



## Automating fish age estimation combining otolith images and deep learning: The role of multitask learning

Dimitris V. Politikos <sup>a,\*</sup>, Georgios Petasis <sup>b</sup>, Archontia Chatzispyrou <sup>a</sup>, Chryssi Mytilineou <sup>a</sup>, Aikaterini Anastasopoulou <sup>a</sup>

<sup>a</sup> Institute of Marine Biological Resources and Inland Waters, Hellenic Centre for Marine Research, 16452, Argyroupoli, Greece

<sup>b</sup> Institute of Informatics and Telecommunications, National Centre for Scientific Research "Demokritos", 60228, Agia Paraskevi, Greece

### ARTICLE INFO

Handled by: B. Morales-Nin

**Keywords:**

Otolith images  
Deep learning  
Fish age estimation  
Multitask learning  
Classification

### ABSTRACT

Knowledge on the age of fish is vital for assessing the status of fish stocks and proposing management actions to ensure their sustainability. Prevalent methods of fish ageing are based on the readings of otolith images by experts, a process that is often time-consuming and costly. This suggests the need for automatic and cost-effective approaches. Herein, we investigate the feasibility of using deep learning to provide an automatic estimation of fish age from otolith images through a convolutional neural network designed for image analysis. On top of this network, we propose an enhanced - with multitask learning - network to better estimate fish age by introducing as an auxiliary training task the prediction of fish length from otolith images. The proposed approach is applied on a collection of 5027 otolith images of red mullet (*Mullus barbatus*), considering fish age estimation as a multi-class classification task with six age groups (Age-0, Age-1, Age-2, Age-3, Age-4, Age-5+). Results showed that the network without multitask learning predicted fish age correctly by 64.4 %, attaining high performance for younger age groups (Age-0 and Age-1, F1 score > 0.8) and moderate performance for older age groups (Age-2 to Age-5+, F1 score: 0.50–0.54). The network with multitask learning increased correctness in age prediction reaching 69.2 % and proved efficient to leverage its predictive performance for older age groups (Age-2 to Age-5+, F1 score: 0.57–0.64). Our findings suggest that deep learning has the potential to support the automation of fish age reading, though further research is required to build an operational tool useful in routine fish aging protocols for age reading experts.

### 1. Introduction

Fish age structure is a key tool in the study of fish populations, providing useful knowledge on their growth rate, mortality, age at maturity and recruitment (Campana, 2001). Such knowledge has proven also vital for assessing their current status (e.g. level of exploitation) and proposing management actions to ensure their sustainability (Bianchini and Ragonese, 2011; Campana, 2001; Carbonara and Follesa, 2019).

Otoliths are widely used as a tool of teleost fish age determination, due to their rhythmic growth properties that reflect age at daily and annual levels. Otoliths are incremental structures, which are related to fish growth, and thus constitute a well-established recorder of fish age (Panfili et al., 2002; Wang et al., 2019; Williams and Bedford, 1974). However, determining fish age is not an easy task. Prevalent methods of fish ageing are based on readings of otolith images by experts, a process that is often time-consuming, costly and prone to biases due to reader

subjectivity (Carbonara and Follesa, 2019), suggesting the need for automatic and cost-effective approaches.

Deep learning has been widely used for automating various domains, such as image classification, image segmentation, object detection, face and speech recognition and text processing (Davies, 2017; LeCun et al., 2015). Convolutional Neural Networks (CNNs) are a class of deep learning neural networks designed for image analysis (Krizhevsky et al., 2012; Goodfellow et al., 2015). In Fisheries science, CNNs have been utilized to automate fish species detection and classification (Allken et al., 2019; Cui et al., 2020; Deep and Dash, 2019; Salman et al., 2016; Villon et al., 2018), fish counting from surveillance videos (French et al., 2015) and fish size estimation (Álvarez-Ellacuría et al., 2020; Monkman et al., 2019).

Machine learning and deep learning approaches have already been proposed to automate fish ageing. Robertson and Morison (2001) combined feed forward neural networks with signal (brightness values

\* Corresponding author at: Hellenic Centre for Marine Research, Greece.

E-mail address: [dimpolit@hcmr.gr](mailto:dimpolit@hcmr.gr) (D.V. Politikos).

along transects within sectioned otolith images) and biological data (fish length, sex, date of capture, otolith weight, date of capture) to automate the ageing procedure of nine fish species in Australian waters. Their work demonstrated that biological data can have a positive effect on age prediction. Fablet and Le Josse (2005) combined otolith images as well as biological and shape features of plaice (*Pleuronectes platessa*) with statistical learning, achieving a predictive accuracy in age estimation of 88 %, compared to expert ground truth. Moen et al. (2018) used the “Inception v3” CNN network (Szegedy et al., 2016) to predict the age of Greenland halibut (*Reinhardtius hippoglossoides*) from otolith images. Age was correctly estimated in 29 % of the tested cases, with an additional 38 % being within one year of difference between prediction and expert readings. The implementation of a pre-trained CNN to estimate the age of snapper (*Pagrus auratus*) and hoki (*Macruronus novaezealandiae*) in New Zealand waters using otolith images demonstrated an accuracy slightly below 50 % (Moore et al., 2019). These studies indicate that automatic fish age estimation is feasible, however further research is needed to improve the predictability of models involved. The observed differences in accuracy among the studies can be attributed to the adopted methodology, type of data, shape features of otoliths and lifespan of fish species, illustrating the difficulty to develop a general automatic method for fish ageing.

Typically, a CNN network receives images as input and is trained to accomplish a single task (e.g. to predict gender from a face image) using an optimization process (training), which minimizes model error through a loss function (typically a distance between model predictions and ground truth). Then, the trained network can be used to make predictions on a new, unseen during training, set of images. Multitask learning (MTL) is a subfield of machine learning, where instead of focusing on a single task, multiple auxiliary tasks are used simultaneously. The motivation behind MTL is to share knowledge among the auxiliary tasks through “hard parameter sharing” (Caruana, 1997; Ruder, 2017), causing the network to prefer hypotheses that explain more than one task. This generally reduces the risk of overfitting<sup>1</sup> because the auxiliary tasks act as regularizers (Chen et al., 2019) and improves generalization during training (Ruder, 2017). MTL can be applied on a wide range of algorithms, including CNNs, in which a CNN is trained on multiple related tasks (e.g. to predict gender and age from a face image), with the aim to leverage the generalization performance and prediction accuracy of all tasks compared with learning them individually (Caruana, 1997; Ruder, 2017; Zhang and Yang, 2018). To train a CNN with multitask learning, a joint loss function is defined as the weighted sum of individual loss functions from each task (Ruder, 2017).

In this study, we explore the feasibility of using deep learning to automate fish age estimation from otolith images using a CNN network designed for image classification. On top of this network, a second objective is to investigate for the first time an enhanced with multitask learning CNN network to better estimate fish age by introducing as an auxiliary task the prediction of fish length from otolith images. The two resulting CNN networks (one that employs MTL and one that is not) are trained, tested and compared on a collection of otolith images of red mullet (*Mullus barbatus*) caught in Greek waters. Overall, we aim to expand the applicability of deep learning in fisheries community and identify potential improvements needed to build an operational tool useful in routine fish aging protocols by age reading experts.

## 2. Material and methods

### 2.1. Data acquisition and preprocessing

For the purposes of this work, red mullet *M. barbatus* Linnaeus, 1758 was selected as a case study, since this species is of high commercial importance in Mediterranean fisheries (Tserpes et al., 2019) and every year thousands of otoliths of this species are aged in the frame of the national data collection fisheries programs. A set of 5027 otolith images of the red mullet *M. barbatus* was provided by the Hellenic Centre for Marine Research (HCMR) otolith database along with the age readings and fish length (in mm) of each individual. Specimens were collected within the framework of the Greek National Data Collection Program from the south Aegean and the Ionian Seas over the period 2016–2018. Both otoliths were extracted from the cranial cavity of each individual, cleaned in running water, dried and stored in tubes, and were photographed using the Image Analysis Pro-Plus software. Digital images were taken with different magnifications depending on the otolith size, so that the reader could detect more precisely the annual rings. This process resulted in images with variable resolution ranging between 600–800 pixels in width and 400–600 pixels in height.

