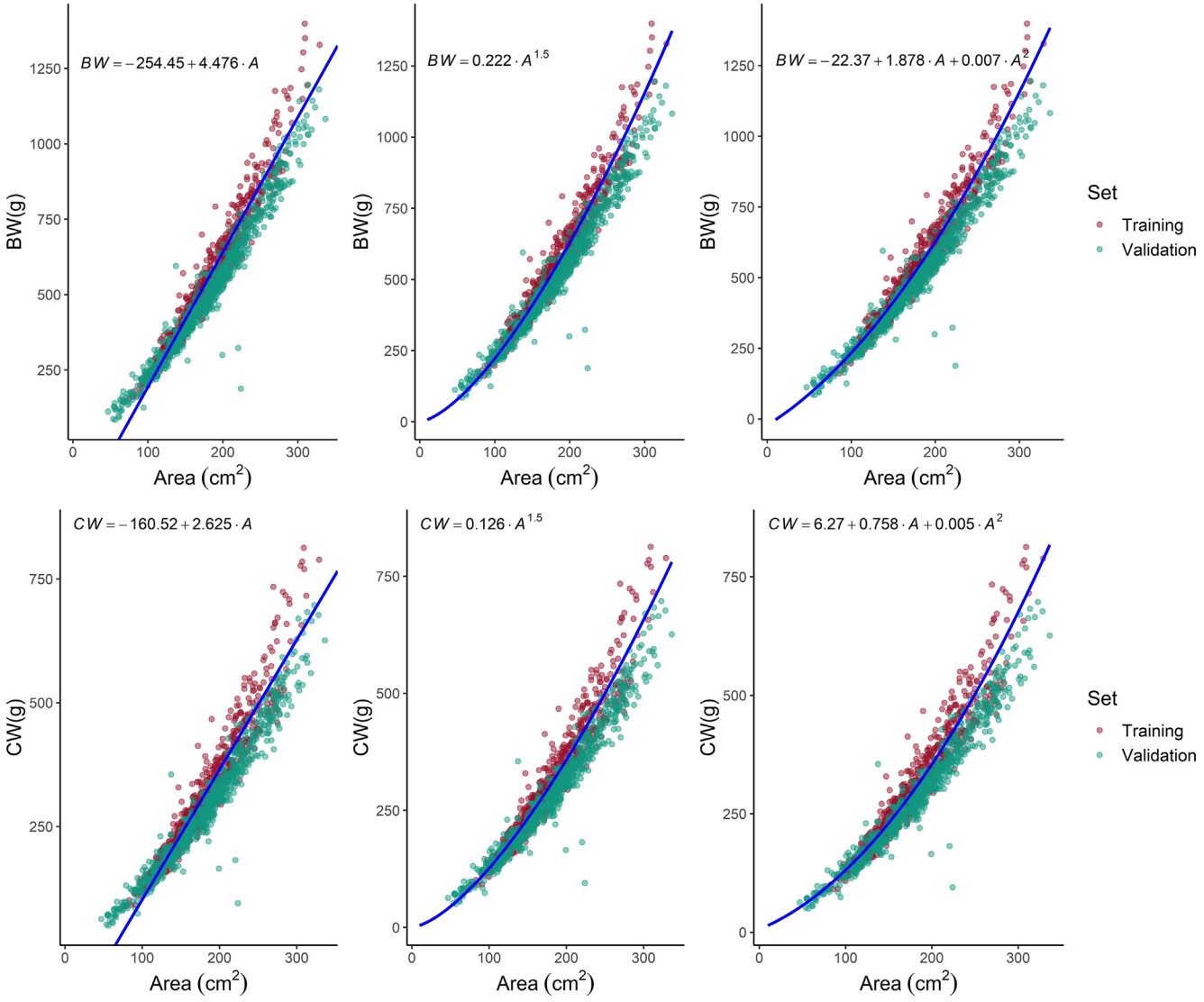

Fig. 9. Visual evaluation of goodness of fit for live body weight (BW) and carcass weight (CW) for models that considered only the segmented fish body area (A); the blue line in each figure represents the predicted regression line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and sonars, since different technologies may be better suited to different water conditions (e.g., salinity and turbidity). Also, an apparent body volume can be estimated with 3D sensors, which may be more correlated to BW than A as shown in pigs (Fernandes et al., 2019; Kongsro, 2014). Although 3D image systems are still in general more expensive and more difficult to implement than those based on 2D cameras, the technology is constantly evolving and access to 3D-based CVS should improve. In any case, for any CVS application to be successful in aquaculture, its implementation cost must be justified by its benefits, may they be in terms of increased productivity, improved health assessment, or facilitated management practices.

