Ecological Informatics 63 (2021) 101322

Contents lists available at ScienceDirect

Ecological Informatics

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

Automatic interpretation of salmon scales using deep learning

Rune Vabø [a,1,*], Endre Moen [a,1], Szymon Smoliński [a,b], Åse Husebø [a], Nils Olav Handegard [a], Ketil Malde [a,c]

[a] Institute of Marine Research, Bergen, Norway
[b] Department of Fisheries Resources, National Marine Fisheries Research Institute, Gdynia, Poland
[c] Department of Informatics, University of Bergen, Norway

ARTICLE INFO

Keywords:
Fish scales
Deep learning
EfficientNet
Transfer learning
Age reading
Maturity staging

ABSTRACT

For several fish species, age and other important biological information is manually inferred from visual scrutinization of scales, and reliable automatic methods are not widely available. Here, we apply Convolutional Neural Networks (CNN) with transfer learning on a novel dataset of 9056 images of Atlantic salmon scales for four different prediction tasks. We predicted fish origin (wild/farmed), spawning history (previous spawner/non-spawner), river age, and sea age. We obtained high prediction accuracy for fish origin (96.70%), spawning history (96.40%), and sea age (86.99%), but lower accuracy for river age (63.20%). Against six human expert readers with an additional dataset of 150 scales, the CNN showed the second-highest percentage agreement for sea age (94.00%, range 87.25±97.30%), but the lowest agreement for river age (66.00%, range 66.00– 84.68%). Estimates of river age by expert readers exhibited higher variance and lower levels of agreement compared to sea age and may indicate why this task is also more difficult for the CNN. Automatic interpretation of scales may provide a cost- and time-efficient method of predicting fish age and life-history traits.

1. Introduction

Aquatic science is based on collecting and analysing large volumes of data. Management of large marine ecosystems depend increasingly on efficient analysis of these data, and it has been argued that the heavy reliance on traditional manual data processing methods may be a major bottleneck in ecosystem assessment frameworks (Malde et al., 2020). Parallel to the increase in data volume and complexity, it is paramount that efficient data processing and automated analysis methods are developed.

The age structure of a fish population provides important information relating to population dynamics, which is essential for fisheries and conservation management (Niemelä et al., 2006b; Ricker, 1975). For many fish species, hard structures such as their scales or otoliths, can be analysed to infer age. For example, the complex life history of Atlantic salmon (*Salmo salar* L. 1758) can be inferred from analysing fish scale patterns. Salmon scales have been used for over a century to estimate age and growth (Dahl, 1911), but the extraction of biological information still requires manual expertise, since there are no reliable automated techniques presently in use.

Atlantic salmon occur in the temperate and subarctic regions of the North Atlantic ecosystem (Aas et al., 2011) and provide a range of ecosystem services and are considered an economically valuable species (Butler et al., 2009). Complex life history and high divergence of their habitats make the management and conservation of this species difficult (Crozier et al., 2004). Most salmon populations are anadromous, adapted to living in both fresh and seawater, with a juvenile phase in rivers along various Atlantic coastlines lasting one to six years (Otero et al., 2012a, 2012b) before migrating into the ocean for feeding (Erkinaro et al., 2019; Hansen and Quinn, 1998). After between one and eight years in the sea, Atlantic salmon migrate back to their juvenile freshwater habitat to spawn (Niemelä et al., 2006a). Many Atlantic salmon populations have been negatively affected by several factors (e.g. freshwater habitat degradation, pollution, diseases and over-exploitation), influencing both the river and marine phase of their life cycles (Hansen and Quinn, 1998). Some of these changes are related to the salmon farming industry, which has grown dramatically on a global scale since the late 1970s (Ford and Myers, 2008; McGinnity et al., 2003). The ecological and economical importance of this species makes efficient methods facilitating the monitoring of wild salmon stocks even

* Corresponding author.
E-mail address: rune.vaboe@hi.no (R. Vabø).
[1] Equal contribution.

https://doi.org/10.1016/j.ecoinf.2021.101322
Received 22 January 2021; Received in revised form 9 May 2021; Accepted 9 May 2021
Available online 15 May 2021

more important.

1.1. Scales as indicators

The complex life history of Atlantic salmon is reflected in how their scales grow. By careful examination of their scales, researchers can determine important characteristics such as river and sea age (i.e. the years spent in the respective phases), spawning history, and whether or not they originated in wild or in farmed fish stocks (Francis, 1990). As scales grow, concentric rings are formed on the surface of each scale, and the growth rate of these rings is proportional to the somatic growth of the fish (Fisher and Pearcy, 1990; Panfili et al., 2002). The scale has two main growth zones, the freshwater zone and the marine zone (Fig. 1). In the temperate regions, seasonal changes in the somatic growth cause differences in the patterns on the scale: narrow winter circuli appear as darker bands formed when water temperature and food supply are low, and wider summer circuli appear as brighter bands, characteristic of fast growth, when water temperatures and food availability are high (Shearer, 1989, but see Thomas et al., 2019). Paired darker and brighter bands form annual 'growth marks', which can be used for reliable and accurate fish ageing (Ibáñez et al., 2008, Spurgeon et al., 2015). Fish age in years is estimated based on the number of winter bands (Fig. 1).

Scales can also provide information relating to the life history of the fish. For instance, previous spawning activity can be inferred from spawning marks identified on scales (Niemelä et al., 2006a). Recognition of certain scale features can also help discriminate between wild and farmed fish (Stokesbury, 1997). For example, circuli spacings tend to be more uniform in farmed Atlantic salmon due to more regular food supply when compared to their wild counterparts (ICES, 2011).

1.2. Automated scale analysis

Considering all the factors contributing to shaping the complex patterns that form on scales, it is easy to appreciate that in-depth analyses of these patterns are nontrivial. Manually extracting biological information from scales requires skilled expertise that takes years to acquire. Automating the extraction of information from scale images could, as a first step, complement current practices. The usefulness of automated fish ageing based on scales or otoliths has long been recognized as an important goal in fishery science (Boehlert, 1985). Different pattern recognition systems have been proposed for the ageing purposes based on growth increments in scales, otoliths, or other biological hard structures. One-dimensional analysis of the intensity profiles extracted along a given reading axis or methods incorporating a two-dimensional perspective have already been developed (Fablet, 2006a; Fisher and Hunter, 2018). Besides image-based information, some methods additionally incorporate independent growth data to constrain the classification algorithms and discriminate true growth increment patterns from false 'checks' (Fablet, 2006b, 2006a; Robertson and Morison, 1999). These techniques all require explicit incorporation of specific knowledge from biological expertise. Historically, AI systems have also primarily relied on explicit inbuilt expert knowledge until the successes of deep learning during the last decade broke this paradigm. The remarkable advantage of deep neural networks are their ability to develop their own expertise by learning to detect and respond to relevant features at many levels of abstraction, obviating the need for explicit feature engineering (LeCun et al., 2015; Schmidhuber, 2015).

In recent years, deep learning techniques, especially deep Convolutional Neural Networks (CNN), have become the dominating computer vision technology and have successfully been applied to a range of complex image classification problems (LeCun et al., 2015). After Krizhevsky et al. in 2012 (Krizhevsky et al., 2012) won the annual *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) competition (Russakovsky et al., 2015), CNNs have become the dominant class of deep neural network architectures for image classification, and subsequent winners have employed increasingly complex network architectures with larger numbers of parameters, e.g. the 2014 ILSVRC winner, GoogleNet (Szegedy et al., 2015), used 6.8 million parameters, whilst the 2018 winner, GPipe, used 557 million parameters (Huang et al., 2018). This implies a correlation between accuracy and network size, limited only by available memory and computational resources. A recently proposed network architecture called EfficientNet aims to scale up networks more efficiently and provide state-of-the-art accuracy within a given computing budget (Tan and Quoc, 2019).