Age determination is based mainly on counting the annuli formation (opaque and dark rings surrounding the centre of the otolith) on the otolith surface (Campana, 2001). In general, higher growth rate is observed during the first two years of fish life, whereas as fish get older, growth rate and the distance of the annuli decreases, making the estimation of age more difficult. Otolith images of different ages are presented in Fig. 1, showing the annuli location. Age interpretation in red mullet is closely related to the number of annuli, the date of sampling and the date of birth of each fish (ICES, 2012, 2017; Mahé et al., 2012, 2016). In this study, for simplification reasons, the number of annuli identified on the otoliths of each fish was used as the age of this fish. The age of each otolith was estimated by two expert readers separately. Carbonara et al. (2015) reported that age estimation of red mullet may vary significantly due to reader experience, geographical area, environmental conditions, and differences in the reading methods and interpretation of growth pattern. For this reason, to minimize bias in age estimation, we only used those otoliths for which age readings of both readers were in agreement. Moreover, the dataset used was exclusively from one country (Greek waters) to avoid geographical variability and the effect of environmental factors in growth.

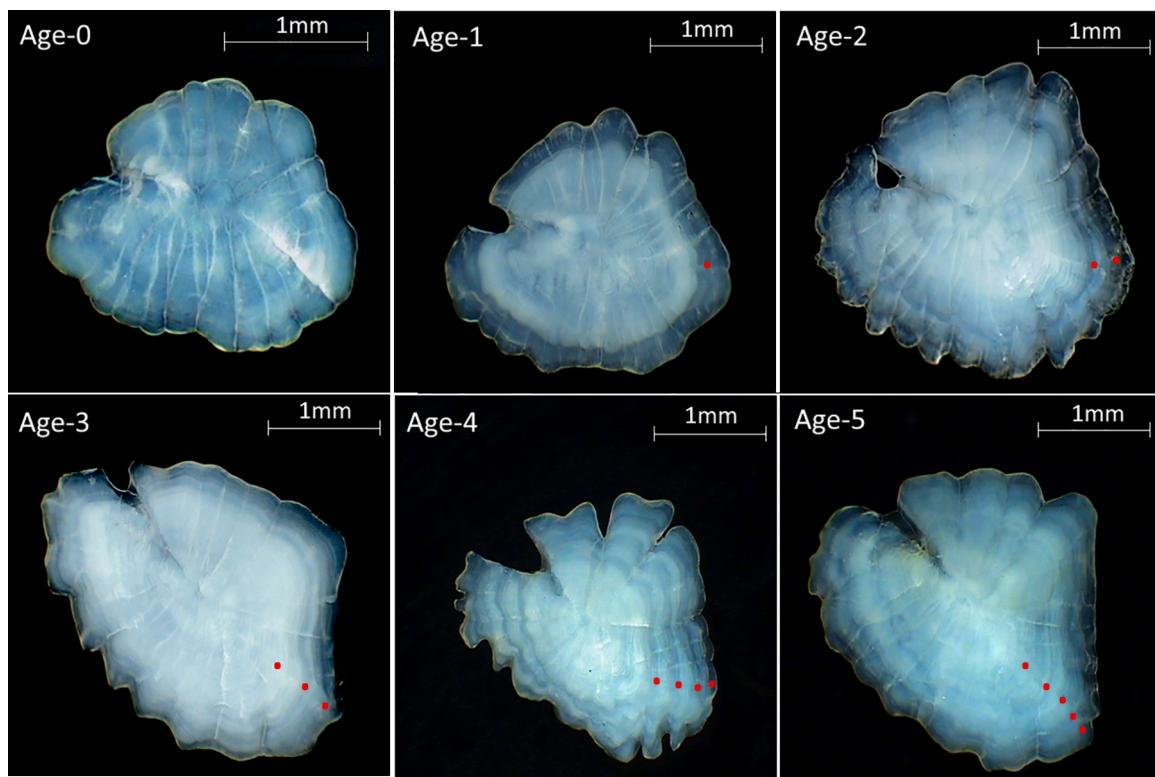
Red mullet age in the dataset was estimated and recorded by readers as discrete values, ranging between 0 and 11 years. Due to the low number of specimens with age >5 years old (~6% of the dataset), they were merged into the 5+ age group and their age was predicted as 5 years old. In total, we considered six age groups: Age-0, Age-1, Age-2, Age-3, Age-4 and Age-5+. Dataset distribution showed higher frequencies for Age-1 and Age-2 age groups and a gradual decrease in frequencies from Age-3 to Age-5+ (Fig. 2). The length of the red mullet distributed in the age groups showed that older fish tend to reach greater lengths (Fig. 3). Despite the high correlation between fish length and age (~0.9 in our dataset), fish length is not usually used as a recorder of fish age. This can be explained by i) the differences in the individual growth rate among fish of the same species and ii) the extended spawning period that can occur as in the case of red mullet (Carbonara et al., 2015; Chatzispyrou et al., 2016), both resulting in important overlap in size among fish of different age groups (Fig. 3).

### 2.2. Convolutional neural network

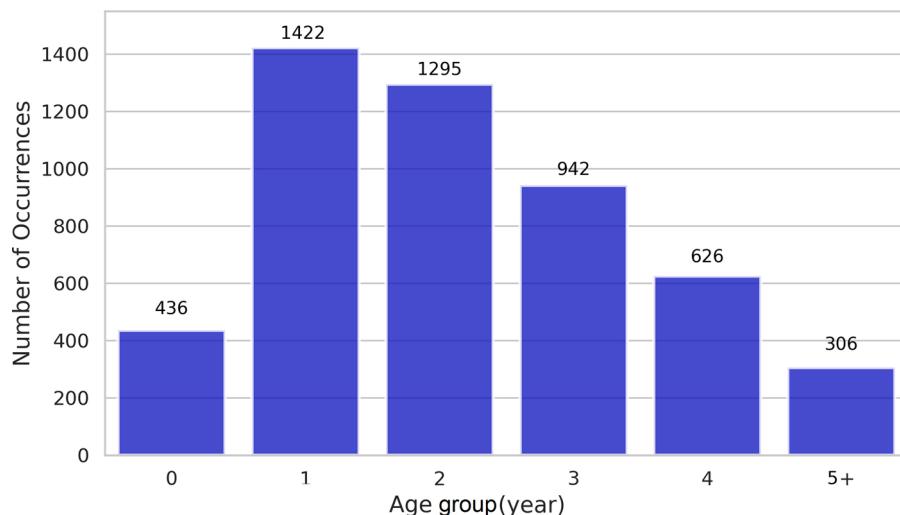
#### 2.2.1. Architecture

The CNN architecture used in this study is an extension of Inception v3 developed by Szegedy et al. (2016). Inception v3 is a state-of-the-art architecture for image classification being pretrained on ImageNet, a large dataset of over ~1.2 million training images and 1000 object categories (Russakovsky et al., 2015). It is built as a sequence of

<sup>1</sup> Overfitting occurs when a deep learning neural network achieves a good performance on the training data and a poor performance on new, unseen data. In other words, an overfitted network suffers from low prediction accuracy.



**Fig. 1.** Digital images of the red mullet (*Mullus barbatus*) otoliths from various ages. Each red dot indicates one year of age. Images were taken with different magnifications depending on the otolith size, so that the reader could detect more precisely the annual rings.



**Fig. 2.** Age distribution of the examined red mullet (*Mullus barbatus*) otoliths.

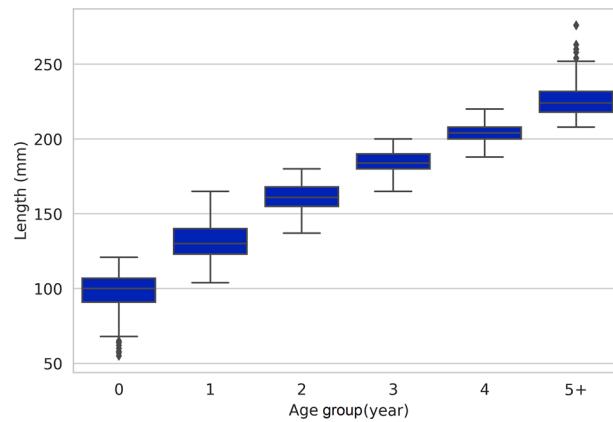
convolutional and pooling layers and can classify images with a size of  $299 \times 299$  pixels into one of 1000 categories. More details about Inception v3 architecture can be found in the Appendix A.