5. Conclusion

In the current paper, we presented the application of DL models for the development of a computer vision system for measurement of live Nile tilapia morphometric traits and prediction of BW and CW. The current method achieved high predictive accuracy, with R^2 of 0.96 and 0.95 for BW and CW respectively, while measuring other traits of interest at a satisfactory speed (0.21 s for evaluation of a single image for the best segmentation model). However, the current CVS is still not general enough to be applied to other fish species or with more texturized backgrounds.

CRediT authorship contribution statement

Arthur F.A. Fernandes: Conceptualization, Methodology, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Eduardo M. Turra:** Conceptualization, Resources, Writing - original draft. **Érika R. de Alvarenga:** Conceptualization, Resources, Data curation, Writing - original draft. **Tiago L. Passafaro:** Conceptualization, Methodology, Writing - original draft, Visualization. **Fernando B. Lopes:** Methodology, Writing - original draft. **Gabriel F.O. Alves:** Conceptualization, Data curation, Writing - original draft. **Vikas Singh:** Conceptualization, Methodology, Writing - original draft. **Guilherme J.M. Rosa:** Conceptualization, Resources, Methodology, Writing - original draft, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research received support from CAPES – Science without Borders and CNPq (National Council for Scientific and Technological Development, Brazil - 472081/2012-8 and 472149/2012-1).

Declaration of Competing Interest

None.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2020.105274.

References

Abu Alhaija, H., Mustikovela, S.K., Mescheder, L., Geiger, A., Rother, C., 2018. Augmented reality meets computer vision: efficient data generation for urban driving scenes. Int. J. Comput. Vis. 126, 961–972. https://doi.org/10.1007/s11263-018-1070-x.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2015. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. arXiv.

Calcagno, V., de Mazancourt, C., 2010. glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models. J. Stat. Softw. 34, 1–29. https://doi.org/10.18637/jss.v034.i12.

CONCEA, 2016. Normativas do CONCEA para produção, manutenção ou utilização de animais em atividades de ensino ou pesquisa científica, 3rd ed. Conselho Nacional de Controle de Experimentação Animal, Brasília, Brazil.

Costa, C., Antonucci, F., Boglione, C., Menesatti, P., Vandeputte, M., Chatain, B., 2013. Automated sorting for size, sex and skeletal anomalies of cultured seabass using external shape analysis. Aquac. Eng. 52, 58–64. https://doi.org/10.1016/J.AQUAENG.2012.09.001.

FAO, 2018. The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals. FAO, Rome.

Fernandes, A.F.A., Dórea, J.R.R., Fitzgerald, R., Herring, W., Rosa, G.J.M., 2019. A novel automated system to acquire biometric and morphological measurements and predict body weight of pigs via 3D computer vision. J. Anim. Sci 97, 496–508. https://doi.org/10.1093/jas/sky418.

Fernandes, A.F.A., Silva, M. de A., Alvarenga, E.R. de, Teixeira, E. de A., Silva Junior, A.F. da, Alves, G.F. de O., Salles, S.C.M. de, Manduca, L.G., Turra, E.M., 2015. Morphometric traits as selection criteria for carcass yield and body weight in Nile tilapia (Oreochromis niloticus L.) at five ages. Aquaculture 446, 303–309. https://doi.org/10.1016/j.aquaculture.2015.05.009.

Gjedrem, T., Robinson, N., Rye, M., 2012. The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. Aquaculture 350–353, 117–129. https://doi.org/10.1016/J.AQUACULTURE.2012.04.008.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Cambridge, MS.