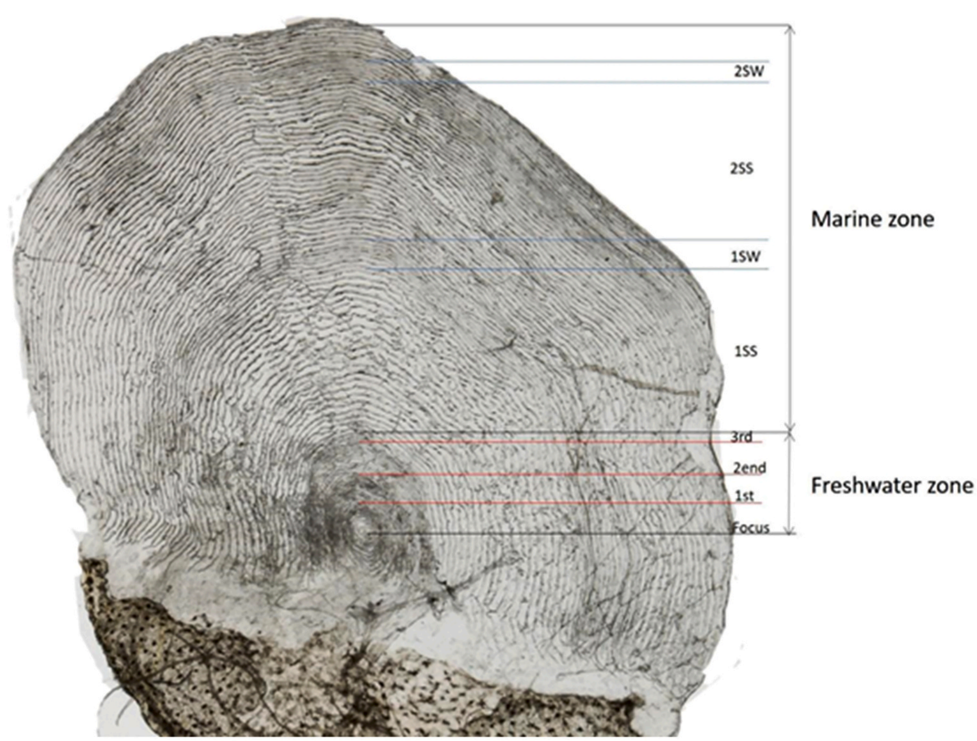


Fig. 1. An example of an expert reader analysis of an adult salmon scale image. Image taken from the 150-image test set. The freshwater zone and the marine zone are shown. In the freshwater zone, the year zones are labelled with red lines. In the marine zone, both winter and summer growth zones are shown. Where 1SS (1. Sea Summer), SW (1. Sea Winter), 2SS (2. Sea Summer) 2SW (2. Sea Winter). CNNs predicted a river age of 2.68 and sea age of 1.998. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In this paper, we utilise recent advances in deep CNNs and use an implementation of EfficientNet to analyse salmon scale images. Our main objective is to explore whether extraction of biological information from scale images can be automated using a CNN. This task differs from typical image classification tasks that most network architectures were developed for, but a previous work successfully used a neural network for age classification of Greenland Halibut otoliths (Moen et al., 2018), indicating that the flexibility of these systems is sufficient. We take advantage of transfer learning by starting our training process with an EfficientNet based CNN pre-trained on an existing open-access image database (ImageNet) and train different CNN models to predict the age, origin (wild/farm escapee) and spawning history of Atlantic salmon based on images of scales. We then evaluate the performance of the CNNs by comparing CNN-based predictions with manual-based estimates made by six expert human readers.

2. Methods

2.1. CNN training data

Norwegian authorities have initiated a large-scale surveillance program comprising eight research institutes and commercial actors to monitor the impact of farmed salmon on wild salmon populations (Anon., 2019). The program samples salmon annually from around 200 Norwegian rivers. The dataset used in this study consist of a total of 9056 high-resolution images of salmon scales sampled by the Institute of Marine Research in Bergen (IMR), Norway (from 2015 to 2018), and Rådgivende Biologer (from 2016 to 2017) in rivers along the coast of Norway. These images were already analysed and labelled by expert readers with biological information, including the origin of the fish (wild or farmed), the spawning history (previous spawner or non-spawner) and the number of years spent in rivers (river age) and at sea (sea age), within those respective years. These readings followed standardised procedures for age reading. In general the IMR scales are always analysed independently by at least two age readers with high agreement. For cases where readers do not agree, an expert reader at Rådgivende Biologer is sometimes consulted to reach agreement, calibrate readings between institutes and maintain consistent age reading. Scales with low quality are discarded. There were three different age readers involved in establishing the 9056 image dataset (one from Rådgivende Biologer and two from IMR). Technical details of this dataset and the reading can be found through the link given in the last section of this paper (Data availability).

However, not all of the biological information was provided for every image. In cases where some information can be interpreted, but not all, this information was included in the dataset. In total, 8286 images were annotated with sea age, 6238 images were annotated with river age, 5919 were annotated as either wild or farmed, and all 9056 images were annotated as either previous spawner or non-spawner. This provided four separate and partly overlapping datasets (Fig. 2). The dataset is dominated by non-spawners (97.4% vs 2.6% previous spawners) and by wild (91.7% vs 8.3% farmed) and contain a large fraction of unknown fish (unknown origin). The sea age dataset is dominated by two years (50.6%), one years (27.9%) and three years (17.4%) while the river age dataset is dominated by three years (56.5%) and two years (33.6%). Images that did not include estimates of river/sea age (not available (NA) in Fig. 2) were used in the binary classifications (origin and the spawning history) but omitted from the regression analysis of fish age.

2.2. Convolutional neural network architecture

In this study, we used both classification and regression CNNs to automate four manual tasks that each extracted biological information from salmon scale images: a classification task distinguishing wild salmon from farmed, a classification task distinguishing previous spawners from non-spawners, and two regression tasks predicting river and sea age. For all tasks, we used the EfficientNet–B4 architecture and each CNN was trained using transfer learning as EfficientNet–B4 was available with pre-trained weights, pre-trained using ImageNet data (Deng et al., 2009). EfficientNet uses a compound scaling method for network depth, width and image input resolution by first finding an optimal baseline combination called EfficientNet–B0 and then scaling up to bigger networks denoted EfficientNet–B1 through EfficientNet–B7 with increasing numbers of parameters. This process determines an optimal image resolution and by default, EfficientNet–B4 uses an input image resolution of 380 × 380 pixels. Despite the possibility benefits of using higher image resolution in order to reveal finer scale details in the salmon scales, we decided to follow the EfficientNet-B4 default architecture and the images were scaled accordingly. For binary classification of farmed/wild and previous spawner/non-spawner salmon, we used an output layer with a two-value softmax output and cross entropy loss. For the regression tasks of sea and river age, we used a linear output unit and

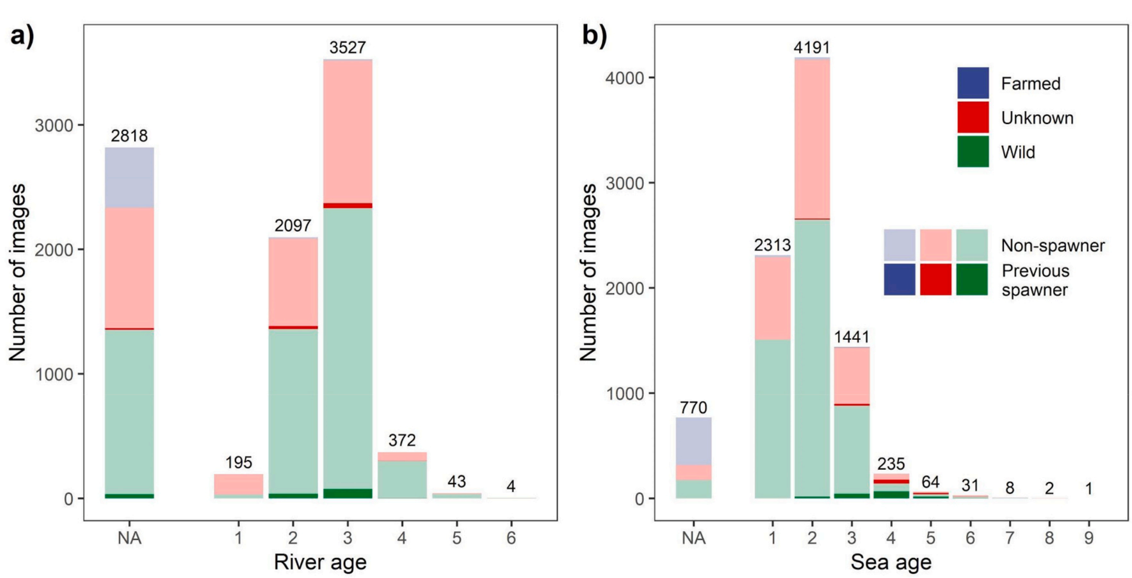


Fig. 2. Age distribution for river (a) and sea (b) age. The colour of the bar indicates the origin of the individual fish and the shade shows spawning history. NA indicates that age information was not available. The total number of samples in each age class is given.