To use Inception v3 in our dataset, specific modifications were made. Initially, the otolith images (input layer for the CNN) were resized to a common size of  $400 \times 400$  pixels using PIL python package (Clark, 2015) and standardized from 0 to 255 pixel values to values between 0.0 and 1.0 to fit into the CNN. The increase of the input resolution to  $400 \times 400$  pixels (default size in Inception v3:  $299 \times 299$ ) was chosen to balance good quality of images during network training and limitations in computational resources. The fully-connected image classification network of Inception-v3 was removed and replaced by a feed-forward

network with two hidden layers and the default output layer of 1000 categories was replaced by a fully-connected layer with six hidden nodes, corresponding to the six age groups Age-0, Age-1, Age-2, Age-3, Age-4 and Age-5+, considering fish age prediction as a multiclass classification task.<sup>2</sup> For the purposes of this study, henceforth this network is named as “Fish Age Network (FAN)”. A schematic view of FAN architecture is shown in Fig. 4.

Fish age estimation takes discrete values in the range [0–5],

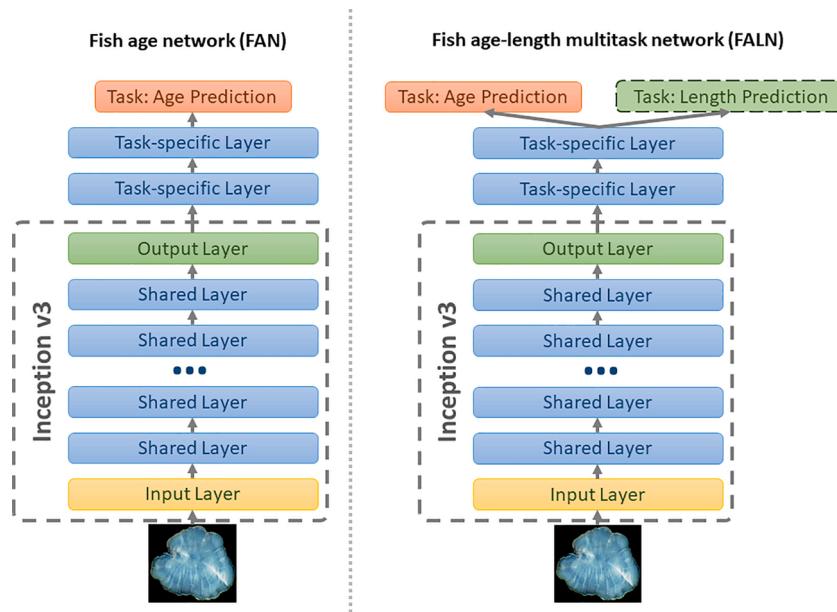
<sup>2</sup> The classes returned by the model are [0, 1, 2, 3, 4, 5], which are directly mapped to age groups [Age-0 – Age-5+].



**Fig. 3.** Total length of the red mullet (*Mullus barbatus*) distributed in the age groups.

where  $m$  is the size of the dataset and  $\hat{y}_c^{(i)}$  is the ground truth for the correct prediction of class  $c$ .

A typical deep learning neural network can be seen as a sequence of layers, with each layer approximating a function  $f(x)$ , forming a representation of its input  $x$ . Each intermediate layer in a deep network is a transformation, learned from data during training. Due to the sequential nature of neural networks, the output of each layer is the input of the next layer, and as a result the input features of a layer are the learned transformation of the previous unit, suggesting that useful features must be learned in the various layers during training. As a result, Inception v3 learns different representations from an image through its sequential layers, which act as feature identifiers with increasing complexity. The first layers can learn representations that concern basic features, such as edges, colors, corners and curves, while layers in the middle of the network and last layers tend to learn more composite features, i.e. recognize “object parts” or whole “objects” of an image. These representations act synergistically and eventually produce a single vector that



**Fig. 4.** Schematic view of fish age network (FAN) and fish age-length multitask neural network (FALN).

corresponding to the age groups [Age-0 – Age-5+]). This motivated us to follow a classification approach, which also returns a discrete value in the range [0–5], when the neural network makes the prediction from an otolith image. To do so, the network outputs a list of probabilities using the softmax<sup>3</sup> function, one for each age value within [0, 5]. Then, the age value with the highest probability is selected as the age prediction. Finally, ages are mapped to age groups [Age-0 - Age-5+]. The metrics used for evaluating the performance of classification tasks (subsection 2.2.3) allow us to quantify the network efficiency for each class separately and subsequently identify limitations in predictions.

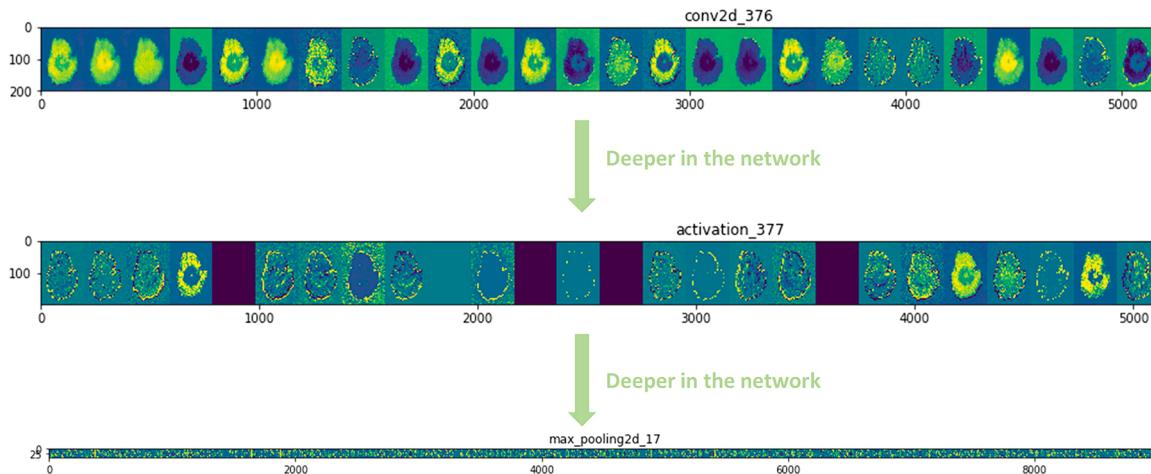
For training FAN, we used the categorical cross-entropy loss function (typical function for multi-class classification tasks) (Goodfellow et al., 2015), defined as,

$$\text{Loss}(\text{age}) = \frac{1}{m} \sum_{i=1}^m \log(\hat{y}_c^{(i)}),$$

summarizes the most meaningful features of the image. For illustration purposes, Fig. 4 shows how fish age network visualizes different viewpoints of an otolith image through the learned representations, for three layers, one close to the input layer, one from the middle of the network, and one towards the output of the network. On the top of Fig. 5, the transformations learned by various kernels can be seen, with each kernel transforming the input image differently, concentrating on recognising specific aspects of the input image. Similarly, an activation layer deeper in the network focuses on elements like edges or specific areas of the input image.

On top of FAN, we employed multitask learning using as an auxiliary task the prediction of fish length from an otolith image. The new network, called henceforth “Fish Age-Length multitask Network (FALN)”, receives as input the otolith images and predicts simultaneously fish age and length. The fish length prediction is a regression task (continuous values in prediction), derived from a fully connected layer with a single hidden node (a rectified linear unit, ReLU). It should be noted that the auxiliary task of fish length prediction is used only during the training process of the network. When the network is applied on otolith images to predict fish ages, the output of the auxiliary task is ignored. We emphasize that the input in FALN is solely otolith images and we don't use fish length to predict age. A schematic view of FALN

<sup>3</sup> A function which performs multi-class classification and transforms the input values to the range [0, 1] with a sum of 1, representing the probability distributions of a list of potential outcomes.