Hao, M., Yu, H., Li, D., 2016. The Measurement of Fish Size by Machine Vision - A Review, in: Computer and Computing Technologies in Agriculture IX. CCTA 2015. Springer, Cham, pp. 15–32. doi: 10.1007/978-3-319-48354-2_2.

He, J., Zhao, Y., Zhao, J., Gao, J., Han, D., Xu, P., Yang, R., 2017. Multivariate random regression analysis for body weight and main morphological traits in genetically improved farmed tilapia (*Oreochromis niloticus*). Genet. Sel. Evol. 49, 80. https://doi.org/10.1186/s12711-017-0357-7.

Jeong, S.-J., Yang, Y.-S., Lee, K., Kang, J.-G., Lee, D.-G., 2013. Vision-based automatic system for non-contact measurement of morphometric characteristics of flatfish. J. Electr. Eng. Technol. 8, 1194–1201. https://doi.org/10.5370/JEET.2013.8.5.1194.

Kongsro, J., 2014. Estimation of pig weight using a Microsoft Kinect prototype imaging system. Comput. Electron. Agric. 109, 32–35. https://doi.org/10.1016/j.compag.2014.08.008.

Konovalov, D.A., Saleh, A., Domingos, J.A., White, R.D., Jerry, D.R., 2018. Estimating Mass of Harvested Asian Seabass Lates calcarifer from Images. World J. Eng. Technol. 06, 15–23. https://doi.org/10.4236/wjet.2018.63B003.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Neural Inform. Process. Syst. 1097–1105.

Lane, N.D., Georgiev, P., 2015. Can Deep Learning Revolutionize Mobile Sensing? In: Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications - HotMobile '15. ACM Press, New York, New York, USA, pp. 117–122. doi: 10.1145/2699343.2699349.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324. https://doi.org/10.1109/5.726791.

Muñoz-Benavent, P., Andreu-García, G., Valiente-González, J.M., Atienza-Vanacloig, V., Puig-Pons, V., Espinosa, V., 2018. Enhanced fish bending model for automatic tuna sizing using computer vision. Comput. Electron. Agric. 150, 52–61. https://doi.org/10.1016/J.COMPAG.2018.04.005.

Navarro, A., Lee-Montero, I., Santana, D., Henríquez, P., Ferrer, M.A., Morales, A., Soula, M., Badilla, R., Negrín-Báez, D., Zamorano, M.J., Afonso, J.M., 2016. IMAFISH_ML: A fully-automated image analysis software for assessing fish morphometric traits on gilthead seabream (Sparus aurata L.), meagre (Argyrosomus regius) and red porgy (Pagrus pagrus). Comput. Electron. Agric. 121, 66–73. https://doi.org/10.1016/j.compag.2015.11.015.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

Saberioon, M., Gholizadeh, A., Cisar, P., Pautsina, A., Urban, J., 2017. Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues. Rev. Aquac. 9, 369–387. https://doi.org/10.1111/raq.12143.

Shafait, F., Harvey, E.S., Shortis, M.R., Mian, A., Ravanbakhsh, M., Seager, J.W., Culverhouse, P.F., Cline, D.E., Edgington, D.R., 2017. Towards automating underwater measurement of fish length: a comparison of semi-automatic and manual stereo–video measurements. ICES J. Mar. Sci. 74, 1690–1701. https://doi.org/10.1093/icesjms/fsx007.

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv.

Turra, E.M., Fernandes, A.F.A., de Alvarenga, E.R., Teixeira, E.A., Alves, G.F.O., Manduca, L.G., Murphy, T.W., Silva, M.A., 2018. Longitudinal analyses of correlated response efficiencies of fillet traits in Nile tilapia. Animal 12, 445–453. https://doi.org/10.1017/S1751731117001768.

Wang, G., Hwang, J.-N., Wallace, F., Rose, C., 2019. Multi-scale fish segmentation refinement and missing shape recovery. IEEE Access 7, 52836–52845. https://doi.org/10.1109/ACCESS.2019.2912612.

Zhou, S.K., Greenspan, H., Shen, D., 2017. Deep Learning for Medical Image Analysis. Elsevier.