MSE loss. Each of the four distinct CNNs were trained separately, and in each case, scale images were pruned from the dataset when the output class/value was missing.

2.3. Implementation and training

The CNN was implemented using the Keras (Chollet, 2015) and TensorFlow (Abadi et al., 2016) software packages implemented in Python, and computation was performed using CUDA version 9.1 and CuDNN with Nvidia (Nvidia Corp., Santa Clara, California) P100 accelerator cards with 12 GB of GPU memory. We used a Keras implementation of EfficientNet–B4 called EfficientNet V1.1.0 (https://github.com/qubvel/efficientnet). The pre-trained weights used for transfer learning were available through this API.

Augmentation was applied to the entire training dataset. The images were augmented using rotations between 0 and 360 degrees, reflected by the vertical axis, and vertically shifted by +/− 5 pixels. In addition, standard image normalization for CNNs was applied, mapping the 8 bit pixel values to floating-point values between 0 and 1. The dataset was then randomly split into training, validation, and test sets, containing 70%, 15% and 15% of the images, respectively. The validation set was used to control (and terminate) the training process, while final performance metrics were estimated using the test set.

To compensate for unbalanced class abundances, the loss function was weighted, using weights obtained from the compute_class_weight function in the sklearn python package (Pedregosa et al., 2011). For instance, since only 8.5% of the scale images in the origin dataset were labelled as farmed salmon, a weight of 5.85 was assigned to the images of farmed fish and a weight of 0.54 to wild salmon images. Similarly, previous spawner and non-spawner salmon were weighted 19 and 0.5, respectively, as the dataset contains only 2.6% scales from previous spawners. All layers were set to trainable during training.

Performance of the CNN models were assessed using four different metrics (test loss, mean squared error - MSE, mean average percentage error – MAPE and accuracy - Acc). Accuracy of the regression tasks was calculated by rounding the prediction to the nearest integer age and comparing it with the ground truth. The labelling provided by human readers are treated as the ground truth, and accuracy and other performance metrics relates directly to this.

The CNN hyperparameters configurations used for all four networks during training are shown in Table 1. During training we use minibatch gradient descent.

2.4. Comparing the CNN and six expert readers using an independent dataset

To evaluate the performance of the trained CNNs relative to human experts, an additional dataset of 150 salmon scale images were scored simultaneously by six expert readers and by the regression CNNs. The images were selected from 12 different salmon rivers in Norway and were sampled between May and October 2019. Following the standard age reading exchange method for Atlantic salmon (Anon., 2008; ICES, 2013; Shearer, 1989; Shearer, 1992), these 150 images were read independently by each participant to determine river age and sea age. In addition to the scale images, auxiliary background information was utilised during the reading, including catch location and date, body length and sex of the fish. The expert readers were also allowed to use a second magnified image when encountering difficult samples. This dataset and the results of the human readings are available (Husebø et al., 2020). The river and sea age regression CNNs only used the images, with no auxiliary information. We investigated the precision of the age estimates made by all the expert readers and the predictions by the CNNs (rounded to the nearest integer) using the coefficient of variation (CV) and percentage agreement (PA) (Campana, 2001). We evaluated relative bias of the expert-derived age estimates by comparing them with their modal age (calculated from all the expert readers and the CNNs) and visualized the results using age bias plots (Campana et al., 1995). In addition, we calculated mean squared error using non-rounded CNN output.

Table 1
Hyperparameter configuration used. An epoch is one cycle through the training data.

Hyper parameter	Value	Description
Batch size	8	The number of images processed in parallel when training the network
Learning rate	$7{*}10^{-5}$	The step size for parameter updates.
Optimiser	Adam	Algorithm for using the gradient to update parameters.
Steps per epoch	1600	Number of batches run before validation error is calculated
Maximum epochs	150	Maximum epochs before training terminates
Patience	20	Number of epochs without improvement in validation error before training will terminate early

3. Results

3.1. Results of classification and regression tasks

The two binary classification tasks achieved high accuracy, with origin (farmed/wild) reaching an accuracy of 96.7% and the spawning history 96.4%. The sea age CNN performed excellent with MSE of 0.157 and an accuracy of 87%, far better than the river age CNN with MSE of 0.336 and 63.2% accuracy (Table 2, Fig. 3, Fig. 4). Note that during training we use class weighting in the loss calculation for the binary classification CNNs (see 2.4). This prevents the CNNs to “detect” the skewed distributions (non-spawners and farmed dominating), and the different classes are therefore perceived by the CNN as uniformly distributed. The performance of the CNNs for the binary classification tasks are therefore very good compared to a random classifier of 50% accuracy. No class weighting was done for the regression CNNs during training; therefore, their performance should be compared to a random weighted sampler. While a purely random classifier would have an accuracy of 1/(NClasses), a weighted random sampler would randomly guess classes according to their occurring frequency. Calculating this for the regression CNNs gave the following accuracies: river age (43.7%) and sea age (36.5%). Both regression CNNs therefore perform much better than what can be expected by weighted random sampling.

To explore the benefits of transfer learning, we also trained the CNN for predicting sea age using random initialization. This gave an accuracy of 49.6%, a MAPE of 54% and a MSE of 1.1, which is relatively poor performance compared to our results using transfer learning (Table 2, 87%, 8.6% and 0.157 respectively). Training with pre-trained weights was also twice as fast.

Table 2
CNN performance metrics. MSE is the mean square error, MAPE is the mean average percentage error, and Acc% is the average accuracy. In addition, the number of images in the dataset (set size) for each task and the weighting of classes are given. The × indicates that the performance metric is not applicable to the specific CNN.

Predicting	Test loss	MSE	MAPE	Acc %	Set size	Weighting classes
River Age	0.336	0.336	17.34	63.20	6238	–
Sea Age	0.157	0.157	8.64	86.99	8286	–
Spawning	0.113	×	×	96.40	9056	Non-spawner: 0.5, Previous spawner:19
Farmed	0.187	×	×	96.70	5919	Farmed: 5.87, Wild:0.54

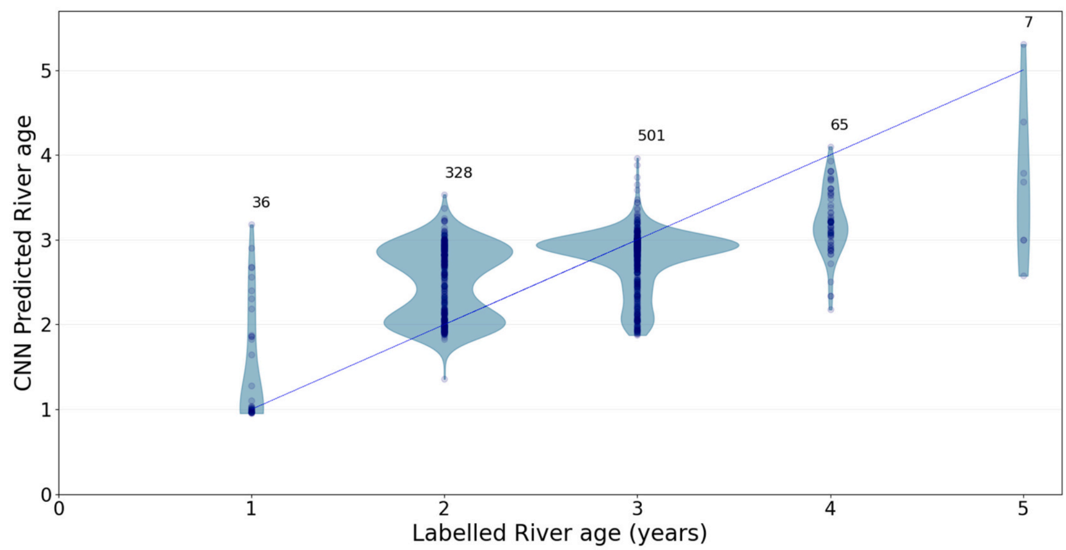


Fig. 3. Predicted versus labelled river age in years. The violin shape takes its form from a smoothed probability density of the values. The straight line indicates correct predictions. The numbers indicated above each distribution is the test data sample size for each age group.