**Fig. 5.** Features of an otolith image identified through the Inception V3 network layers.

architecture is shown in Fig. 4.

The composite loss function for training the FALN network is described by the equation:

$$\text{Total Loss} = a_1 * \text{Loss}(\text{age}) + a_2 * \text{Loss}(\text{length})$$

where  $\text{Loss}(\text{age})$  is the loss function for age prediction and  $\text{Loss}(\text{length})$  is the loss function for length prediction. The loss weights  $a_1$  and  $a_2$  are hyperparameters that need to be tuned during network training.

The Mean Squared Error (MSE) loss function was used for fish length prediction (typical loss function for regression tasks):

$$\text{Loss}(\text{length}) = \frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2,$$

where  $\hat{y}_i$  is the measured length and  $y_i$  is the predicted length over the  $m$  samples of the dataset.

### 2.2.2. Training

For the purpose of training FAN and FALN, the convolutional layers of Inception v3 were set to trainable, i.e. weights in the hidden layers were allowed to be updated. The dataset was randomly split into three subsets with a 70 %–15 %–15 % ratio: the training set, the validation set and the testing set, which consisted of 3517, 755 and 754 otolith images, respectively. The training and validation sets were used to train the fish networks over sequential epochs, whereas the testing set was used to evaluate their performance. More technical information of the way network training is conducted can be found in the Appendix A.

Data augmentation is often applied in small datasets to generate more images through various transformations (flip, rotation, zoom, brightness). This process improves the training process by decreasing the risk of overfitting and enhancing the generalizability of the model (Shorten and Khoshgoftaar, 2019). In our dataset, we applied random rotation of otolith images between 0–360° to artificially increase the size of training and validation subsets. Preliminary results showed that other type of transformations didn't have a positive effect and thus were excluded from our analysis. This is in agreement with Shorten and Khoshgoftaar (2019) who pointed out that proper choice of data augmentation is necessary to attain optimal performance of a network.

The configuration of the training process is determined by a set of hyperparameters (Table 1). A search was conducted for different batch sizes (8, 12, 16, 20, 28, 32), learning rates (0.0001, 0.0004, 0.00001, 0.00004, 0.001, 0.01, 0.1), optimizers (Adam, SGD, RMSprop), dropouts (0.2, 0.3, 0.35, 0.4), number of hidden layers (1,2,3), number of nodes (128, 256, 512), L1 regularization (0.5, 0.1, 0.05, 0.025), patience (10, 20, 25, 50) and loss weights ( $a_1 = 1$ ,  $a_2$  was set to 1, 100, 450, 475, 500, 750, 1000, 2000). The optimal values of the hyperparameters were

**Table 1**

Configuration adopted to train the fish age and fish age-length multitask neural networks.

Attributes of networks	Description	Fish age network	Fish age-length multitask network
Adam optimizer	Default values with categorical cross entropy loss function	Default	Default
Epoch	Number of times that the training dataset is passed forward and backward through the network to refine model weights.	120	120
Learning rate	A parameter used to train a network via gradient descent. The gradient descent algorithm multiplies the learning rate by the gradient.	0.0004	0.0004
Batch size	The number of images used in one iteration of network training.	16	16
Patience	Number of epochs that the model is not improved during training. It's used for early stopping of neural network training.	25	25
Loss weights	Individual loss contributions of age and length to the final loss in the fish age-length multitask network.	–	1, 475
Fully connected layer (FC)	- Two hidden layers and number of nodes - L1 regularization = add penalty to loss function - Dropout = drop out neuron connections between hidden layers	1024, 512 0.025 0.35	1024, 512 0.025 0.35

found using a learning rate of 0.0004, a batch size of 16 and the Adam optimization algorithm. Additionally, the patience was set to 25, the loss weights were set  $a_1 = 1$  and  $a_2 = 475$ , hidden layers were set to 2, number of nodes was set to 1024 for the first layer and 512 for the second layers, dropout was set to 0.35 and L1 regularization was set to 0.025.

The strategy followed for training the fish networks was to train initially FAN and find the optimal hyperparameters. Then, we used the same hyperparameters for the FALN and we further optimized the parameters  $a_1$  and  $a_2$ . By this way, we were able to identify that any further improvement in FALN prediction occurred due to multitask learning. For optimizing  $a_1$  and  $a_2$ , we fixed the value of  $a_1 = 1$  and then we performed a Grid search provided by scikit (Pedregosa et al., 2011) to

estimate  $a_2$ .

### 2.2.3. Performance

For evaluating the performance of fish networks on the testing set, we used the standard evaluation metrics of accuracy, precision, recall and F1-score (Powers, 2011):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

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

where TP and TN are true positives and true negatives, and FP and FN are false positives and false negatives. These metrics are computed separately for each of the age groups, as well as averages over all age groups. Accuracy is the fraction of true predictions out of all predictions. Precision is the fraction of correctly-predicted ages out of all predicted cases of this age-group. Recall is the fraction of correctly predicted ages out of the actual cases of this age-group. The F1 score is a composite measure (harmonic mean) of precision and recall and tends to decline when one of the two metrics is low, providing a better indication of the overall performance of the model. To summarize the above metrics over all age-groups, we use their macro and weighted-averages. Macro-average calculates metrics for each class individually and then takes unweighted mean of the measures. In contrast, weighted-average takes a weighted mean of the metrics based on the number of data instances in the testing set. The evaluation metrics are calculated and listed in a table using scikit-learn Python module (Pedregosa et al., 2011). The Mean Squared Error (MSE) is used to estimate the error in fish length predictions.

The design, training and evaluation of fish networks were conducted with Keras 2.3.1 and Tensorflow 2.2.0 backend (Abadi et al., 2016). The code and samples of the otolith images can be found at GitHub<sup>4</sup>.

### 2.2.4. Alternative approach

Prior relevant works (Moen et al., 2018; Moore et al., 2019) have formulated fish age estimation as a regression task, where age is predicted as a continuous variable. In this case, the MSE metric is commonly used to evaluate the performance of the networks over all age groups. For comparison reasons with these works, we conducted two additional runs with FAN and FALN architectures, considering fish age as a regression problem. This resulted in predicting age on a continuous scale. Therefore, the loss function for age was set to be the MSE, i.e.

$$\text{Loss}(\text{age}) = \frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2,$$

where  $\hat{y}_i$  is the measured age and  $y_i$  is the predicted age over the  $m$  images of the dataset.

## 3. Results

The performance of FAN and FALN networks during training with respect to loss and accuracy is presented in Fig. 6. Concerning the FAN, training age-loss gradually declined over the 120 epochs to a minimum of  $\approx 0.22$  (Fig. 6A, blue line), whereas validation age-loss followed a declining trend till epoch 50 to a minimum of  $\approx 1$  and then slightly started to fluctuate between 1 and 2, signifying overfitting (Fig. 6A, red

line). Training accuracy reached a maximum of  $\approx 0.95$  (Fig. 6B, blue line), whereas validation accuracy topped at  $\approx 0.67$  over the 120 epochs (Fig. 6B, red line).

In the case of FALN, training age-loss was continually declining over the 120 epochs to a minimum of  $\approx 0.42$  (Fig. 6C, blue line). Validation age-loss aligned with training age-loss till epoch  $\approx 85$ , decreasing to a minimum of  $\approx 0.75$  and thereafter started slightly to increase, fluctuating around 1 (Fig. 6C, red line). Training accuracy gradually increased and reached  $\approx 0.83$ , whereas validation accuracy reached a maximum of  $\approx 0.7$  at around 80 epochs (Fig. 6D).

The predictive performance of FAN is shown in Table 2. Precision, Recall and F1 score were high for Age-0 (0.90) and Age-1 (0.82) (Table 2). Interestingly, F1 score was very high for Age-0 despite the fact that the instances in the testing set were relatively few (support = 64), compared to other age groups (Table 2). In contrast, F1 score performed less well for older ages: Age-2 (0.52), Age-3 (0.54), Age-4 (0.50) and Age-5+ (0.52). The overall performance of FAN, according to macro and weighted F1 scores was 0.63 and 0.64, respectively.

The predictive performance of FALN is shown in Table 3. Similar to FAN, Precision, Recall and F1 were high for the younger age groups, i.e. Age-0 (0.83) and Age-1 (0.85). Additionally, F1 score was less effective on predicting older ages, i.e. Age-2 (0.64), Age-3 (0.57), Age-4 (0.60) and Age-5+ (0.59). The overall performance of F1 score, based on macro and weighted averages, was 0.68 and 0.70, respectively.

The comparison of F1 scores per age group between FAN and FALN networks is illustrated in Fig. 7. F1 score was notably higher in FALN compared to FAN for Age-2 (increase by 23.00 %) and Age-4 (increase by 20.13 %) groups, whereas F1 score in FALN was slightly lower for Age-0 compared to FAN (decrease by 7.77 %). Over all age groups, macro-average F1 score was 0.63 for FAN (Table 2) and increased to 0.68 for FALN (Table 3). Similarly, weighted-average F1 score increased from 0.64 for FAN (Table 2) to 0.70 for FALN (Table 3).

On the testing set of 754 otolith images, FAN predicted correctly the age of 486 out of 754 images (64.4 %), while in 241 out of 754 images (31.9 %) showed one year of difference compared to the expert estimated age (135 less and 106 plus one year than the ground truth) (Fig. 8, blue bars). For the FALN, age was correctly estimated in 522 out of 754 images (69.2 %) with 220 out of 754 images (29.1 %) being with one year of difference (70 less and 150 plus one year than the ground truth) (Fig. 8, orange bar).

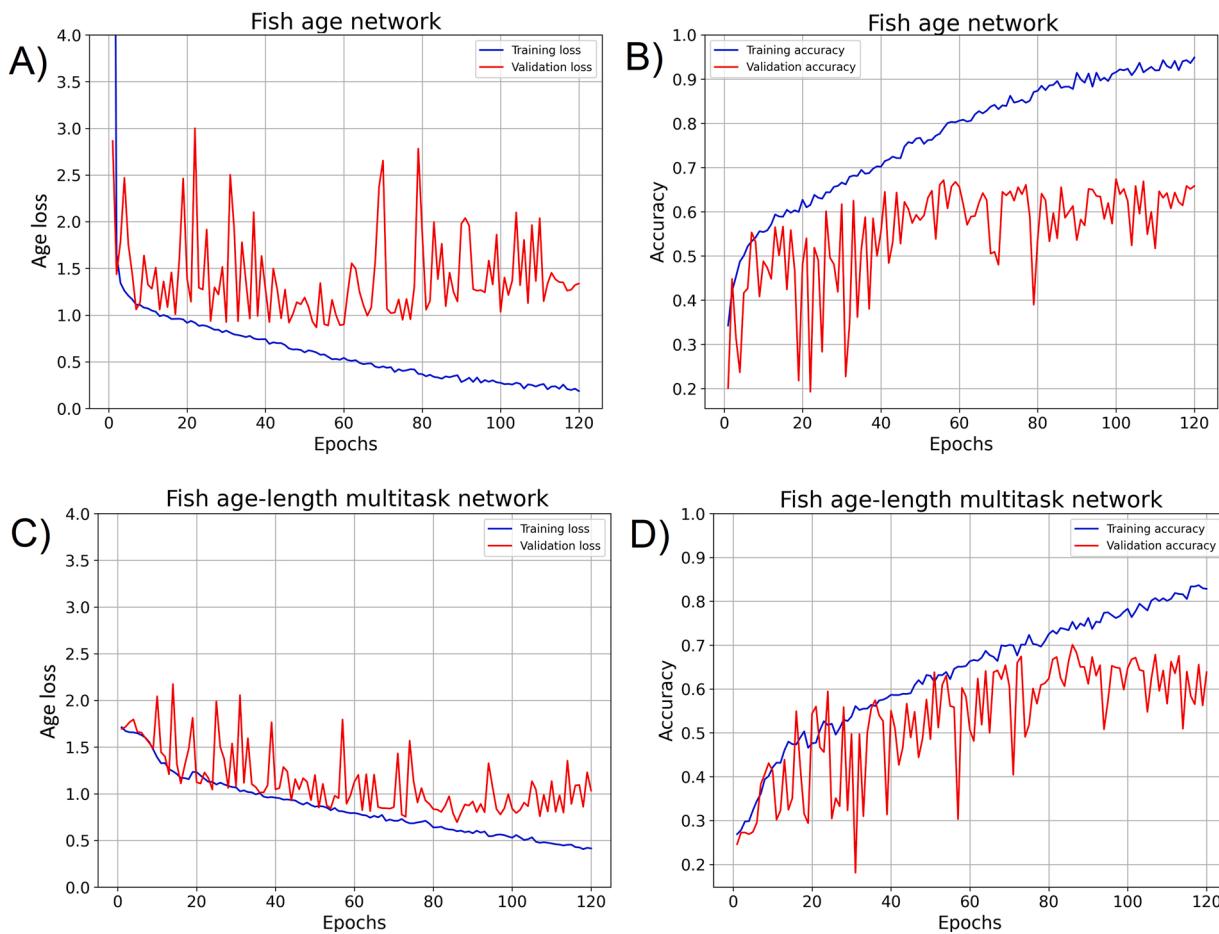
When treating fish age estimation as a regression problem, the MSE was 0.58 years for FAN and 0.56 years for FALN.

## 4. Discussion

The innovation of this research was to introduce a deep learning methodology to accomplish fish age prediction from otolith images and showcase the benefit of multitask learning in leveraging the predictive performance of proposed fish CNN networks. As a case study, we used a dataset of otolith images of the red mullet species collected in Greek waters. To accomplish this experimental trial, we first used a pre-trained convolutional neural network as a baseline and configured it for our task through the FAN. FAN demonstrated a correctness in age prediction by 64.4 % performing well on the younger age groups and moderately on older ones. The inclusion of multitask learning in our analysis resulted to FALN, which performed better compared to FAN. More specifically, FALN improved correct age prediction to 69.2 % (increase by 4.7 % compared to FAN) and proved more efficient to identify older age groups increasing F1 score between 3.6 % and 23.0 %, compared to FAN.

The high predictive performance of networks for the young age groups aligns with the low variability in red mullet age estimation reported in workshops (ICES, 2012, 2017). Accordingly, the low accuracy in older ages found in our analysis agrees with the findings of these workshops, showing a high coefficient of variation between readers for the older ages, reflecting the difficulty in distinguishing the annuli on the otoliths in older individuals.

<sup>4</sup> <https://github.com/dimpolitik/DeepOtolith>



**Fig. 6.** Plots displaying the age loss and accuracy of training and validation sets over 120 epochs for fish age network and fish age-length multitask networks. Fish age network: A) training (blue line) and validation (red line) age-loss, B) training (blue line) and validation accuracy (red line); Fish age-length multitask network: C) training (blue line) and validation (red line) age-loss, D) training (blue line) and validation accuracy (red line).

**Table 2**

Classification report of the fish age neural network (FAN) for the testing set (754 instances). Accuracy, Precision, Recall, and F1 score metrics were computed per age group as well averaged over all age groups through macro and weighted-averages. The Support column indicates the number of data instances per age group at the testing set.

Age group	Precision	Recall	F1 Score	Support
Age-0	0.93	0.88	0.90	64
Age-1	0.76	0.88	0.82	216
Age-2	0.68	0.43	0.52	181
Age-3	0.50	0.59	0.54	152
Age-4	0.44	0.57	0.50	98
Age-5+	0.74	0.40	0.52	43
accuracy			0.64	754
macro average	0.68	0.62	0.63	754
weighted average	0.66	0.44	0.64	754

The extended decline of validation age loss in FLAN till epoch 85 when validation age loss of FAN declined till epoch 50 (Fig. 6A,C; red lines), the slowdown of FLAN training accuracy to 0.84 till epoch 120 when FAN topped at 0.95 (Fig. 6B,D; blue lines), and the better performance of FALN on validation set compared to FAN (Fig. 6B,D; red lines) illustrated the regularization effect of multitask learning to prevent overfitting and finally improve fish age prediction on the test set.