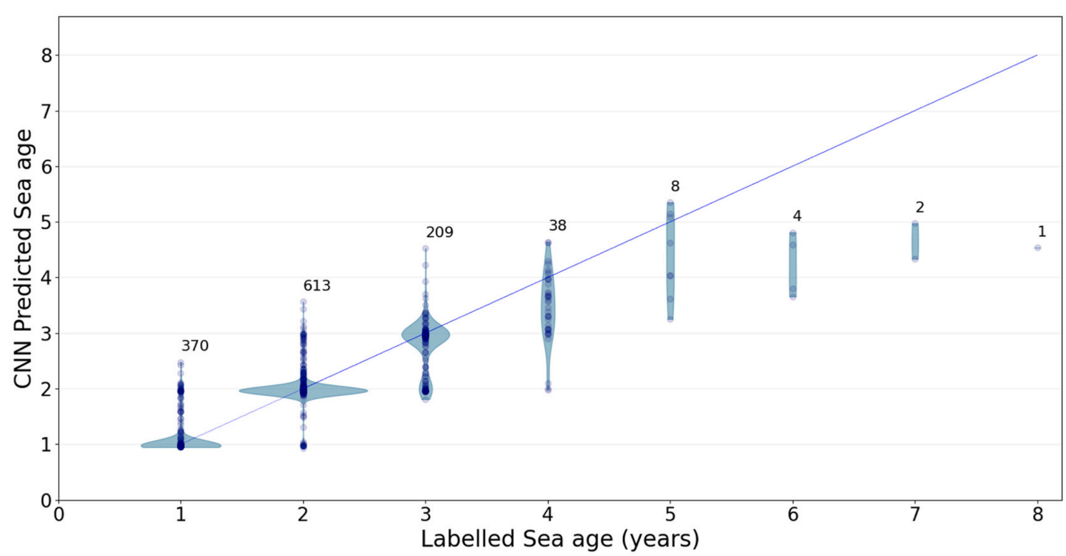


Fig. 4. Predicted versus labelled sea age in years. The violin shape takes its form from a smoothed probability density of the values. The straight line indicates correct predictions. The numbers indicated above each distribution is the test data sample size for each age group.

3.2. Experts estimates and CNN predictions for river and sea age

Estimates of sea and river age made by human readers using the additional 150 image dataset were compared with predictions from the regression CNNs. The sea age CNN performed better than the river age CNN (Table 3, Fig. 5), which is consistent with the main analysis (Table 2 and Figs. 3 and 4). All human expert readers performed better than the CNN with respect to predicting river age (Fig. 5), with agreement among expert readers varying between 73.72% and 84.68%. CNN agreement (rounding predictions to the nearest integer) was lower with a value of 66%. On average, expert reader agreement was 78% for river age and 92% for sea age. The sea age CNN performed better than many

Table 3
Performance metrics for six expert readers and CNN for sea and river age estimation. The mean square error (MSE) is calculated using the total mean (all expert readers and CNN) and the CNN prediction is rounded to the nearest integer when calculating MSE. The coefficient of variation (CV) and the percentage agreement (PA) is calculated using the modal age of all readers and the CNN.

Estimator	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	CNN
MSE River	0.121	0.145	0.130	0.133	0.125	0.159	0.219
CV River	4.19	4.70	6.32	4.78	4.31	7.04	8.97
PA River	84.68	80.31	76.42	80.74	81.20	73.72	66.00
MSE Sea	0.068	0.050	0.033	0.098	0.051	0.059	0.058
CV Sea	2.75	2.64	0.84	3.9	2.01	2.08	1.87
PA Sea	89.86	91.28	97.30	87.25	93.33	93.96	94.00

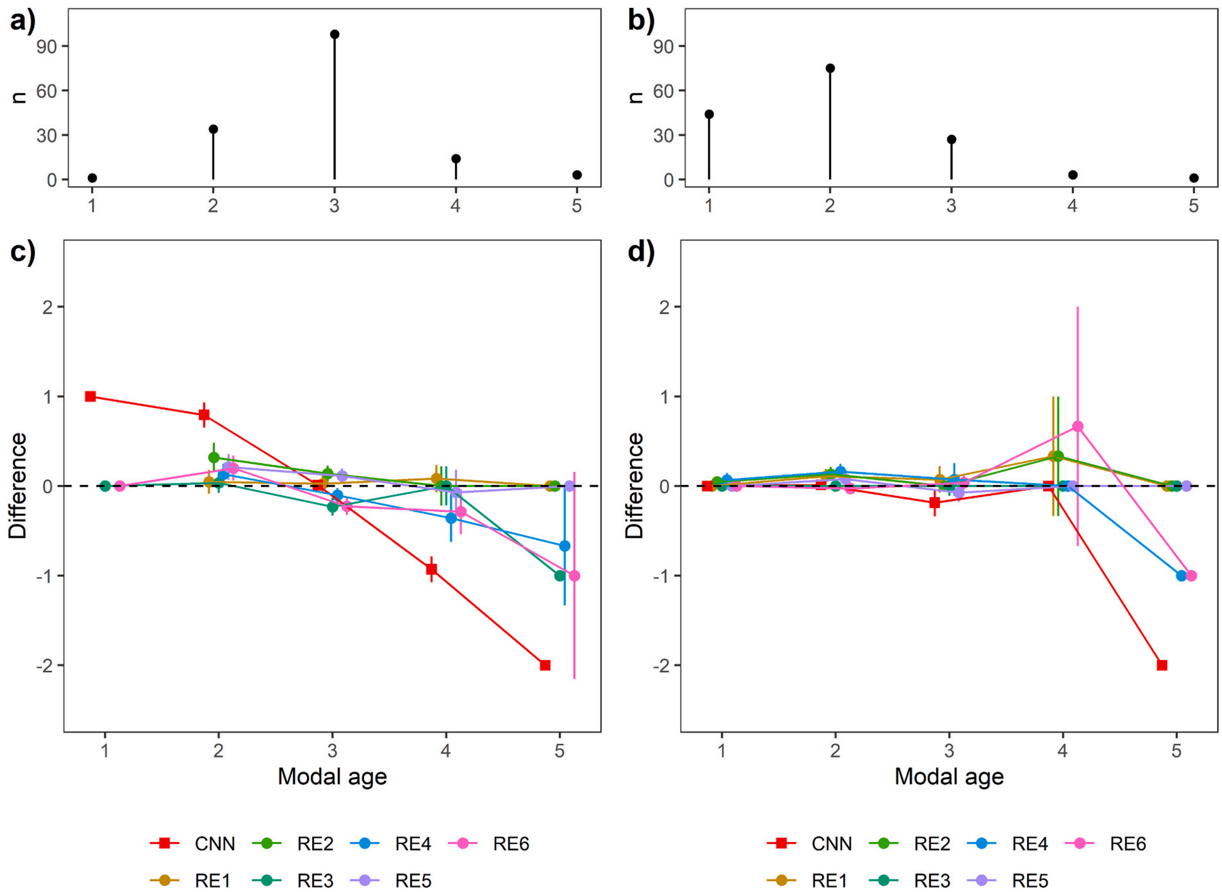


Fig. 5. Age bias plots. Distribution of the modal river age (a), modal sea age (b), and age bias plots for river (c) and sea age (d) reading. Mean difference between expert reader interpretation or CNN rounded prediction and modal age calculated from all readers and CNN is indicated (±2 s.e.) in (c) and (d).

of the expert readers (94% agreement), ranking second overall after expert number 3 (Table 3). The results suggest that when a reading task is difficult for humans (high variance between readers) the CNN performs poorly, whilst when a task is easier (low variance between readers), the CNN performs at a level that is consistent with the best expert readers.