The current results are encouraging, though the fish networks need to be further improved before being adopted in routine fish aging protocols. The small dataset size of otolith images for older age groups as well the grouping of ages 5–11 into the Age 5+ class did not allow an

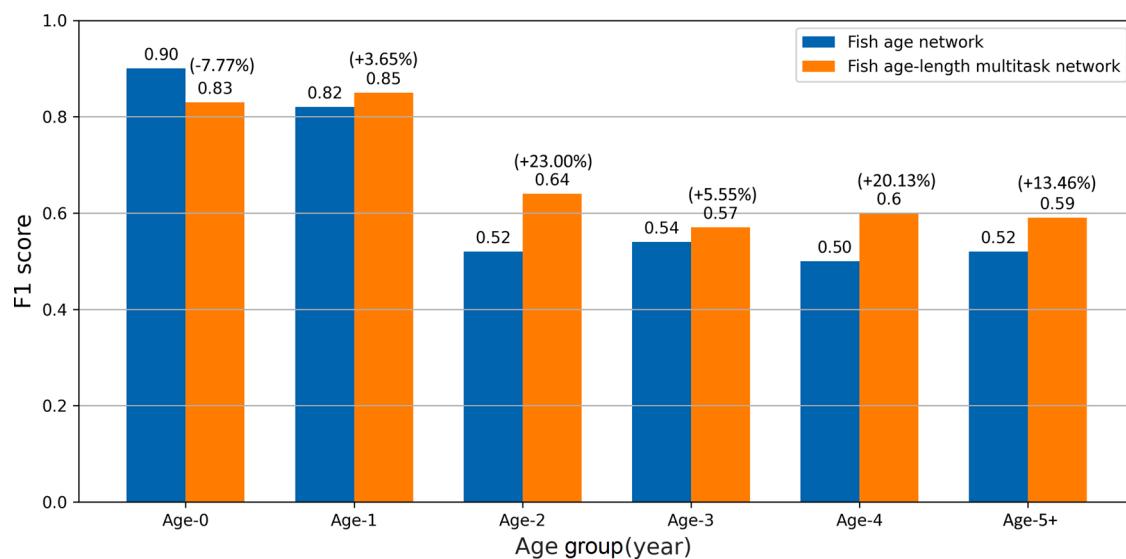
**Table 3**

Classification report for the fish-age-length multitask neural network (FALN) for the testing set (754 instances). Accuracy, Precision, Recall, and F1 score metrics were computed per age group and averaged over all age groups through macro and weighted-averages. The Support column indicates the number of data instances per age group at the testing set.

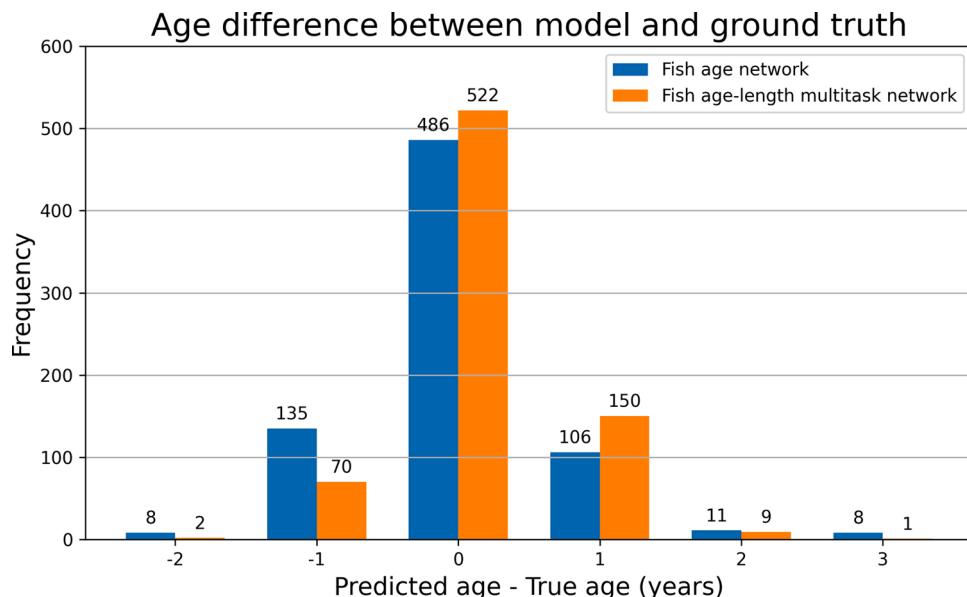
Age group	Precision	Recall	F1 Score	Support
Age-0	0.91	0.77	0.83	64
Age-1	0.85	0.84	0.85	216
Age-2	0.70	0.60	0.64	181
Age-3	0.58	0.57	0.57	152
Age-4	0.50	0.76	0.60	98
Age-5+	0.66	0.53	0.59	43
accuracy			0.69	754
macro average	0.70	0.68	0.68	754
weighted average	0.71	0.69	0.70	754

efficient training of the networks, which was in turn reflected in moderate predictions. Moreover, the real difficulty to distinguish the annuli in older ages may also be reflected in network results, restricting further their accuracy. Future work should include more images of older fish in order to facilitate the training process. The positive effect of multitask learning on fish age estimation was encouraging, suggesting that exploring the possibility of introducing additional fish characteristics (e.g. length, weight, gender) or morphometric characters related to the otolith structure (e.g. surface area, perimeter diameter) into the networks may be valuable. These are worthy issues for future consideration.

The MSEs of the present work from the regression approach and



**Fig. 7.** F1 scores per age group for fish age and fish age-length multitask networks. Percentages in parentheses show the increase/decrease of macro-average F1 of fish age-length multitask network (FALN) compared to the fish age network (FAN).



**Fig. 8.** Histogram of the age difference: Predicted age – True age (years) calculated in the testing set (in total 754 instances) for fish age network (FAN) and fish age-length multitask neural network (FALN).

those from similar studies are presented in Table 4. Our study adopted 6

**Table 4**

Mean Square Error (MSE) for fish age predictions from present work and related studies. N-age groups is the total number of age groups of the tested fish species. MSE values from our study were computed when treating fish age estimation as a regression problem.

References	Species	N-age groups	MSE
Present study	Red mullet	6	0.58 <sup>1</sup> 0.56 <sup>2</sup>
Moen et al. (2018)	Greenland halibut	26	2.6
Moore et al. (2019)	Snapper	27	1.2
	Hoki	8	1.3

<sup>1</sup> Fish age network.

<sup>2</sup> Fish age-length multitask network.

age groups and MSE was 0.58 for FAN and 0.56 for FALN. The study of Moen et al. (2018) included 26 age groups and MSE was 2.6. Moore et al. (2019) findings showed a relatively low MSE = 1.2 for snapper with 27 age groups and a relatively high MSE = 1.3 for Hoki with 8 age groups. The variability observed in MSEs among the studies may be partially explained by the different range of age groups assumed in the networks, also illustrating the challenge to build a general CNN for fish age estimation over multiple species.

When employing MTL, individual tasks may be measured through different loss functions and hence individual losses may be in different scales. Losses that are significantly smaller than other losses may be ignored during training (as their impact in the total loss may be negligible) while larger losses can have a greater impact during training, which may result in the training process ignoring some of the auxiliary tasks, without exploiting the potential advantages of MTL. Typically, we want all tasks to equally contribute to the total loss, thus in cases of

auxiliary task losses imbalance, the imbalance has to be corrected through weights that scale all losses to the same value range (loss weights). Thus, the proper tuning of loss weights assigned in individual losses during network training is a key factor for the success of MTL. In our case, the optimal loss weights were found through hyperparameter tuning to be  $a_1 = 1$  for age prediction and  $a_2 = 475$  for length prediction. The main role of these parameters was to balance the different scales of age-loss and length-loss. The ratio age-loss/length-loss during training (Fig. 9) showed that age-loss is 100–500 times higher than length-loss. By selecting  $a_2 = 475$ , the losses of fish age and length prediction tasks ended up being roughly on the same scale during epochs 100–120, where the mean ratio of the two losses is around 460, concluding that the balance among individual task losses is more beneficial to occur during the final stages of model training (Fig. 9). It is also worth noting that the loss weights do not have necessarily an ecological interpretation.

During data pre-processing, we merged the otolith images, which corresponded to fish age older than 5 years old, into the Age-5+ age class due to their very small dataset size (< 20 images). This was inevitable because network training with very few images is practically pointless, leading to almost zero prediction accuracy. Thus, the collection of more images for these age groups is necessary to improve fish age estimations.

We also note that our approach did not employ fish length data as an additional signal for predicting fish age, as already has been proposed for other biological data (Robertson and Morison, 2001; Fablet and Le Josse, 2005). Instead, fish length data were used as a regularizer through MTL during the training phase only, leading to a better network in terms of fish age prediction, which is done using only otolith images.

The fish network architectures proposed in this study can become the starting point for further development. Our approach may be equally tested to other fish species. Testing other state-of-the-art pretrained models such as EfficientNet (Tan and Le, 2019) or performing hyperparameter tuning with an optimisation algorithm (Bergstra and Bengio, 2012; Wu et al., 2019) could be also of interest. Sharing CNN networks, exclusively pre-trained on fish otolith images (Moen et al., 2018; Moore et al., 2019), would be a quick way to explore how fish networks respond to domain-specific transfer learning. Moreover, setting a long-term goal for collecting images over thousands of fish images in a common platform could result in a dataset similar to ImageNet, contributing tremendously to this kind of research as previously highlighted (Moen et al., 2018).

Fish age estimation from otolith images is a key component for the management of fish stocks, providing valuable information to estimate

vital rates of fish biology and population dynamics (Wang et al., 2019). High costs in data collection and age interpretation have forced fisheries programmes to limit the number of fish specimens analysed and rely on model estimations to determine the age composition of fish stocks (Carbonara and Follesa, 2019). This urges the need to come up with automated methods to estimate fish age from large datasets of otolith images. Our approach suggests that deep learning can contribute towards this direction and constitute a valuable tool for age reading experts in the near future.

## Funding

No funding was received for this work.

## Intellectual property

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

## Research ethics

We further confirm that any aspect of the work covered in this manuscript that has involved human patients has been conducted with the ethical approval of all relevant bodies and that such approvals are acknowledged within the manuscript.

Written consent to publish potentially identifying information, such as details or the case and photographs, was obtained from the patient(s) or their legal guardian(s).

## Authorship

All listed authors meet the ICMJE criteria. We attest that all authors contributed significantly to the creation of this manuscript, each having fulfilled criteria as established by the ICMJE.

We confirm that the manuscript has been read and approved by all named authors.

We confirm that the order of authors listed in the manuscript has been approved by all named authors.

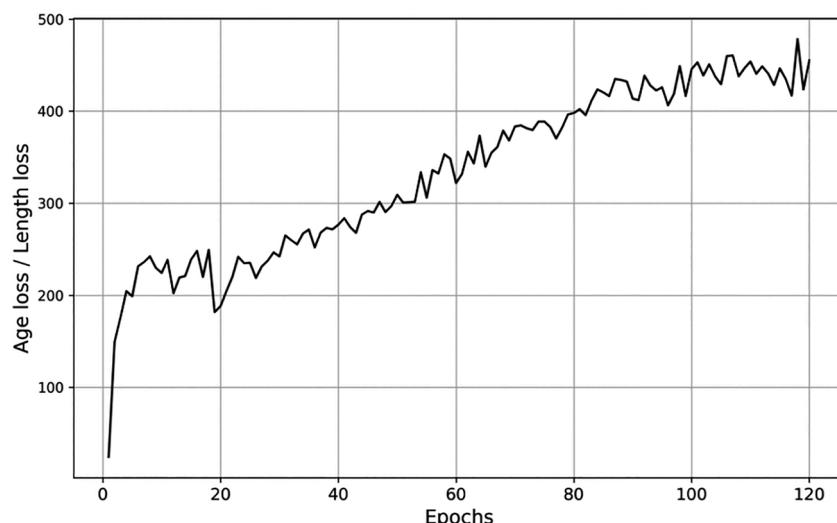


Fig. 9. Plot of age-loss/length-loss ratio over 120 epochs during the training of fish age-length multitask network (FALN).

## Contact with the editorial office

This author submitted this manuscript using his/her account in EVISE.

We understand that this Corresponding Author is the sole contact for the Editorial process (including EVISE and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

We confirm that the email address shown below is accessible by the Corresponding Author, is the address to which Corresponding Author's EVISE account is linked, and has been configured to accept email from the editorial office of American Journal of Ophthalmology Case Reports: [dimpol@hcmr.gr](mailto:dimpol@hcmr.gr)

## CRediT authorship contribution statement

**Dimitris V. Politikos:** Conceptualization, Methodology, Software, Formal analysis, Writing - original draft. **Georgios Petasis:** Methodology, Software, Writing - review & editing. **Archontia Chatzispyrou:** Data curation, Conceptualization, Writing - review & editing. **Chryssi Mytilineou:** Data curation, Conceptualization, Writing - review & editing. **Aikaterini Anastasopoulou:** Data curation, Conceptualization, Writing - review & editing.

## Conflict of Interest

No conflict of interest exists.

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

## Acknowledgements

The authors would like to thank Ketil Malde and Endre Moen for their useful advice during the configuration of the neural network architectures, the IMBRIW of HCMR database and the DCF Hellenic Programme for providing the data. We express our great appreciation to the Editor and the two reviewers for their insightful comments and their effort to follow the technical details of our work.

## Appendix A

### Inception v3 architecture

Inception v3 architecture is built on a series of layers, which consist of i) convolutional and pooling filters with rectified linear units as activation function, ii) three fully-connected layers to the final concatenation layer and iii) a softmax classifier activation for predicting the correct category. Convolutional layers may be understood as an efficient way to transform an input image into another image, highlighting meaningful patterns. Fully-connected layers take a vector as input and produce another vector as output after passing it through hidden layers, which are made up from neurons. These neurons are fully-connected to all neurons in the previous layer. For a general introduction to CNNs, see Goodfellow et al. (2015).

### Network Training

During network training, the performance of fish networks is monitored in terms of training and validation loss, and training and validation accuracy over sequential epochs. Epochs represent the number of times that the training dataset is passed forward and backward through the network to refine model weights. Accuracy is the proportion of age predictions that match the ground truth. The goal of training is to decrease loss and increase accuracy through the tuning of

hyperparameters. These measurements are monitored per epoch, both for training and validation sets. Whenever the validation loss decreases the trained model is saved, ending up with the network that corresponds to the minimum loss and highest accuracy on the validation set. The trained network is then evaluated on the testing set. For a general introduction how to train a deep learning model, see Goodfellow et al. (2015).