CNN predictions were plotted against the overall mean (all expert readers and CNN) and the mean of the expert readers only (Fig. 6). CNN predictions are shown together with the deviation from the mean estimated age. CNN predictions are clustered around the mode of the river age distribution (three years), overpredicting the age of one- and two-year-olds and underpredicting the age of four-year-olds.

There was notable disagreement between readers in some instances of sea age (scale #40–50 and #95–125 in Fig. 6). Expert readers and CNN predictions agree about sea age on around 45 images (#50-95 in Fig. 6). The deviation of CNN predictions from the mean highlights how the CNN performs relative to the mean. It is interesting to note that CNN predictions are clustered around integer ages even when there is disagreement between expert readers. This was also evident in the main analysis (Fig. 4). The CNN predicts one- and two-year-olds well for sea age (except for a few outliers), but the deviation increases substantially for three-year-olds. It is worth noting that the deviation from the mean for even the best expert reader increased at relatively high sea age estimates (see Fig. 5).

CNN predictions are clearly biased towards a river age of 3 years, over- and underpredicting the age of two- and four-year-olds, respectively (Fig. 6), and if rounded to the nearest integer, all but 15 of the scales would be predicted as having a river age of three years.

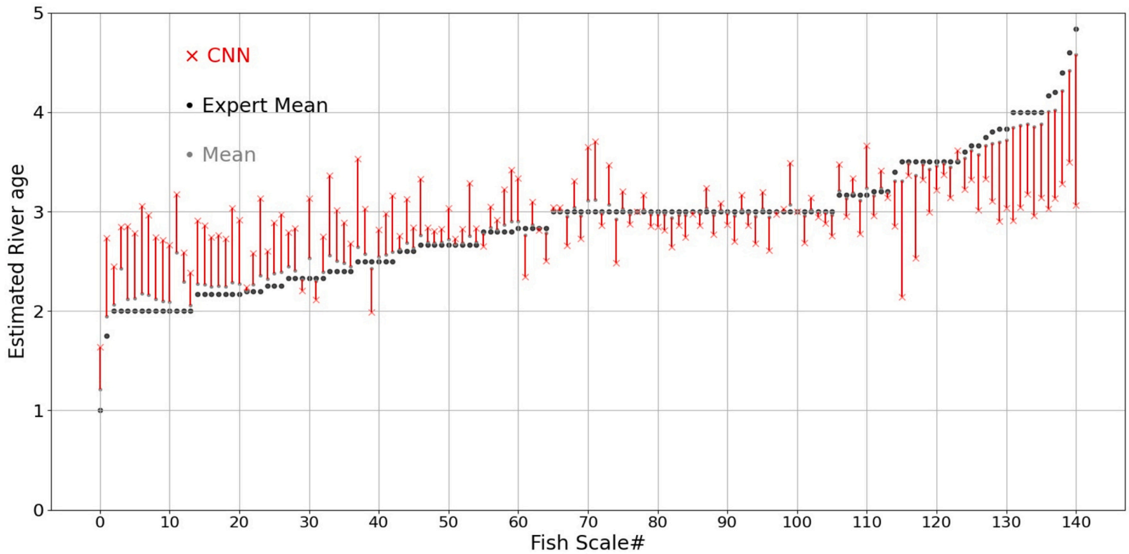
4. Discussion

We have used a state-of-the-art Convolutional Neural Network architecture, EfficientNet–B4, and adapted it to the task of classifying images of salmon scales. The CNN performs well on the binary classifications with a high accuracy in predicting both the origin (wild or farmed) and the spawning history (previous spawner or non-spawner). Regression CNNs performed well when predicting sea age (86.99% accuracy) but performed relatively poorly when predicting river age (63.20% accuracy).

4.1. Effect of nonuniform age distributions

Both regression CNNs underpredicted age in the high age classes (Figs. 3 and 4). The river age predicting CNN also overpredicted the age of one-year olds. The predictions are best for one, two- and three-years sea age and three years river age. This is likely caused by the training data not being uniformly distributed over ages. In our data set (Fig. 2), 90.2% have a river age of two or three years, while only 6% are four-year olds and 3% one-year olds. Sea age distribution is similarly nonuniform, where only 5% of the salmon are four years or older. The dataset, however, is not incomplete but rather reflects the natural age distribution of Atlantic salmon, related to specific ecological and biological factors (Otero et al., 2012a, 2012b, Wedemeyer et al., 1980). For instance, Atlantic salmon typically spend two or three years in the river, and more rarely spend one year or more than four years in the river. The relatively small number of training examples in infrequent age classes can lead to lower accuracy when compared with the more abundant age classes.

From our own experience, we frequently observe that CNN output is



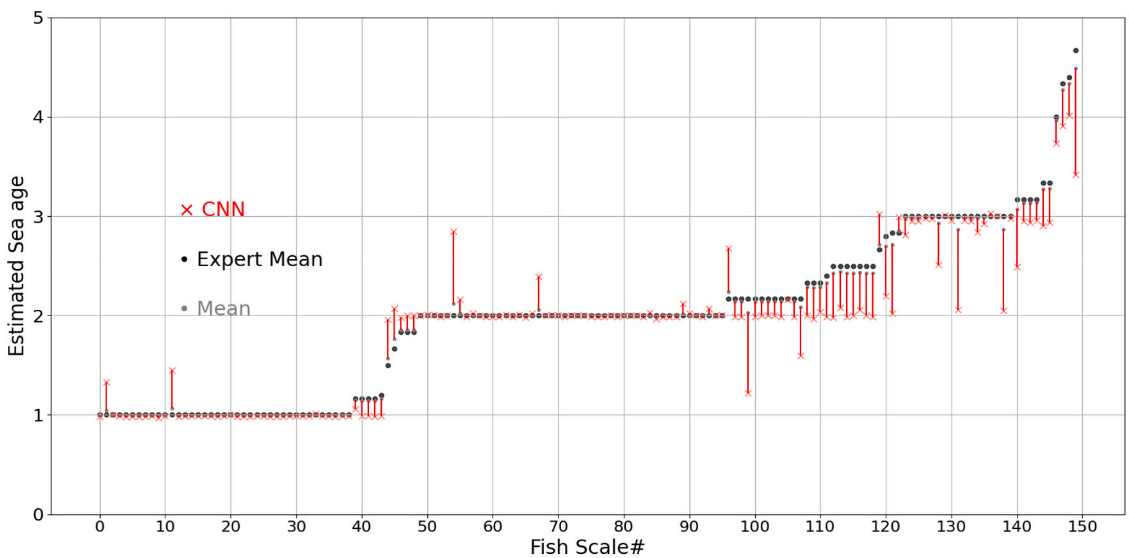


Fig. 6. CNN predictions and mean age estimates for each scale in the independent 150-image test set. Top: River age estimates. Bottom: Sea age estimates. The fish scales are sorted by the mean of the expert readers age estimates (black dots). The scattering of the CNN predictions around the mean is indicated with crosses. The vertical red lines are drawn between the CNN predictions and the mean (grey dots) to highlight how the CNN deviates from the mean. For river age, around 10 scales are not read by any expert reader and have therefore been omitted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

biased towards the mean of the training data and we suspect that this bias is stronger when there is less information in the input image. The river age CNN predicts an age close to the average in most cases. Nevertheless, the CNN shows a preference for integer values (Fig. 3), so while it may be unable to extract useful information from individual images, it has learned a prior distribution around integer ages from the aggregate data.