## References

- Abadi, M., et al., 2016. TensorFlow: Large-scale Machine Learning on Heterogeneous Distributed Systems. available from. Software. <https://www.tensorflow.org/>.
- Allken, V., Handegard, N.O., Rosen, S., Schreyeck, T., Mahiot, T., Malde, K., 2019. Fish species identification using a convolutional neural network trained on synthetic data. - ICES. J. Mar. Sci. 76 (1), 342–349. <https://doi.org/10.1093/icesjms/fsy147>.
- Álvarez-Ellacuría, A., Palmer, M., Catalán, I.A., Lisaní, J.-L., 2020. Image-based, unsupervised estimation of fish size from commercial landings using deep learning. ICES J. Mar. Sci. 77 (4), 1330–1339. <https://doi.org/10.1093/icesjms/fsz216>.
- Bergstra, J., Bengio, J.Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13 (1), 281–305. <http://jmlr.org/papers/v13/bergstra12a.html>.
- Bianchini, M.L., Ragonese, S., 2011. Establishing length-at-age references in the red mullet, *Mullus barbatus* L. 1758 (*Pisces, Mullidae*), a case study for growth assessments in the Mediterranean Geographical Sub-Areas (GSA). Mediterr. Mar. Sci. 12 (2), 316–332. <https://doi.org/10.12681/mms.35>.
- Campana, S.E., 2001. Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. J. Fish Biol. 59, 197–242. <https://doi.org/10.1006/jfib.2001.1668>.
- Carbonara, P., Intini, S., Modugno, E., Maradonna, F., Spedicato, M.T., Lembo, G., Zupa, W., Carnevali, O., 2015. Reproductive biology characteristics of red mullet (*Mullus barbatus* L., 1758) in Southern Adriatic Sea and management implications. Aquat. Living Resour. 28, 21–31. <https://doi.org/10.1051/alr/2015005>.
- Carbonara, P., Follesa, M.C. (Eds.), 2019. Handbook on Fish Age Determination: a Mediterranean Experience. Studies and Reviews. No. 98. FAO, Rome, p. 192.
- Caruana, R., 1997. Multitask learning. Mach. Learn. 28 (1), 41–75. <https://doi.org/10.1023/A:1007379606734>.
- Chatzispyrou, A., Anastasopoulou, A., Mytilineou, C., Kavadas, S., Papantoniou, V., Panagiotou, A., 2016. Preliminary study of the reproductive cycle of the red mullet (*mullus barbatus*) in the aegean Sea. In: 16th Hellenic Conference of Ichthyologists. Kavala, Greece.
- Chen, Q., Peng, Y., Keenan, T., Dharssi, S., Agro, E., Wong, W., Chew, E., Lu, Z., 2019. A multi-task deep learning model for the classification of age-related macular degeneration. AMIA Summits Transl. Sci. Proc. 1–10 arXiv:1812.00422. <https://arxiv.org/abs/1812.00422>.
- Clark, A., 2015. Pillow (PIL Fork) Documentation. Readthedocs. Retrieved from. <https://buildmedia.readthedocs.org/media/pdf/pillow/latest/pillow.pdf>.
- Cui, S., Zhou, Y., Wang, Y., Zhai, L., 2020. Fish detection using deep learning. Appl. Comput. Intell. Soft Comput. 1–13. <https://doi.org/10.1155/2020/3738108>.
- Davies, E.R., 2017. Computer Vision. Principles, Algorithms, Applications, Learning, 5th edition. Academic Press, p. 900.
- Deep, B.V., Dash, R., 2019. Underwater fish species recognition using deep learning techniques. In: 6<sup>th</sup> International Conference on Signal Processing and Integrated Networks (SPIN). Noida, India, pp. 665–669.
- Fablet, R., Le Josse, N., 2005. Automated fish age estimation from otolith images using statistical learning. Fish. Res. 72 (2-3), 279–290. <https://doi.org/10.1016/j.fishres.2004.10.008>.
- French, G., Fisher, M.H., Mackiewicz, M., Needle, C.L., 2015. Convolutional neural networks for counting fish in fisheries surveillance video. Proceedings of the Machine Vision of Animals and Their Behaviour (MVAB) 1–7.
- Goodfellow, I.J., Bengio, Y., Courville, A., 2015. Deep Learning. MIT Press, p. 433.
- ICES, 2012. Report of the Workshop on Age Reading of Red Mullet and Striped Red Mullet, 2–6 July 2012, Boulogne-sur-Mer. ICES CM 2012/ACOM:60, France, p. 52.
- ICES, 2017. Workshop on Ageing Validation Methodology of *Mullus* species (WKVALMU), 15–19 May 2017, Conversano. ICES CM 2017/SSGIEOM:31, Italy, p. 74.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. <https://doi.org/10.1038/nature14539>.
- Mahé, K., Elleboode, R., Charilaou, C., Ligas, A., Carbonara, P., Simona, I., 2012. Striped Red Mullet (*Mullus surmuletus*) and Red Mullet (*Mullus barbatus*) Otolith and Scale Exchange, 2011, p. 30.
- Mahé, K., Anastasopoulou, A., Bekas, P., Carbonara, P., Casciaro, L., Charilaou, C., Elleboode, R., Gonzalez, N., Guijarro, B., Indennata, A., Kousteni, V., Massaro, A., Mytilineou, Ch., Ordines, F., Palmisano, M., Panfilii, M., Pesci, P., 2016. Report of the Striped Red Mullet (*Mullus surmuletus*) and Red Mullet (*Mullus barbatus*), p. 21.
- Moen, E., Handegard, N.O., Allken, V., Albert, O.T., Harbitz, A., Malde, K., 2018. Automatic interpretation of otoliths using deep learning. PLoS One 13 (12), e0204713. <https://doi.org/10.1371/journal.pone.0204713>.
- Monkman, G.G., Hyder, K., Kaiser, M.J., Vidal, F.P., 2019. Using machine vision to estimate fish length from images using regional convolutional neural networks. Methods Ecol. Evol. 10, 2045–2056. <https://doi.org/10.1111/2041-210x.13282>.
- Moore, B.R., McLaren, J., Peat, C., Anjomrouz, M., Horn, P.L., Hoyle, S., 2019. Feasibility of Automating Otolith Ageing Using CT Scanning and Machine Learning. New Zealand Fisheries Assessment Report 2019/58, p. 23.

- Panfili, J., Pontual, H.D., Troadec, H., Wright, P.J. (Eds.), 2002. Manual of Fish Sclerochronology. Ifremer-IRD coedition, Brest, France.
- Pedregosa, F., et al., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Powers, D.M.W., 2011. Evaluation: from precision, recall and F-Measure to ROC, informedness, markedness & Correlation. *J. Mach. Learn. Technol.* 2 (1), 37–63. <https://10.9735/2229-3981>.
- Robertson, S., Morison, A., 2001. Development of an artificial neural network for automated age estimation. Department of Natural Resources and Environment. ISBN 0-7311-5038-4.
- Ruder, S., 2017. An overview of multi-task learning in deep neural networks. arXiv 1706 (05098). <https://arxiv.org/abs/1706.05098>.
- Russakovsky, O., et al., 2015. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252. <https://doi.org/10.1007/s11263-015-0816-y>.
- Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., Harvey, E., 2016. Fish species classification in unconstrained underwater environments based on deep learning. *Limnol. Oceanogr.-Meth.* 14 (9), 570–585. <https://doi.org/10.1002/lom3.10113>.
- Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *J. Big Data* 6 (1), 48. <https://doi.org/10.1186/s40537-019-0197-0>, 60.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). arXiv, 1512.00567. <https://arxiv.org/abs/1512.00567>.
- Tan, M., Le, Q., 2019. EfficientNet: rethinking model scaling for convolutional neural networks. In: Proceedings of 36th International Conference of Machine Learning 6105–6114.
- Tserpes, G., Massuti, E., Fiorentino, F., Facchini, M.T., Viva, C., Jadaud, A., Joksimovic, A., Pesci, P., Piccinetti, C., Sion, L., Thasitis, I., Vrgoc, N., 2019. Distribution and spatio-temporal biomass trends of red mullets across the Mediterranean. *Sci. Mar.* 43–55. <https://doi.org/10.3989/scimar.04888.21A>.
- Villon, S., et al., 2018. A deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecol. Inform.* 48, 238–244. <https://doi.org/10.1016/j.ecoinf.2018.09.007>.
- Wang, C.-H., Benjamin, D., Walther, B.D., Gillanders, B.M., 2019. Introduction to the 6th international otolith symposium. *Mar. Freshw. Res.* 70, i–iii. [https://doi.org/10.1071/MFv70n12\\_ED](https://doi.org/10.1071/MFv70n12_ED).
- Williams, T., Bedford, B.C., 1974. The use of otoliths for age determination. In: Bagemal, T.B. (Ed.), *The Ageing of Fish. Proceedings of the International Symposium*. Allen & Unwin, London, pp. 114–123.
- Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, 575H., Deng, S.-H., 2019. Hyperparameter optimization for machine learning models based on Bayesian optimization. *JETSC* 17 (1), 26–40. <https://doi.org/10.11989/JEST.1674-862X.80904120>.
- Zhang, Y., Yang, Q., 2018. An overview of multi-task learning. *Nat. Sci. Rev.* 5, 0–43. <https://doi.org/10.1093/nsr/nwx105>.