4.2. Comparing expert readers and CNNs performance

The sea age CNN performs on par with expert readers scoring second best on PA and CV measures (Table 3). The relatively poorer performance of the river age CNN is likely related to the compactness and pattern complexity of the inner core of the scale (Figs. 1 and 7). The core is shaped by river growth and has more irregular growth zones. In the river, many additional environmental factors (like river flow fluctuations, density-dependent effects, higher temperature amplitudes etc.) can also play a role. In contrast, during foraging in the sea where food availability is more abundant, salmon grow both faster and more regularly, making the growth zones larger and easier to distinguish. The poor performance of the river age CNN may therefore be a consequence of reducing image resolution to the default EfficientNet–B4 value of 380 × 380 pixels, resulting in a blurred core, insufficient in detail to enable extraction of pertinent features. The EfficientNet architecture is constructed to optimize the combination of input image resolution, layer depth and width. Here, we adhere to this design, and thus we did not consider alternative configurations. However, this inner core density of the scale creates problems even for expert readers, as it becomes difficult to distinguish the annual growth zones (Fig. 7) (ICES, 2011). This seems to be the case even when expert readers have access to a much higher image resolution of 2560 × 1920 pixels and access to magnified images. This is evident from the much larger disagreement among expert readers

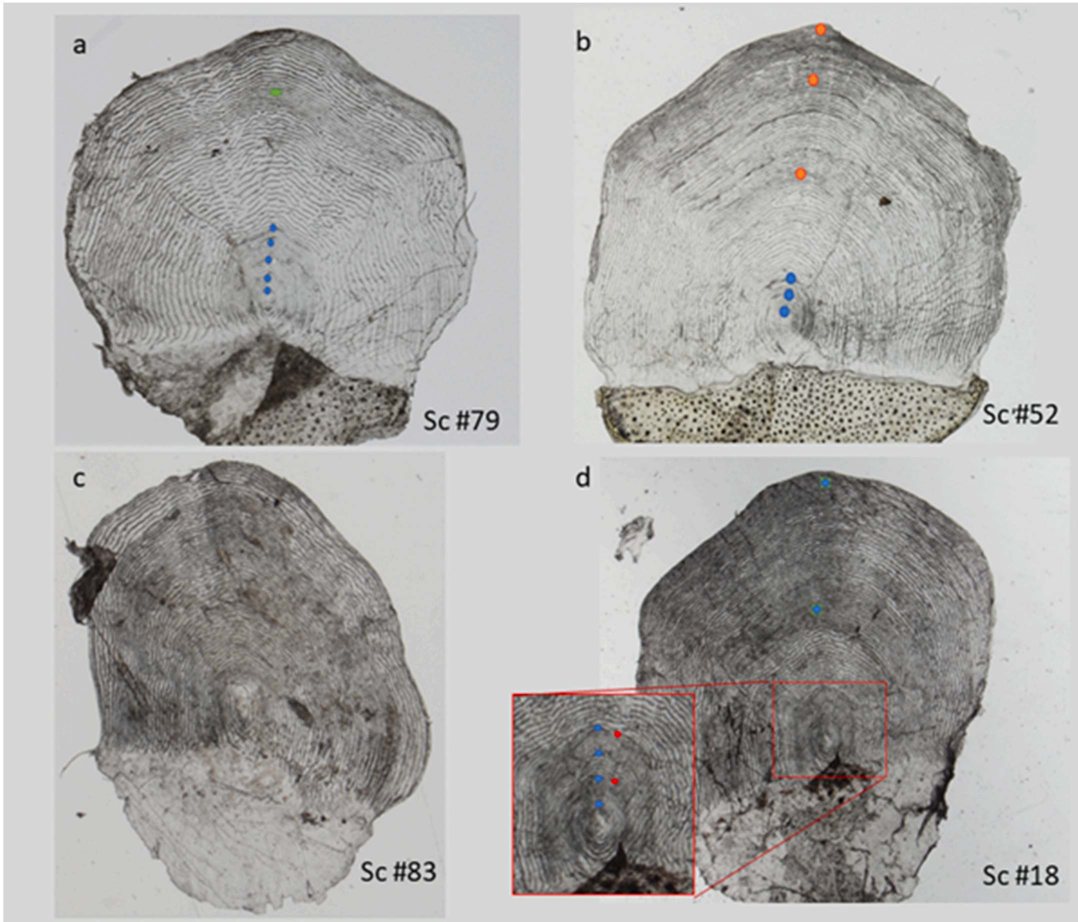


Fig. 7. Examples of salmon scale images with expert readers (ER) annotations of growth zones. Various combinations of bad/good CNN predictions from the 150-image test set is shown. The annotated coloured dots are indications of growth zones, i.e. number of years in the river and the sea, shown for one of the expert readers and for two expert readers in the inner zone of scale #18. **a**) Scale #79. Bad CNN and Good ER for river age. CNN predicts 3.06 years and expert reader modal age is 5 years. One reader estimates 4 years (83% agreement). **b**) Scale #52. Bad CNN and Good ER for sea age. CNN predicts 2.02 years and expert reader modal age is 3 years. 5 readers estimate 3 years, one estimates 2 years (83% agreement). **c**) Scale #83. Good CNN and Good ER for sea age. CNN prediction is 1.00 year and expert reader modal age was 1 year with 100% agreement. No annotation indicated. **d**) Scale #18. Controversial. CNN predicts 3.22 years river age. Only 40% agreement among expert readers, varying between 2, 3 and 4 years.

when estimating river age compared to sea age, as the highest agreement on reading river age (84.68%, reader 1) was less than the lowest (87.25%, reader 4) for sea age (Table 3). In Fig. 7d, a scale with low disagreement among expert readers on river age is shown including a close-up of the inner core of the scale, where two different readers have indicated their evaluation of yearly growth zones. A testimony from one of the readers demonstrates the difficulty of estimating river age: Paraphrasing: “My estimate (red dots) is two years in the river because the river is on the south west coast of Norway where growth conditions are good.”. The other reader (blue dots) identified four growth zones, i.e. four years in the river (Fig. 7). It is reasonable to assume that the expert readers would have had even lower agreement if they only had access to a 380 × 380 resolution image.

In addition, the expert readers had the advantage of using auxiliary information, including catch date, body length, back calculated smolt size, weight, sex and river location. Typically, growth conditions differ significantly from south to north and smolt size differs accordingly (Aronsen et al., 2019). Further, classification of a scale as farmed/wild, for instance, might guide the age reading because of differences in expected growth patterns. No such information was given to the CNNs during training but it is reasonable to expect that the CNNs performance would increase if some extra input information were provided during training and prediction e.g. CNNs predictions of wild/farmed or spawned/not spawned were fed into the regression CNNs. It is encouraging to note that the sea age CNN performed relatively well when compared to the mean estimates of the expert readers without using auxiliary background information (Table 3).

4.3. The importance of transfer learning

We found that the use of transfer learning when training our CNNs was crucial for their performance. Transfer learning is a common technique within deep learning, where networks trained to perform well on one task can be adapted with some training towards similar tasks (Yosinski et al., 2014). This is especially useful when datasets are sparse and more narrowly distributed. Often, a significant benefit can be gained by starting from a pre-trained network where the first and intermediate layers in the network already have been shaped by training on a large and diverse dataset. EfficientNet is available pretrained on the large benchmark dataset, ImageNet, containing around 1.4 million images in 1000 classes (Deng et al., 2009), more than 100 times the number of labelled salmon scale images used in this study. The network's ability to abstract lower level but more generally useful features has then already been established and new datasets presented to the network will only have to shape the upper layers where higher feature abstraction and categorisation takes place.

Transfer learning is generally more effective when the network has been pre-trained on a similar task, and we expect our trained CNNs can provide useful starting points for automating analysis of scales from other species, as well as for other hard parts of organisms (e.g., otoliths, vertebrae). Future fish ageing laboratories handling larger datasets of salmon scales could also benefit from using our trained CNNs as a basis for improved CNNs and automated systems.

4.4. The implications of deep learning

Several factors contribute to the ability of a machine learning system to solve a particular problem. The quality and quantity of training data, the appropriate neural network architecture and the available GPU resources, are all essential. Whenever supervised learning is used, the quality of the labelling is also integral to the data quality. If the majority of the training data is based on one expert readers labelling, this may not be ideal. Our study shows that even highly skilled experts disagree, and this disagreement is likely to reflect subjective biases arising from different experiences and maybe different emphasis on various background information. A CNN captures one reader's tendency to interpret images, thereby introducing some subjective bias indirectly into the network. In an ideal world with more resources the best labelling of an image would be gained from the modal reading from multiple expert readers. This is based on the statistical phenomena that the group average of several estimators – without systematic bias – of some unknown quantity tends to be closer to the truth than most of the single estimates. Using such a labelling approach would smoothen single reader biases and the trained CNN should tend to predict the group average.

Given abundant GPU resources an ideal approach would also be to combine predictions from an ensemble of architectural similar or even identical CNNs. These CNNs could use the same training data but altering the training process slightly by shuffling the sequence of data or by using different batch sizes. This approach of using multi-model ensembles to produce more reliable predictions have been explored in modelling studies within various fields (Kindt, 2018; Liao et al., 2014; Olsen et al., 2016). Future studies could exploit this, and it is expected that the group average prediction from several CNNs would outperform the performance of any single CNN in the very same group.

But do deep learning techniques that automate manual fish scale analysis improve efficiency? It can be argued that the process of collecting and pre-processing scales is truly the time-consuming part compared to the manual image interpretation. However, training a human reader to be a skilled expert reader is very time and resource demanding and an automated process can be scaled up more easily to handle larger datasets (Mahé, 2009). One major advantage of using deep learning instead of classical image analysis methods is that a deep neural network is partly able to capture the expertise of the readers and thereby function as a consistent future representative of this knowledge base. This can serve as an important reference for quality control, as an aid for training new expert readers and to reduce inherent vulnerability in depending on a small number of highly trained experts. The search for more objective techniques (Robertson and Morison, 1999), reducing the subjective nature of the human interpretation, also comply well with automation. Moreover, assuming long-term accuracy of the automatic classification, it may allow for the identification of age-reading ‘drifts’ through time, which result in increasing biases relative to the earlier determinations (Campana, 2001). Conversely, a system for interpreting scales that is stable over time can help to identify long term changes in the biology. For instance, the appearance of ‘false checks’ in salmon scales is reportedly becoming more common (see e.g., Todd et al., 2021). These interrelations between potential errors in automatic interpretation and potential errors due to human age-reading ‘drifts’ should be investigated with caution, and any automatic system must be regularly updated as data evolves.

The CNNs trained in this study are the first deep learning system to be able to automate the analysis of salmon scale images. They are prototypes with high potential to be further improved. Improvements can come from new available datasets, new CNN architectures and better quality assurance of labelling to ensure unbiased input data. Improved CNNs may also need to incorporate axillary biological information in the training process, and base analysis on several CNN predictions to overcome the challenges that were identified in this study. As a second step such automated systems could be deployed to become an integral part of future fish ageing laboratories adding complementary methodology which could help facilitate a more streamlined analysis.

5. Conclusions

Conservation and management of wild Atlantic salmon stocks are highly dependent on the biological and ecological information that can be extracted from fish scale features (Niemelä et al., 2006b). These natural markers in scales, integrated with automatic scale interpretation methods, offer potentially a cost-efficient and effective way of investigating salmon age and life-history traits. We have shown that convolutional neural networks can be successfully applied to extract age, origin, and spawning history from salmon scales, with a performance in prediction of sea age rivalling that of highly trained human experts. Although river age proved to be more difficult to predict, we believe that with suitable adaptations to the network and training procedure, it too can be adequately addressed. Deep learning offers a promising automation methodology for the analysis of salmon scale images, providing many benefits which could improve the quality of fish age estimation and support the management of these biological resources.

Data availability

Salmon scale 9056 image dataset: doi:10.21335/NMDC-1050865887

Expert reading of extra 150 image dataset: doi:10.21335/NMDC-1462728994

CNN based architecture: EfficientNet: https://github.com/qubvel/efficientnet

Pre-trained network weights (EfficientNet–B4): https://github.com/qubvel/efficientnet/blob/master/efficientnet/weights.py

Declaration of Competing Interest

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

Acknowledgements

We gratefully acknowledge valuable advice on the CNN architecture provided by Hyeongji Kim from the Institute of Marine Research, Bergen, Norway. This project was funded in part by Research Council of Norway projects 270966/O70 (COGMAR - Ubiquitous cognitive computer vision for marine services).

References

Aas, Ø., Einum, S., Klemetsen, A., Skurdal, J., 2011. Atlantic Salmon Ecology. John Wiley & Sons, pp. 1–467.
Abadi, M., et al., 2016. TensorFlow: A System for Large-Scale Machine Learning. OSDI 2016, Savannah, GA, 2-4 November 2016, pp. 265–283.
Anon, 2008. SALSEA-Merge - Workshop on Digital Scale Reading Methodology, Trondheim, Norway, 8th to 10th September 2008, pp. 1–23.
Anon, 2019. Rømt oppdrettslaks i vassdrag i 2018. Rapport fra det nasjonale overvåkningsprogrammet. Fisken og havet, særnr, pp. 4–2019. ISSN: 1894–5031.
Aronsen, T., Bakke, G.O., Barlaup, B.T., Berntsen, J.H.H., Diserud, O.H., Fiske, P., Fjeldheim, P.T., Larsen, B.F., Glover, K., Heino, M.P., Husebø, Å., 2019. Rømt oppdrettslaks i vassdrag i 2018. Rapport fra det nasjonale overvåkningsprogrammet. Fisken og havet, særnr, 4–2019. ISSN: 1894–5031.
Boehlert, G.W., 1985. Using objective criteria and multiple regression models. Fish. Bull. 83, 103–117.

Butler, J.R.A., Radford, A., Riddington, G., Laughton, R., 2009. Evaluating an ecosystem service provided by Atlantic salmon, sea trout and other fish species in the river Spey, Scotland: the economic impact of recreational rod fisheries. Fish. Res. 96, 259–266.

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.

Campana, S.E., Annand, M.C., Mcmillan, J.I., 1995. Graphical and statistical methods for determining the consistency of age determinations. Trans. Am. Fish. Soc. 124, 131–138.

Chollet, F., 2015. Keras. Available at: https://github.com/fchollet/keras.

Crozier, W.W., Schön, P.J., Chaput, G., Potter, E.C.E., Maoiléidigh, Ó., N., and MacLean, J. C., 2004. Managing Atlantic salmon (Salmo salar L.) in the mixed stock environment: challenges and considerations. ICES J. Mar. Sci. 61, 1344–1358.

Dahl, K., 1911. The Age and Growth of Salmon and Trout in Norway as Shown by their Scales, IX. London Salmon Trout Association, p. 141.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A Large-Scale Hierarchical Image Database (In CVPR09).

Erkinaro, J., Czorlich, Y., Orell, P., Kuusela, J., Falkegård, M., Länsman, M., Pulkkinen, H., et al., 2019. Life history variation across four decades in a diverse population complex of Atlantic salmon in a large subarctic river. Can. J. Fish. Aquat. Sci. 76, 42–55.

Fablet, R., 2006a. Semi-local extraction of ring structures in images of biological hard tissues: application to the Bayesian interpretation of fish otoliths for age and growth estimation. Can. J. Fish. Aquat. Sci. 63, 1414–1428. https://doi.org/10.1139/F06-001.

Fablet, R., 2006b. Statistical learning applied to computer-assisted fish age and growth estimation from otolith images. Fish. Res. 81, 219–228. https://doi.org/10.1016/j.fishres.2006.07.013.

Fisher, M., Hunter, E., 2018. Digital Imaging Techniques in Otolith Data Capture, Analysis and Interpretation, 598, pp. 213–231.

Fisher, J.P., Pearcy, W.G., 1990. Spacing of scale circuli versusgrowth rate in young coho salmon. Fish. Bull. 88, 637–643.

Ford, J.S., Myers, R.A., 2008. A global assessment of salmon aquaculture impacts on wild salmonids. PLoS Biol. 6, 0411–0417.

Francis, R.I.C.C., 1990. Back-calculation of fish length: a critical review. J. Fish Biol. 36, 883–902.

Hansen, L.P., Quinn, T.P., 1998. The marine phase of the Atlantic salmon (Salmo salar) life cycle, with comparisons to Pacific salmon. Can. J. Fish. Aquat. Sci. 55, 104–118.

Huang, Yanping, et al., 2018. GPipe: Efficient Training of Giant Neural Networks Using Pipeline Parallelism. ArXiv Preprint ArXiv:1811.06965.

Husebø, Å., Andersen, K.C., Bakke, G., Fjeldheim, P.T., Knutar, T.S., Østbø, B.I., 2020. Atlantic Salmon Scale Age Reading Data. https://doi.org/10.21335/NMDC-1462728994.

Ibáñez, A.L., Britton, J.R., Cowx, I.G., 2008. Relationship between scale growth checks, circuli formation rate and somatic growth in Rutilus rutilus (L.) a fish farm-reared cyprinid. J. Fish Biol. 72, 1023–1034.

ICES, 2011. Report of the Workshop on Age Determination of Salmon (WKADS), p. 66.

ICES, 2013. Report of the Second Workshop on Age Determination of Salmon (WKADS2). 4th-6th September 2012, Derry, Northern Ireland. ICES WKADS2 report 2012, ICES CM 2012/ACOM:61ICES CM 2012/ACOM:61, pp. 1–28.

Kindt, Roeland, 2018. Ensemble species distribution modelling with transformed suitability values. Environ. Model. Softw. 100, 136–145.

Krizhevsky, Alex, Sutskever, Ilya, Hinton, Geoffrey E., 2012. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, 2, pp. 1097–1105.

LeCun, Yann, Bengio, Yoshua, Hinton, Geoffrey, May 28, 2015. Deep learning. Nature 521 (7553), 436–444.

Liao, K., Xua, F., Zheng, J., Zhu, Q., Yang, G., 2014. Using different multimodel ensemble approaches to simulate soil moisture in a forest site with six traditional pedotransfer functions. Environ. Model. Softw. 57, 27–32.

Mahé, K., 2009. AFISA Automated FISh Ageing.

Malde, K., Handegard, N.O., Eikvil, L., Salberg, A.B., July-August 2020. Machine intelligence and the data-driven future of marine science. ICES J. Mar. Sci. 77 (4), 1274–1285. https://doi.org/10.1093/icesjms/fsz057.

McGinnity, P., Prodöhl, P., Ferguson, A., Hynes, R., Maoiléidigh, N.Ó., Baker, N., Cotter, D., et al., 2003. Fitness reduction and potential extinction of wild populations of Atlantic salmon, Salmo salar, as a result of interactions with escaped farm salmon. Proc. R. Soc. B Biol. Sci. 270, 2443–2450.

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, 87. https://doi.org/10.1371/journal.pone.0204713.

Niemelä, E., Erkinaro, J., Julkunen, M., Hassinen, E., Länsman, M., Brørs, S., 2006a. Temporal variation in abundance, return rate and life histories of previously spawned Atlantic salmon in a large subarctic river. J. Fish Biol. 68, 1222–1240.

Niemelä, E., Orell, P., Erkinaro, J., Dempson, J.B., BrØrs, S., Svenning, M.A., Hassinen, E., 2006b. Previously spawned Atlantic salmon ascend a large subarctic river earlier than their maiden counterparts. J. Fish Biol. 69, 1151–1163.

Olsen, Erik, Fay, Gavin, Gaichas, Sarah, Gamble, Robert, Lucey, Sean, Link, Jason S., 2016. Ecosystem model skill assessment. Yes we can! PLoS One 11 (1), e0146467.

Otero, J., Jensen, A.J., L'Abeé-Lund, J.H., Stenseth, N.C., Storvik, G.O., Vøllestad, L.A., 2012a. Contemporary Ocean warming and freshwater conditions are related to later sea age at maturity in Atlantic salmon spawning in Norwegian rivers. Ecol. Evol. 2, 2192–2203.

Otero, J., Jensen, A.J., L'Abée-Lund, J.H., Stenseth, N.C., Storvik, G.O., Vøllestad, L.A., 2012b. Contemporary Ocean warming and freshwater conditions are related to later sea age at maturity in Atlantic salmon spawning in Norwegian rivers. Ecol. Evol. 2 (9), 2192–2203.

Panfili, J., de Pontual, H., Troadec, H., Wright, P.J., 2002. Manual of Fish Sclerochronology. IFREMER-IRD, Brest,France.

Pedregosa, F., Varoquaux, Gaël., Gramfort, A., Michel, V., Thirion, B., Grisel, O.et al., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12 (Oct), 2825–2830.

Ricker, W., 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Canada 191, 382. https://doi.org/10.1038/108070b0.

Robertson, S.G., Morison, A.K., 1999. A trial of artificial neural networks for automatically estimating the age of fish. Mar. Freshw. Res. 50, 73–82. https://doi.org/10.1071/MF98039.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. IJCV.

Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117.

Shearer, W.M., 1989. Report of the second Atlantic Salmon Scale Reading Workshop, Aberdeen, Scotland, 12-14 October 1989. ICES rapport nr. CM, M, 7, p. 18.

Atlantic salmon scale reading guidelines. In: Shearer, W.M. (Ed.), 1992. ICES Cooperative Research Report, vol. 188. ICES, Copenhagen, p. 46.

Spurgeon, Jonathan, Hamel, Marty, Pope, Kevin, Pegg, Mark, 2015. The global status of freshwater fish age validation studies and a prioritization framework for further research. Rev. Fish. Sci. Aquac. 23 https://doi.org/10.1080/23308249.2015.1068737.

Stokesbury, M., 1997. High incidence of hatchery origin Atlantic salmon in the smolt output of a Canadian River. ICES J. Mar. Sci. 54, 1074–1081.

Szegedy, Christian, Liu, Wei, Jia, Yangqing, Sermanet, Pierre, Reed, Scott, Anguelov, Dragomir, Erhan, Dumitru, Vanhoucke, Vincent, Rabinovich, Andrew, 2015. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

Tan, Mingxing, Quoc, V.Le., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In: arXiv preprint arXiv:1905.11946.

Thomas, K., Hansen, T., Brophy, D., O Maoiléidigh, N., Fjelldal, P.G., 2019. Experimental investigation of the effects of temperature and feeding regime on scale growth in Atlantic salmon Salmo salar post-smolts. J. Fish Biol. 94, 896–908. https://doi.org/10.1111/jfb.13971.

Todd, C.D., Hanson, N.N., Boehme, L., Revie, C.W., Marques, A.R., 2021. Variation in the post-smolt growth pattern of wild one sea-winter salmon (Salmo salar L.), and its linkage to surface warming in the eastern North Atlantic Ocean. J. Fish Biol. 98, 6–16. https://doi.org/10.1111/jfb.14552.

Wedemeyer, G.A., Saunders, R.L., Clarke, W.C., 1980. Environmental factors affecting smoltification and early marine survival of anadromous salmonids. North Am. Salmon Res. Cent. Rep. 5.

Yosinski, Jason, Clune, Jeff, Bengio, Yoshua, Lipson, Hod, 2014. How transferable are features in deep neural networks?. In: Advances in neural Information Processing Systems, pp. 3320–3328.